Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 138 (2007) 24–46
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Testing with many weak instruments

Donald W.K. Andrews[a,*], James H. Stock[b]

[a]Cowles Foundation for Research in Economics, Yale University, P.O. Box 208281, New Haven, CT 06520-8281, USA
[b]Department of Economics, Harvard University, MA, USA

Available online 10 July 2006

Abstract

This paper establishes the asymptotic distributions of the likelihood ratio (LR), Anderson–Rubin (AR), and Lagrange multiplier (LM) test statistics under "many weak IV asymptotics." These asymptotics are relevant when the number of IVs is large and the coefficients on the IVs are relatively small. The asymptotic results hold under the null and under suitable alternatives. Hence, power comparisons can be made.

Provided $k^3/n \to 0$ as $n \to \infty$, where n is the sample size and k is the number of instruments, these tests have correct asymptotic size. This holds no matter how weak the instruments are. Hence, the tests are robust to the strength of the instruments. The asymptotic power results show that the conditional LR test is more powerful asymptotically than the AR and LM tests under many weak IV asymptotics.

JEL classification: C12; C30

Keywords: Anderson–Rubin test; Conditional likelihood ratio test; Instrumental variables; Lagrange multiplier test; Many instrumental variables; Weak instruments

1. Introduction

This paper contributes to the literature on weak instrumental variables (IVs) in linear IV models. The weak IV literature documents that standard procedures, such as two-stage least squares-based t tests and confidence intervals, perform poorly when the IVs are weak

*Corresponding author. Tel.: +1 203 432 3698; fax: +1 203 432 6167.
E-mail address: donald.andrews@yale.edu (D.W.K. Andrews).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2006.05.012

(i.e., when the IVs are only weakly correlated with the right-hand side endogenous variables). In consequence, alternative testing procedures have been developed whose size is robust to the strength of the IVs. Such tests include the Anderson and Rubin (1949) (AR) test, the Lagrange multiplier (LM) test introduced in Kleibergen (2002) and Moreira (2001), and the conditional likelihood ratio (CLR) test introduced in Moreira (2003). Andrews et al. (2006a) have shown that the CLR test has near optimal power properties in models with Gaussian errors within a class of invariant similar tests. Furthermore, the robustness of the asymptotic size and power properties of the AR, LM, and CLR tests to non-normality has been established under the "weak IV asymptotics" of Staiger and Stock (1997), see the references above.

This paper contributes to the literature by analyzing the behavior of the AR, LM, and CLR tests when the IVs may be weak, the number of IVs, k, may be relatively large, and the equation errors may be non-normal. Specifically, the paper presents new results for these tests in the linear IV regression model under "many weak IV asymptotics" in which $k \to \infty$ as the sample size, n, goes to infinity and the strength of the IVs may be weak. Asymptotics of this type have been considered recently by Chao and Swanson (2005), Stock and Yogo (2005), Han and Phillips (2006), Anderson et al. (2005), Hansen et al. (2005), Newey and Windmeijer (2005), and Andrews and Stock (2006). Most of these papers focus on the properties of estimators. In contrast, we are interested in the properties of tests—both for testing purposes and for obtaining confidence intervals via inversion. In particular, we are interested in the properties of tests when the equation errors are non-normal.

We find that in the many weak IV asymptotic setup the CLR, AR, and LM tests are completely robust asymptotically to weak IVs with normal and non-normal errors. That is, the asymptotic levels of the tests are correct no matter how weak are the IVs. On the other hand, the asymptotic levels of the CLR, AR, and LM tests are not completely robust to the magnitude of k relative to n. One does not want to take k too large relative to n. Results of Andrews and Stock (2006) for the case of normal errors indicate that the condition $k^{3/2}/n \to 0$ as $n \to \infty$ is necessary for correct asymptotic size.[1] With non-normal errors, the results of this paper show that a sufficient condition for correct asymptotic size is $k^3/n \to 0$ as $n \to \infty$. Although this condition covers many cases of interest, it can be restrictive. For example, it is not suitable for the Angrist and Krueger (1991) example when one interacts the quarter of birth IV with state dummies to yield $k = 180$ and $n = 329,509$. Whether the condition $k^3/n \to 0$ is necessary is an open question (see the discussion below).

Andrews and Stock (2006) show that the CLR test is essentially on the asymptotic power envelope for normal errors under many weak IV asymptotics—regardless of the relative strength of the IVs to k in the asymptotics. In addition, the AR and LM tests are found not to be on the power envelope. In the present paper, we show that the asymptotic power properties of the CLR, AR, and LM tests are the same under non-normal errors as under normal errors given the $k^3/n \to 0$ condition. The aforementioned results combine to establish that the CLR test has power advantages over the AR and LM tests for non-normal as well as normal errors.

[1] This condition is necessary for the estimator of the reduced-form variance matrix to be $k^{1/2}$-consistent, and $k^{1/2}$-consistency of this estimator is necessary for the effect of estimation of the variance matrix to be asymptotically negligible.

We conclude that the "many weak IV" asymptotic results for non-normal errors given in the present paper buttress the arguments in Andrews et al. (2006a) and Andrews and Stock (2006) for employing the CLR test over the AR, LM, and other tests in model scenarios with potentially weak IVs.

The proofs of the results given here make use of the degenerate U-statistic central limit theorem of Hall's (1984), as in Newey and Windmeijer (2005).

Other papers in the literature that consider many weak IVs, include Chamberlain and Imbens (2004) and Chao and Swanson (2006). Bekker and Kleibergen (2003) consider "many irrelevant IVs asymptotics," in which $k \to \infty$ as $n \to \infty$ and the reduced-form coefficient matrix on the IVs is zero. Weak IV asymptotics (with k fixed) were introduced in Staiger and Stock (1997). Many IV asymptotics (with strong IVs) have been employed in Anderson (1976), Kunitomo (1980), Morimune (1983), Bekker (1994), Donald and Newey (2001), Hahn (2002), Hahn et al. (2004), and Hansen et al. (2005) among others.

This paper is organized as follows. Section 2 introduces the model and assumptions employed. Section 3 defines the CLR, AR, and LM tests. Section 4 gives the results. An Appendix provides the proofs.

2. Model and assumptions

The model we consider is an IV regression model with one endogenous right-hand side (rhs) variable, p exogenous variables, and k IVs. The sample size is n. The number of IVs, k, depends on n, i.e., $k = k_n$. We note that the case of a single rhs endogenous variable is by far the most important in empirical applications.

The model consists of a structural equation and a reduced-form equation:

$$\begin{aligned} y_1 &= y_2\beta + X\gamma_1 + u, \\ y_2 &= \widetilde{Z}\pi + X\xi_1 + v_2, \end{aligned} \tag{1}$$

where $y_1, y_2 \in R^n$, $X \in R^{n\times p}$, and $\widetilde{Z} \in R^{n\times k}$ are observed variables; $u, v_2 \in R^n$ are unobserved errors; and $\beta \in R$, $\pi \in R^k$, $\gamma_1 \in R^p$, and $\xi_1 \in R^p$ are unknown parameters. The exogenous variable matrix X and the IV matrix $\widetilde{Z}$ are random. The $n \times 2$ matrix of errors $[u : v_2]$ is iid across rows. The variable y_2 is endogenous in the equation for y_1 (i.e., y_2 and u may be correlated). Endogeneity may be due to simultaneity, left-out variables, or mismeasurement of an exogenous variable.

The two reduced-form equations are

$$\begin{aligned} y_1 &= \widetilde{Z}\pi\beta + X\gamma + v_1, \\ y_2 &= \widetilde{Z}\pi + X\xi_1 + v_2, \end{aligned}$$

where

$$v_1 = u + v_2\beta \quad \text{and} \quad \gamma = \gamma_1 + \xi_1\beta. \tag{2}$$

The reduced-form errors $[v_1 : v_2]$ are iid across rows with each row having mean zero and 2×2 non-singular covariance matrix Ω.

Let $Y = [y_1 : y_2] \in R^{n\times 2}$ and $V = [v_1 : v_2] \in R^{n\times 2}$ denote the matrices of endogenous variables and reduced-form errors, respectively. We write the ith rows of Y, V, X, and $\widetilde{Z}$ as the column vectors $Y_i, V_i \in R^2$, $X_i \in R^p$, and $\widetilde{Z}_i \in R^k$, respectively. The two equation reduced-form model can be written as

$$Y_i = a\pi'\widetilde{Z}_i + \eta' X_i + V_i \quad \text{for } i \leqslant n,$$

where

$$a = (\beta, 1)' \quad \text{and} \quad \eta = [\gamma : \xi_1] \in R^{p\times 2}. \tag{3}$$

Define

$$Z_i^* = \widetilde{Z}_i - E\widetilde{Z}_i X_i'(EX_iX_i')^{-1}X_i$$

and

$$\lambda_{n,k}^* = n\pi' EZ_i^* Z_i^{*'}\pi. \tag{4}$$

$\lambda_{n,k}^*$ indicates the strength of the IVs (and is proportional to the concentration parameter).

We use the following assumptions.

Assumption 1. $\{(V_i, X_i, \widetilde{Z}_i) : i \leqslant n\}$ are iid across i for each n and $\{(V_i, X_i) : i \leqslant n; n \geqslant 1\}$ are identically distributed across i and n.

Assumption 2. $EV_i = 0$, $EV_i\widetilde{Z}_i' = 0$, $EV_iX_i' = 0$, EX_iX_i' is pd, $\liminf_{n\to\infty} \lambda_{\min}(EZ_i^*Z_i^{*'}) > 0$, and $\sup_{j\leqslant k; n\geqslant 1}(E\|V_i\|^4\widetilde{Z}_{ij}^4 + E\|V_i\|^4 + E\widetilde{Z}_{ij}^4 + E\|X_i\|^4) < \infty$, where $\widetilde{Z}_i = (\widetilde{Z}_{i1}, \ldots, \widetilde{Z}_{ik})'$.

Assumption 3. $EV_iV_i' = \Omega$, $E(V_iV_i' \otimes Z_i^*Z_i^{*'}) = \Omega \otimes \mathrm{E}Z_i^*Z_i^{*'}$ for all $n \geqslant 1$, and Ω is pd.

Assumption 4. $k \to \infty$ and $k^3/n \to 0$ as $n \to \infty$, and p does not depend on n.

Assumption 5. $\lambda_{n,k}^*/k^\tau \to r_\tau$ as $n \to \infty$ for some constants $r_\tau \in [0, \infty)$ and $\tau \in (0, \infty)$.

Assumption 6. β is fixed for all n when $\tau \leqslant \frac{1}{2}$; $\beta = \beta_0 + Bk^{1/2-\tau}$ when $\tau \in (\frac{1}{2}, 1]$; and $\beta = \beta_0 + Bk^{-\tau/2}$ when $\tau \geqslant 1$.

Assumption 1 states that the errors, exogenous variables, and IVs are random and iid across $i \leqslant n$. Note that $(V_i, X_i, \widetilde{Z}_i)$ cannot be iid across n because the dimension, k, of $\widetilde{Z}_i$ depends on n.

Assumption 2 requires that the IVs and exogenous variables are uncorrelated with the reduced-form errors and satisfy standard moment conditions.

Assumption 3 implies that the reduced-form errors are homoskedastic.

Assumption 4 states that the number of IVs goes to infinity as $n \to \infty$, but not too quickly, and the number of exogenous variables is fixed. As noted in Section 1, the condition $k^3/n \to 0$ can be restrictive, and may not be necessary (at least in the presence of suitable additional assumptions). This condition is not used in the proof to obtain a central limit theorem. Rather, it is used to show that one obtains the same limit distribution of a weighted quadratic form in a k-vector sample average when (i) a population weight matrix is replaced by its sample version (see Lemmas 5 and 7 in the Appendix), and (ii) the k-vector sample average based on a population projection onto an intercept and other exogenous variables is replaced by the corresponding k-vector sample average based on the sample projection (see Lemmas 6 and 7 in the Appendix). The argument in Hansen et al. (2005, Lemma A9) probably can be used to show that the $k^3/n \to 0$ condition is stronger than necessary for the purpose (i) above. However, weakening the $k^3/n \to 0$ condition and still showing (ii) is problematic.

An alternative approach to that considered in this paper is to treat the IVs and exogenous variables as fixed rather than random. With this approach, one does not need to establish (i) and (ii) above because a CLT can be applied directly to the quantities that involve the sample weight matrix and the sample projections. The drawback of this approach is that sufficient conditions for a CLT involve conditions that depend on the relative magnitudes of k and n in an opaque way and are difficult to verify. For example, conditions arise on the magnitudes of the elements of the projection matrix onto the $n \times k$ space spanned by the IVs Z. Rather than rely on conditions of this sort, in this paper we use an approach that yields a condition on k that is clean and transparent, but may be stronger than necessary.

Assumption 5 controls the relative magnitude of the IV strength, as measured by $\lambda^*_{n,k}$, to the number of IVs k. For example, Assumption 5 holds if $\pi = C(k^\tau/n)^{1/2}$ for some $C \in R^k$ with $\|C\| = 1$ and $CEZ_i^* Z_i^{*'} C \to r_\tau$. The smaller is τ, the weaker are the IVs relative to k. Andrews and Stock (2006) find that the key value of τ for inference concerning β is $\tau = 1/2$. For $\tau = 1/2$, some tests (such as the CLR, AR, and LM tests) have non-trivial power asymptotically against fixed alternatives. For $\tau > 1/2$, these tests have asymptotic power equal to one against any fixed alternative. Many of the papers in the many weak IV literature only consider the case of $\tau = 1$. Note that Assumptions 2 and 5 imply that $\pi'\pi = O(k^\tau/n)$, see (15) below.

Assumption 6 specifies the true value of β that is considered in the results below. Assumption 6 takes β such that the asymptotic distributions of the test statistics considered are non-degenerate. It is shown that this requires that β is a fixed value when $\tau \leqslant 1/2$ and β is a sequence of local alternatives to the null value β_0 when $\tau > 1/2$. Of course, $\beta = \beta_0$ is allowed when $\tau \leqslant 1/2$ or $\tau > 1/2$.

3. Tests

In applications, interest often is focused on the parameter β on the rhs endogenous variable y_2. Hence, our interest is in the null and alternative hypotheses

$$\mathrm{H}_0 : \beta = \beta_0 \quad \text{and} \quad \mathrm{H}_1 : \beta \neq \beta_0. \tag{5}$$

The parameter π, which determines the strength of the IVs, is a nuisance parameter that appears under the null and alternative hypotheses. The parameters γ_1, ξ_1, and Ω also are nuisance parameters, but are of lesser importance because tests concerning β typically are invariant to γ_1 and ξ_1 and the behavior of standard tests, such as t tests, are much less sensitive to Ω than to π.

We now define the AR, LM, and CLR tests. We estimate Ω ($\in R^{2\times 2}$) via

$$\widehat{\Omega}_n = (n-k-p)^{-1}\widehat{V}'\widehat{V}, \quad \text{where } \widehat{V} = Y - P_{[\widetilde{Z}:X]}Y \tag{6}$$

and $P_A = A(A'A)^{-1}A'$ for a non-singular matrix A. We define

$$\widehat{S}_n = (Z'Z)^{-1/2}Z'Yb_0 \cdot (b_0'\widehat{\Omega}_n b_0)^{-1/2}, \quad \text{where } Z = \widetilde{Z} - P_X\widetilde{Z} \quad \text{and} \quad b_0 = (1, -\beta_0)',$$

$$\widehat{T}_n = (Z'Z)^{-1/2}Z'Y\widehat{\Omega}_n^{-1}a_0 \cdot (a_0'\widehat{\Omega}_n^{-1}a_0)^{-1/2}, \quad \text{where } a_0 = (\beta_0, 1)',$$

$$\widehat{Q}_{\lambda,k,n} = [\widehat{S}_n : \widehat{T}_n]'[\widehat{S}_n : \widehat{T}_n] = \begin{bmatrix} \widehat{S}_n'\widehat{S}_n & \widehat{S}_n'\widehat{T}_n \\ \widehat{T}_n'\widehat{S}_n & \widehat{T}_n'\widehat{T}_n \end{bmatrix} = \begin{bmatrix} \widehat{Q}_{S,n} & \widehat{Q}_{ST,n} \\ \widehat{Q}_{ST,n} & \widehat{Q}_{T,n} \end{bmatrix},$$

and

$$\widehat{\widehat{Q}}_{\lambda,k,n} = (\widehat{Q}_{\lambda,k,n} - kI_2)/k^{1/2}. \tag{7}$$

The AR, LM, and LR test statistics can be written as

$$\widehat{AR}_n = \widehat{Q}_{S,n}/k,$$
$$\widehat{LM}_n = \widehat{Q}^2_{ST,n}/\widehat{Q}_{T,n},$$

and

$$\widehat{LR}_n = \frac{1}{2}\left(\widehat{Q}_{S,n} - \widehat{Q}_{T,n} + \sqrt{(\widehat{Q}_{S,n} - \widehat{Q}_{T,n})^2 + 4\widehat{Q}^2_{ST,n}}\right) \tag{8}$$

(see Moreira, 2003 and Andrews et al., 2006a).

Under H_0, $\widehat{AR}_n \to_d \chi^2_k/k$ and $\widehat{LM}_n \to_d \chi^2_1$ as $n \to \infty$ under strong and weak IV asymptotics assuming iid homoskedastic errors and k fixed for all n (e.g., see Andrews et al., 2006a). Under the additional assumption of normal errors, $\widehat{AR}_n \sim F_{k,n-k-p}$. Hence, an F critical value is typically employed with the AR test, and a χ^2_1 critical value is used for the LM test.

The CLR test rejects the null hypothesis when

$$\widehat{LR}_n > \kappa_{LR,\alpha}(\widehat{Q}_{T,n}), \tag{9}$$

where the conditional critical value function $\kappa_{LR,\alpha}(\widehat{Q}_{T,n})$ is defined to satisfy $P_{\beta_0}(\widehat{LR}_n > \kappa_{LR,\alpha}(q_T)|\widehat{Q}_{T,n} = q_T) = \alpha$. Andrews et al. (2006b) gives detailed tables of $\kappa_{LR,\alpha}(q_T)$ values.

4. Asymptotic results

This section contains the results of the paper. We establish the asymptotic distributions of the statistic $\widehat{\widehat{Q}}_{\lambda,k,n}$ and the test statistics $\widehat{AR}_n$, $\widehat{LM}_n$, and $\widehat{LR}_n$, which depend on $\widehat{\widehat{Q}}_{\lambda,k,n}$, under many weak IV asymptotics. In contrast to the assumptions in Andrews and Stock (2006), we do not assume that the errors are normally distributed. We use the asymptotic distributions to show that the AR, LM, and CLR tests have correct asymptotic size under many weak IV asymptotics. The asymptotic distributions also show that the many weak IV asymptotic power of the AR, LM, and CLR tests under non-normality is the same as it is under normality. Hence, the power comparisons given in Andrews and Stock (2006) under normality also hold asymptotically under non-normality.

The asymptotic distribution of $\widehat{\widehat{Q}}_{\lambda,k,n}$ depends on the following quantities:

$$c_\beta = (\beta - \beta_0) \cdot (b_0'\Omega b_0)^{-1/2} \in R,$$
$$d_\beta = a'\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \in R,$$
$$V_{3,\tau} = \begin{cases} Diag\{2,1,2\} & \text{if } 0<\tau\leqslant 1/2, \\ Diag\{2,1,0\} & \text{if } 1/2<\tau<1, \\ Diag\{2,1+d^2_{\beta_0}r_1,0\} & \text{if } \tau = 1, \\ Diag\{2,d^2_{\beta_0}r_\tau,0\} & \text{if } \tau>1, \quad \text{and} \end{cases}$$

$$\gamma_B = (b_0'\Omega b_0)^{-1/2} d_{\beta_0} \quad B \in R, \tag{10}$$

for the scalar constant B given in Assumption 6. Let $\chi_1^2(\delta)$ denote a noncentral chi-square distribution with one degree of freedom and non-centrality parameter δ.

Theorem 1. *Suppose Assumptions* 1–6 *hold. Then, the following results hold*:

(a) *If* $0<\tau<1/2$ *and* β *is fixed*,

$$\begin{pmatrix} (\widehat{S}_n'\widehat{S}_n - k)/k^{1/2} \\ \widehat{S}_n'\widehat{T}_n/k^{1/2} \\ (\widehat{T}_n'\widehat{T}_n - k)/k^{1/2} \end{pmatrix} \to_d \begin{pmatrix} \overline{Q}_{S,\infty} \\ \overline{Q}_{ST,\infty} \\ \overline{Q}_{T,\infty} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, V_{3,\tau} \right),$$

$$(\widehat{AR}_n - 1)k^{1/2} \to_d \overline{Q}_{S,\infty} \sim \mathrm{N}(0,2), \quad \widehat{LM}_n \to_d \overline{Q}_{ST,\infty}^2 \sim \chi_1^2(0), \ \textit{and}$$

$$\widehat{LR}_n/k^{1/2} \to_d \frac{1}{2}\left(\overline{Q}_{S,\infty} - \overline{Q}_{T,\infty} + \sqrt{(\overline{Q}_{T,\infty} - \overline{Q}_{S,\infty})^2 + 4\overline{Q}_{ST,\infty}^2} \right).$$

(b) *If* $\tau = 1/2$ *and* β *is fixed*,

$$\begin{pmatrix} (\widehat{S}_n'\widehat{S}_n - k)/k^{1/2} \\ \widehat{S}_n'\widehat{T}_n/k^{1/2} \\ (\widehat{T}_n'\widehat{T}_n - k)/k^{1/2} \end{pmatrix} \to_d \begin{pmatrix} \overline{Q}_{S,\infty} \\ \overline{Q}_{ST,\infty} \\ \overline{Q}_{T,\infty} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} c_\beta^2 r_{1/2} \\ c_\beta d_\beta r_{1/2} \\ d_\beta^2 r_{1/2} \end{pmatrix}, V_{3,1/2} \right),$$

$$(\widehat{AR}_n - 1)k^{1/2} \to_d \overline{Q}_{S,\infty} \sim \mathrm{N}(c_\beta^2 r_{1/2}, 2),$$

$$\widehat{LM}_n \to_d \overline{Q}_{ST,\infty}^2 \sim \chi_1^2(c_\beta^2 d_\beta^2 r_{1/2}^2), \ \textit{and}$$

$$\widehat{LR}_n/k^{1/2} \to_d \frac{1}{2}\left(\overline{Q}_{S,\infty} - \overline{Q}_{T,\infty} + \sqrt{(\overline{Q}_{T,\infty} - \overline{Q}_{S,\infty})^2 + 4\overline{Q}_{ST,\infty}^2} \right).$$

(c) *If* $1/2<\tau\leqslant 1$ *and* $\beta = \beta_0 + Bk^{1/2-\tau}$ *for a scalar constant* B,

$$\begin{pmatrix} (\widehat{S}_n'\widehat{S}_n - k)/k^{1/2} \\ \widehat{S}_n'\widehat{T}_n/k^{1/2} \\ (\widehat{T}_n'\widehat{T}_n - k)/k^{\tau} \end{pmatrix} \to_d \begin{pmatrix} \overline{Q}_{S,\infty} \\ \overline{Q}_{ST,\infty} \\ \overline{Q}_{T,\infty} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ \gamma_B r_\tau \\ d_{\beta_0}^2 r_\tau \end{pmatrix}, V_{3,\tau} \right),$$

$$(\widehat{AR}_n - 1)k^{1/2} \to_d \overline{Q}_{S,\infty} \sim \mathrm{N}(0,2),$$

$$\widehat{LM}_n \to_d \overline{Q}_{ST,\infty}^2 \sim \chi_1^2(\gamma_B^2 r_\tau^2) \quad \text{when } 1/2<\tau<1,$$

$$\widehat{LM}_n \to_d \overline{Q}_{ST,\infty}^2/(1 + d_{\beta_0}^2 r_1) \sim \chi_1^2(\gamma_B^2 r_\tau^2/(1 + d_{\beta_0}^2 r_1)) \quad \textit{when } \tau = 1,$$

$$\widehat{LR}_n = (1/(d_{\beta_0}^2 r_\tau))k^{1-\tau}\widehat{LM}_n(1 + o_p(1)) \quad \textit{when } 1/2<\tau<1, \ \textit{and}$$

$$\widehat{LR}_n = ((1 + d_{\beta_0}^2 r_1)/(d_{\beta_0}^2 r_1))\widehat{LM}_n + o_p(1) \quad \textit{when } \tau = 1.$$

(d) *If* $\tau \in (1,2]$, $r_\tau > 0$, *and* $\beta = \beta_0 + Bk^{-\tau/2}$,

$$\begin{pmatrix} (\widehat{S}_n'\widehat{S}_n - k)/k^{1/2} \\ \widehat{S}_n'\widehat{T}_n/k^{\tau/2} \\ (\widehat{T}_n'\widehat{T}_n - k)/k^{\tau} \end{pmatrix} \to_d \begin{pmatrix} \overline{Q}_{S,\infty} \\ \overline{Q}_{ST,\infty} \\ \overline{Q}_{T,\infty} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ \gamma_B r_\tau \\ d_{\beta_0}^2 r_\tau \end{pmatrix}, V_{3,\tau} \right),$$

$$(\widehat{AR}_n - 1)k^{1/2} \to_d \overline{Q}_{S,\infty} \sim \mathrm{N}(0,2),$$

$$\widehat{LM}_n \to_d \overline{Q}_{ST,\infty}^2/(d_{\beta_0}^2 r_\tau) \sim \chi_1^2(\gamma_B^2 r_\tau/d_{\beta_0}^2)) \textit{ provided } d_{\beta_0} \neq 0, \textit{ and}$$

$$\widehat{LR}_n = \widehat{LM}_n + \mathrm{o_p}(1).$$

Comments. (1) Theorem 1 shows that one obtains the same limit distribution when the errors are non-normal and Ω is estimated as when the errors are normal and Ω is known. As shown in Theorem 2 below, a consequence of this is that the critical value function for the CLR test given in (9) and χ_1^2 and $F_{k,n-k-p}$ critical values for the AR and LM tests, respectively, yield the correct asymptotic size under many weak IV asymptotics when the errors are not necessarily normally distributed. This holds in spite of the fact that the LR statistic has a non-degenerate asymptotic null distribution only after rescaling by a quantity $k^{1/2}$ or $k^{1-\tau}$ that is unknown to the practitioner.

(2) Given that the asymptotic distributions of the AR, LM, and LR statistics are the same under non-normal errors as under normal errors, the power comparisons of the three tests given in Andrews and Stock (2006) for the case of normal errors also applies to the case of non-normal errors. In particular, when $\tau < 1/2$, all three tests have trivial asymptotic power. (It is shown in Andrews and Stock (2006) for the case of normal errors that no test has non-trivial asymptotic power when $\tau < 1/2$.) In the most interesting case in which $\tau = 1/2$ and the whole range of possible fixed alternatives is considered, the CLR test is essentially uniformly more powerful asymptotically than the AR and LM tests, and the CLR test is essentially on the asymptotic power envelope for two-sided tests for the case of normal errors, see Andrews and Stock (2006).[2] When $\tau > 1/2$, the CLR and LM tests have equal asymptotic power against local alternatives and are on the asymptotic power envelope for two-sided tests for the case of normal errors (i.e., are asymptotically efficient under normality). In contrast, the AR test has trivial power against these alternatives.

(3) Note that the cases considered in Chao and Swanson (2005) and Han and Phillips (2006) correspond to $\tau > 1/2$. Those considered in Stock and Yogo (2005), Anderson et al. (2005), Hansen et al. (2005), and Newey and Windmeijer (2005) correspond to the case where $\tau = 1$.

(4) An interesting feature of Theorem 1 is that the statistics $\widehat{S}_n'\widehat{S}_n$, $\widehat{S}_n'\widehat{T}_n$, and $\widehat{T}_n'\widehat{T}_n$ are asymptotically independent.

The following Theorem shows that the CLR, LM, and AR tests have correct asymptotic size under many weak IV asymptotics and Assumptions 1–6.

Theorem 2. *Suppose Assumptions* 1–6 *hold. For any* $\tau \in (0,2]$, *under* $\mathrm{H}_0 : \beta = \beta_0$, (a) $\lim_{n\to\infty} P(\widehat{LR}_n > \kappa_{LR,\alpha}(\widehat{Q}_{T,n})) = \alpha$, (b) $\lim_{n\to\infty} P(\widehat{LM}_n > \chi_1^2(\alpha)) = \alpha$, *where* $\chi_1^2(\alpha)$ *is the* $1-\alpha$

[2] By "essentially," we mean that exhaustive simulations show that the asymptotic power of the CLR test is on, or very close to, the asymptotic power envelope.

quantile of the χ^2_1 *distribution, and* (c) $\lim_{n\to\infty} P(\widehat{AR}_n > F_{k,n-k-p}(\alpha)) = \alpha$, *where* $F_{k,n-k-p}(\alpha)$ *is the* $1-\alpha$ *quantile of the* $F_{k,n-k-p}$ *distribution.*

To conclude, the many weak IV asymptotic results given in this section show that the significance level of the AR, LM, and CLR tests are asymptotically correct no matter how weak are the IVs with normal or non-normal errors. On the other hand, these tests are not completely robust to many IVs. One cannot employ too many IVs relative to the sample size. For non-normal errors, the tests have correct asymptotic significance levels provided $k^3/n \to 0$ as $n \to \infty$ no matter how weak are the IVs. For normal errors, the results of Andrews and Stock (2006) show that the less restrictive condition $k^{3/2}/n \to 0$ as $n \to \infty$ suffices.

The many weak IV asymptotic results for parameter values in the alternative hypothesis show that the CLR test is more powerful asymptotically than the AR and LM tests for both normal and non-normal errors. The LM test, in turn, is more powerful asymptotically than the AR test.

The level and power results established under many IV asymptotics, combined with the properties of the CLR test under weak IV asymptotics, see Andrews et al. (2006a), lead us to recommend the CLR test (or heteroskedasticity and/or autocorrelation robust versions of it) for general use in scenarios where the IVs may be weak.

5. Appendix of proofs

In this Appendix, we prove the results of Section 4. For brevity, we only prove the results for the case $\tau \in (0,1)$. The results for $\tau \in (0,2]$ are proved in Andrews and Stock (2006) using similar methods. We start by stating several Lemmas, the purposes of which are discussed following Lemma 3 below. Lemma 3 is a CLT for multivariate degenerate U-statistics. The CLT is proved by using the Cramer–Wold device and verifying the conditions of Hall's (1984, Theorem 1) univariate CLT for degenerate U-statistics. Newey and Windmeijer (2005, Lemma A2) makes a similar use of Hall's result when establishing the asymptotic distribution of empirical likelihood estimators with many weak IVs.

Lemma 3. *Let* $\{(\xi_{ni}, \eta_{ni}) : i \leqslant n; n \geqslant 1\}$ *be a triangular array of random vectors that satisfies* (i) $\xi_{ni}, \eta_{ni} \in R^k$, *for all* $i \leqslant n$, *where* $k = k_n$; (ii) *for each* $n \geqslant 1$, (ξ_{ni}, η_{ni}) *are iid across* $i \leqslant n$; (iii) $\mathrm{E}\xi_{ni} = \mathrm{E}\eta_{ni} = 0$; (iv) $Var(\xi_{ni}) = I_k$, $Var(\eta_{ni}) = I_k$, *and* $Cov(\xi_{ni}, \eta_{ni}) = 0$; (v) $\sup_{\ell \leqslant k, n \geqslant 1} (\mathrm{E}\xi^4_{ni\ell} + \mathrm{E}\eta^4_{ni\ell}) < \infty$, *where* $\xi_{ni} = (\xi_{ni1}, \ldots, \xi_{nik})'$ *and* $\eta_{ni} = (\eta_{ni1}, \ldots, \eta_{nik})'$; (vi) $k \to \infty$ *as* $n \to \infty$; *and* (vii) $k^2/n \to 0$ *as* $n \to \infty$. *Then,*

(a)
$$\frac{1}{nk^{1/2}} \sum\sum_{1 \leqslant i < j \leqslant n} \begin{pmatrix} 2\xi'_{ni}\xi_{nj} \\ \xi'_{ni}\eta_{nj} + \xi'_{nj}\eta_{ni} \\ 2\eta'_{ni}\eta_{nj} \end{pmatrix} \to_d \mathrm{N}(0, V_3), \text{ where } V_3 = Diag\{2,1,2\},$$

(b)
$$\frac{1}{nk^{1/2}} \sum_{i=1}^{n} \begin{pmatrix} \xi'_{ni}\xi_{ni} - k \\ \xi'_{ni}\eta_{ni} \\ \eta'_{ni}\eta_{ni} - k \end{pmatrix} \to_p 0, \text{ and}$$

(c)
$$\frac{1}{k^{1/2}} vech\left(\frac{1}{n^{1/2}} \sum_{i=1}^{n} [\xi_{ni} : \eta_{ni}]' \frac{1}{n^{1/2}} \sum_{j=1}^{n} [\xi_{nj} : \eta_{nj}] - kI_2 \right) \to_d \mathrm{N}(0, V_3).$$

Comment. In Lemma 3 (and Lemma 4 below), Assumption (vii) can be relaxed if Assumption (v) is strengthened. We do not state such a result because a stronger condition than Assumption (vii) is needed anyway in Lemma 6 below.

We now summarize the purpose of Lemma 3 and the Lemmas that follow. The result of Theorem 1 concerns $[\widehat{S}_n : \widehat{T}_n]$. The $k \times 2$ matrix $[\widehat{S}_n : \widehat{T}_n]$ is roughly of the form $n^{-1/2}\sum_{i=1}^n [\xi_{ni} : \eta_{ni}]$, which appears in Lemma 3(c), with $\xi_{ni} = Z_i^* \cdot Y_i' b_0 (b_0' \Omega b_0)^{-1/2}$ and $\eta_{ni} = Z_i^* \cdot Y_i' a_0 (a_0' \Omega^{-1} a_0)^{-1/2}$. Since the means of these random vectors are not zero, Assumption (iii) of Lemma 3 does not hold. Hence, Lemma 3 is extended in Lemma 4 below to allow for non-zero means that are of a magnitude that corresponds to $\tau < 1$ in Theorem 1. Since the variance matrices of ξ_{ni} and η_{ni} as defined above are not I_k and are unknown, Assumption (iv) of Lemma 3 does not hold. Hence, Lemma 4 is extended in Lemma 5 below to allow for general variance matrices that are estimated. Next, $[\widehat{S}_n : \widehat{T}_n]$ are based on $Z_i = \widetilde{Z}_i - [n^{-1}\widetilde{Z}' X(n^{-1}X'X)^{-1}]X_i$, not $Z_i^* = \widetilde{Z}_i - [E\widetilde{Z}_i X_i'(EX_i X_i')^{-1}]X_i$, so Lemma 5 is extended in Lemma 6 to allow ξ_{ni} and η_{ni} to be linear combinations of iid random vectors, such as $\widetilde{Z}_i$ and X_i, with coefficient matrices for the linear combinations that converge in probability to constant matrices. Thus, Lemma 6 is needed when the model includes exogenous variables. All of the Lemmas mentioned above apply when the means of ξ_{ni} and η_{ni} are of a magnitude that corresponds to $\tau < 1$. Finally, Lemma 7 provides results on the asymptotic behavior of the sample matrices $\widetilde{Z}'\widetilde{Z}, X'X,\ X'\widetilde{Z}$, and $Z'Z$.

The proof of Theorem 2 is given at the end of this Appendix. It uses the results of Theorem 1, but does not (directly) use any of the lemmas.

Lemma 4. *Let $\{(\xi_{ni}, \eta_{ni}) : i \leqslant n; n \geqslant 1\}$ be a triangular array of random vectors that satisfies the assumptions of Lemma* 3, *but with Assumption* (iii) *replaced by* (iii)′ $\mathrm{E}\xi_{ni} = \mu_{n\xi}, \mathrm{E}\eta_{ni} = \mu_{n\eta}$, *and* $(\lambda_{n\xi} + \lambda_{n\eta})/k \to 0$, *where* $\lambda_{n\xi} = n\mu_{n\xi}'\mu_{n\xi}$, $\lambda_{n\eta} = n\mu_{n\eta}'\mu_{n\eta}$, *and* $\lambda_{n\xi\eta} = n\mu_{n\xi}'\mu_{n\eta}$. *Then,*

(a)
$$\frac{1}{nk^{1/2}} \sum\sum_{1 \leqslant i<j \leqslant n} \begin{pmatrix} 2\xi_{ni}'\xi_{nj} \\ \xi_{ni}'\eta_{nj} + \xi_{nj}'\eta_{ni} \\ 2\eta_{ni}'\eta_{nj} \end{pmatrix} - \begin{pmatrix} \lambda_{n\xi}/k^{1/2} \\ \lambda_{n\xi\eta}/k^{1/2} \\ \lambda_{n\eta}/k^{1/2} \end{pmatrix} \to_d \mathrm{N}(0, V_3),$$

(b)
$$\frac{1}{nk^{1/2}} \sum_{i=1}^n \begin{pmatrix} \xi_{ni}'\xi_{ni} - k \\ \xi_{ni}'\eta_{ni} \\ \eta_{ni}'\eta_{ni} - k \end{pmatrix} \to_p 0, \text{ and}$$

(c)
$$\frac{1}{k^{1/2}} vech\left(\frac{1}{n^{1/2}} \sum_{i=1}^n [\xi_{ni} : \eta_{ni}]' \frac{1}{n^{1/2}} \sum_{j=1}^n [\xi_{nj} : \eta_{nj}] - kI_2\right) - \begin{pmatrix} \lambda_{n\xi}/k^{1/2} \\ \lambda_{n\xi\eta}/k^{1/2} \\ \lambda_{n\eta}/k^{1/2} \end{pmatrix}$$
$$\to_d \mathrm{N}(0, V_3), \quad \text{where } V_3 = Diag\{2, 1, 2\}.$$

Let $\| \cdot \|$ denote the Euclidean norm of a vector or matrix.

Lemma 5. *Let* $\{(\xi_{ni},\eta_{ni}) : i \leqslant n; n \geqslant 1\}$ *be a triangular array of random vectors that satisfies the assumptions of Lemma* 3, *but with Assumption* (iv) *replaced by* (iv)$'$ $Var(\xi_{ni}) = \Sigma_{n\xi} \in R^{k\times k}$, $Var(\eta_{ni}) = \Sigma_{n\eta} \in R^{k\times k}$, $Cov(\xi_{ni},\eta_{ni}) = 0$, *and* $\widehat{\Sigma}_{n\xi}$ *and* $\widehat{\Sigma}_{n\eta}$ *are random* $k\times k$ *matrices that satisfy* $\|\widehat{\Sigma}_{n\xi} - \Sigma_{n\xi}\| = o_p(k^{-1/2})$ *and* $\|\widehat{\Sigma}_{n\eta} - \Sigma_{n\eta}\| = o_p(k^{-1/2})$, *with Assumption* (iii) *replaced by* (iii)$''$ $E\xi_{ni} = \mu_{n\xi}, E\eta_{ni} = \mu_{n\eta}$, *and* $(\lambda^*_{n\xi} + \lambda^*_{n\eta})/k \to 0$, *where* $\lambda^*_{n\xi} = n\mu'_{n\xi}\Sigma^{-1}_{n\xi}\mu_{n\xi}$, $\lambda^*_{n\eta} = n\mu'_{n\eta}\Sigma^{-1}_{n\eta}\mu_{n\eta}$, *and* $\lambda^*_{n\xi\eta} = n\mu'_{n\xi}\Sigma^{-1/2}_{n\xi}\Sigma^{-1/2}_{n\eta}\mu_{n\eta}$, *with Assumption* (vii) *replaced by* (vii)$'$ $k^3/n \to 0$, *and with the addition of Assumption* (viii) $\inf_{n\geqslant 1}\lambda_{\min}(\Sigma_{n\xi}) > 0$ *and* $\inf_{n\geqslant 1}\lambda_{\min}(\Sigma_{n\eta}) > 0$. *Then,*

$$\frac{1}{k^{1/2}}\, vech\left(\frac{1}{n^{1/2}}\sum_{i=1}^{n}[\widehat{\Sigma}^{-1/2}_{n\xi}\xi_{ni} : \widehat{\Sigma}^{-1/2}_{n\eta}\eta_{ni}]'\frac{1}{n^{1/2}}\sum_{j=1}^{n}[\widehat{\Sigma}^{-1/2}_{n\xi}\xi_{nj} : \widehat{\Sigma}^{-1/2}_{n\eta}\eta_{nj}] - kI_2\right) - \begin{pmatrix}\lambda^*_{n\xi}/k^{1/2}\\ \lambda^*_{n\xi\eta}/k^{1/2}\\ \lambda^*_{n\eta}/k^{1/2}\end{pmatrix} \to_d N(0,V_3), \quad \textit{where } V_3 = Diag\{2,1,2\}.$$

Lemma 6. *Suppose* (a) $(\xi_{ni},\eta_{ni},\widehat{\Sigma}_{n\xi},\widehat{\Sigma}_{n\eta})$ *satisfy the conditions of Lemma* 5 (b) $\xi_{ni} = \xi_{n1i} + D_{n\xi}\xi_{n2i}$ *and* $\eta_{ni} = \eta_{n1i} + D_{n\eta}\eta_{n2i}$, *where* $D_{n\xi}, D_{n\eta} \in R^{k\times k}$ *are non-random matrices*, (c) $\{(\xi_{n2i},\eta_{n2i}) : i \leqslant n\}$ *are iid across* $i \leqslant n$ *with* $\|E\xi_{n2i}\|^2 = O(k^2/n)$ *and* $\|E\eta_{n2i}\|^2 = O(k^2/n)$, (d) $\widehat{D}_{n\xi}, \widehat{D}_{n\eta} \in R^{k\times k}$ *are random matrices that satisfy* $\|\widehat{D}_{n\xi} - D_{n\xi}\| = o_p(k^{-1})$ *and* $\|\widehat{D}_{n\eta} - D_{n\eta}\| = o_p(k^{-1})$, *and* (e) $\widehat{\xi}_{ni} = \xi_{n1i} + \widehat{D}_{n\xi}\xi_{n2i}$ *and* $\widehat{\eta}_{ni} = \eta_{n1i} + \widehat{D}_{n\eta}\eta_{n2i}$. *Then, the result of Lemma* 5 *holds with* $(\widehat{\xi}_{ni},\widehat{\eta}_{ni})$ *in place of* (ξ_{ni},η_{ni}).

Lemma 7. *Suppose Assumptions* 1, 2, *and* 4 *hold, then* (a) $\|n^{-1}\widetilde{Z}'\widetilde{Z} - E\widetilde{Z}_i\widetilde{Z}'_i\| = o_p(k^{-1/2})$, (b) $\|(n^{-1}X'X)^{-1} - (EX_iX'_i)^{-1}\| = O_p(n^{-1/2})$, (c) $\|EX_i\widetilde{Z}'_i\| = O(k^{1/2})$, (d) $\|n^{-1}X'\widetilde{Z} - EX_i\widetilde{Z}'_i\| = o_p(k^{-1})$, *and* (e) $\|n^{-1}Z'Z - EZ^*_iZ^{*\prime}_i\| = o_p(k^{-1/2})$.

Proof of Theorem 1. As stated above, we only prove the results of the Theorem for the case where $\tau \in (0,1)$. The proof for $\tau \in (1,2]$ is given in Andrews and Stock (2006). First, we show that the results of the Theorem hold with $\widehat{S}_n$ and $\widehat{T}_n$ defined with the true Ω in place of $\widehat{\Omega}_n$. The statistics $(\widehat{S}_n,\widehat{T}_n)$ are invariant to the coefficient η on X_i. Hence, wlog we take $\eta = 0$. Let $b_* = b_0(b'_0\Omega b_0)^{-1/2}$ and $a_* = \Omega^{-1}a_0(a'_0\Omega^{-1}a_0)^{-1/2}$. We apply Lemma 6 with

$$\begin{aligned}
&\xi_{ni} = Z^*_i \cdot Y'_i b_*, \quad \xi_{n1i} = \widetilde{Z}_i \cdot Y'_i b_*, \quad \xi_{n2i} = (X'_i \cdot Y'_i b_*, 0'_{k-p})' \in R^k,\\
&\widehat{\Sigma}_{n\xi} = \widehat{\Sigma}_{n\eta} = n^{-1}Z'Z, \quad D_{n\xi} = D_{n\eta} = [E\widetilde{Z}_iX'_i(EX_iX'_i)^{-1} : 0_{k\times(k-p)}] \in R^{k\times k},\\
&\widehat{D}_{n\xi} = \widehat{D}_{n\eta} = [n^{-1}\widetilde{Z}'X(n^{-1}X'X)^{-1} : 0_{k\times(k-p)}] \in R^{k\times k},\\
&\eta_{ni} = Z^*_i \cdot Y'_i a_*, \quad \eta_{n1i} = \widetilde{Z}_i \cdot Y'_i a_*, \quad \text{and} \quad \eta_{n2i} = (X'_i \cdot Y'_i a_*, 0'_{k-p})' \in R^k.
\end{aligned} \tag{11}$$

Assumptions (b) and (e) of Lemma 6 follow immediately from (11).

Assumption (a) of Lemma 6 requires that Assumptions (i), (ii), (iii)$''$, (iv)$'$, and (v)–(viii) of Lemmas 3–5 hold. Assumptions (i), (ii), and (v)–(viii) hold immediately by Assumptions 1–4.

Assumption (iv)$'$ holds because $\|\widehat{\Sigma}_{n\xi} - \Sigma_{n\xi}\| = \|\widehat{\Sigma}_{n\eta} - \Sigma_{n\eta}\| = o_p(k^{-1/2})$ by Lemma 7(e), where

$$\begin{aligned}
\Sigma_{n\xi} &= Var(\xi_{ni}) = \mathrm{E}Z_i^*Z_i^{*'} \cdot \mathrm{E}(V_i'b_0)^2(b_0'\Omega b_0)^{-1} = \mathrm{E}Z_i^*Z_i^{*'},\\
\Sigma_{n\eta} &= Var(\eta_{ni}) = \mathrm{E}Z_i^*Z_i^{*'} \cdot \mathrm{E}(V_i'\Omega^{-1}a_0)^2(a_0'\Omega^{-1}a_0)^{-1} = \mathrm{E}Z_i^*Z_i^{*'},\\
Cov(\xi_{ni},\eta_{ni}) &= \mathrm{E}Z_i^*Z_i^{*'} \cdot \mathrm{E}(b_0'V_iV_i'\Omega^{-1}a_0)(b_0'\Omega b_0)^{-1/2}(a_0'\Omega^{-1}a_0)^{-1/2} = 0,
\end{aligned} \tag{12}$$

each equation uses Assumptions 1–3, and the last equality uses $b_0'a_0 = 0$.

Assumption (iii)$''$ holds using Assumption 5 because

$$\begin{aligned}
&\mu_{n\xi} = \mathrm{E}\xi_{ni} = \mathrm{E}Z_i^*Y_i'b_* = \mathrm{E}Z_i^*(\widetilde{Z}_i'\pi a' + X_i'\eta)b_* = \mathrm{E}Z_i^*Z_i^{*'}\pi a'b_* = \mathrm{E}Z_i^*Z_i^{*'}\pi c_\beta,\\
&\mu_{n\eta} = \mathrm{E}\eta_{ni} = \mathrm{E}Z_i^*Y_i'a_* = \mathrm{E}Z_i^*Z_i^{*'}\pi a'\Omega^{-1}a_* = \mathrm{E}Z_i^*Z_i^{*'}\pi d_\beta,\\
&(\lambda_{n\xi}^*, \lambda_{n\xi\eta}^*, \lambda_{n\eta}^*) = n\pi'\mathrm{E}Z_i^*Z_i^{*'}\pi \cdot (c_\beta^2, c_\beta d_\beta, d_\beta^2),\\
&(\lambda_{n\xi}^* + \lambda_{n\eta}^*)/k = (\lambda_{n,k}^*/k^\tau)k^{\tau-1}(c_\beta^2 + d_\beta^2) = \mathrm{O}(k^{\tau-1}) = o(1),
\end{aligned} \tag{13}$$

where $\lambda_{n,k}^*$ is defined in (4) and the last equality holds because $\tau < 1$.

To show Assumption (c) of Lemma 6, we write

$$\|\mathrm{E}\xi_{n2i}\|^2 = \|\mathrm{E}X_i \cdot Y_i'b_*\|^2 = \|\mathrm{E}X_i\widetilde{Z}_i'\pi\|^2(a'b_*)^2 \leqslant \|\mathrm{E}X_i\widetilde{Z}_i'\|^2 \cdot \|\pi\|^2(a'b_*)^2, \tag{14}$$

where the second equality uses the assumption above that the coefficient, η, on X_i is 0 wlog. Now, $\|\mathrm{E}X_i\widetilde{Z}_i\|^2 = \mathrm{O}(k)$ by Lemma 7(c). Also, Assumption 5 gives

$$\mathrm{O}(1) = \lambda_{n,k}^*/k^\tau = n\pi'\mathrm{E}Z_i^*Z_i^{*'}\pi/k^\tau \geqslant n\pi'\pi\lambda_{\min}(\mathrm{E}Z_i^*Z_i^{*'})/k^\tau. \tag{15}$$

This, Assumption 2, and $\tau < 1$ yield $\pi'\pi = \mathrm{O}(k^\tau/n) \leqslant \mathrm{O}(k/n)$. Combining these results gives $\|\mathrm{E}\xi_{n2i}\|^2 = \mathrm{O}(k^2/n)$, as desired. The same argument gives $\|\mathrm{E}\eta_{n2i}\|^2 = \mathrm{O}(k^2/n)$. Thus, Assumption (c) holds.

Assumption (d) of Lemma 6 holds because

$$\begin{aligned}
\|\widehat{D}_{n\xi} - D_{n\xi}\| &= \|n^{-1}\widetilde{Z}'X(n^{-1}X'X)^{-1} - \mathrm{E}\widetilde{Z}_iX_i'(n^{-1}X'X)^{-1}\\
&\quad + \mathrm{E}\widetilde{Z}_iX_i'(n^{-1}X'X)^{-1} - \mathrm{E}\widetilde{Z}_iX_i'(EX_iX_i')^{-1}\|\\
&\leqslant \|n^{-1}\widetilde{Z}'X - \mathrm{E}\widetilde{Z}_iX_i'\| \cdot \|(n^{-1}X'X)^{-1}\|\\
&\quad + \|\mathrm{E}\widetilde{Z}_iX_i'\| \cdot \|(n^{-1}X'X)^{-1} - (\mathrm{E}X_iX_i')^{-1}\|\\
&= o_p(k^{-1})\mathrm{O}_p(1) + \mathrm{O}(k^{1/2})\mathrm{O}_p(n^{-1/2}) = o_p(k^{-1}),
\end{aligned} \tag{16}$$

where the second equality holds by Lemma 7(b)–(d) and the fact that $(n^{-1}X'X)^{-1} = \mathrm{O}_p(1)$ by the WLLNs, Slutsky's Theorem, and $\mathrm{E}X_iX_i' > 0$ and the third equality uses Assumption 4.

The means of the asymptotic normal distributions given in Theorem 1(a)–(c) arise in the present case because, by (13) and Assumptions 5 and 6, we have

$$\begin{aligned}
(\lambda_{n\xi}^*/k^{1/2}, \lambda_{n\xi\eta}^*/k^{1/2}, \lambda_{n\eta}^*/k^{1/2})' &= (\lambda_{n,k}^*/k^{1/2}) \cdot (c_\beta^2, c_\beta d_\beta, d_\beta^2)'\\
&\to \begin{cases} r_{1/2}(c_\beta^2, c_\beta d_\beta, d_\beta^2)' & \text{when } \tau = 1/2\\ (0,0,0)' & \text{when } \tau < 1/2. \end{cases}
\end{aligned} \tag{17}$$

When $\tau \in (1/2, 1)$, (13) and Assumptions 5 and 6 lead to

$$c_\beta = Bk^{1/2-\tau}\overline{b}_0, \quad \overline{b}_0 = (b_0'\Omega b_0)^{-1/2}, \quad d_\beta = d_{\beta_0}(1 + o(1)),$$

and

$$\begin{aligned}(\lambda^*_{n\xi}/k^{1/2}, \lambda^*_{n\xi\eta}/k^{1/2}, \lambda^*_{n\eta}/k^\tau)' &= (c_\beta^2\lambda^*_{n,k}k^{-1/2}, c_\beta d_\beta\lambda^*_{n,k}k^{-1/2}, d_\beta^2\lambda^*_{n,k}k^{-\tau})' \\ &= (B^2\overline{b}_0^2k^{1-2\tau}\lambda^*_{n,k}k^{-1/2}, B\overline{b}_0 d_\beta k^{1/2-\tau}\lambda^*_{n,k}k^{-1/2}, d_\beta^2\lambda^*_{n,k}k^{-\tau})' \\ &\to (0, B\overline{b}_0 d_{\beta_0} r_\tau, d_{\beta_0}^2 r_\tau)' = (0, \gamma_B r_\tau, d_{\beta_0}^2 r_\tau)'. \end{aligned} \tag{18}$$

Hence, when $\tau \in (1/2, 1)$, Lemma 6 shows that $((\widehat{S}_n'\widehat{S}_n - k)/k^{1/2}, \widehat{S}_n'\widehat{T}_n/k^{1/2}) \to_d (\overline{Q}_{S,\infty}, \overline{Q}_{ST,\infty})$ and $(\widehat{T}_n'\widehat{T}_n - k)/k^\tau - d_\beta^2\lambda^*_{n,k}/k^\tau = o_p(1)$ because $\tau > 1/2$. The latter, combined with $d_\beta^2\lambda^*_{n,k}/k^\tau \to d_{\beta_0}^2 r_\tau$, gives the desired result that $(\widehat{T}_n'\widehat{T}_n - k)/k^\tau \to_p d_{\beta_0}^2 r_\tau$ when $\tau \in (1/2, 1)$. Here and below, the stated results for $\widehat{AR}_n$, $\widehat{LM}_n$, and $\widehat{LR}_n$ hold given those for $\widehat{S}_n'\widehat{S}_n$, $\widehat{S}_n'\widehat{T}_n$, and $\widehat{T}_n'\widehat{T}_n$ by the same argument as in (11.10)–(11.14) of the Proof of Theorem 1 of Andrews and Stock (2006).

To complete the proof for $\tau < 1$, we extend the results to the case where $(\widehat{S}_n, \widehat{T}_n)$ are defined with $\widehat{\Omega}_n$, not Ω. This extension holds by the result of Lemma 1 of Andrews and Stock (2006) that $k^{1/2}(\widehat{\Omega}_n - \Omega) = o_p(1)$ (which holds under Assumptions 1–3) and the proof of Theorem 4 of Andrews and Stock (2006). □

Proof of Lemma 3. To prove part (a), by the Cramer–Wold device, it suffices to show that for any $\alpha = (\alpha_1, \alpha_2, \alpha_3)' \in R^3$ with $\alpha \neq 0$,

$$\alpha' U_n = \sum\sum_{1 \leqslant i < j \leqslant n} S_{nij} = \sum_{j=2}^n M_{nj} \to_d \mathrm{N}(0, \alpha' V_3 \alpha), \quad \text{where } M_{nj} = \sum_{i=1}^{j-1} S_{nij}$$

and

$$S_{nij} = \frac{1}{nk^{1/2}}(2\alpha_1\xi_{ni}'\xi_{nj} + \alpha_2(\xi_{ni}'\eta_{nj} + \xi_{nj}'\eta_{ni}) + 2\alpha_3\eta_{ni}'\eta_{nj}). \tag{19}$$

We establish this result using Hall's (1984, Theorem 1) univariate CLT for degenerate U-statistics. Hall's CLT is established by writing the U-statistic as a martingale (with martingale differences $\{M_{nj} : j \geqslant 1\}$) and applying Brown's (1971) martingale CLT.

We apply Hall's Theorem 1 with his $X_{ni} = (\xi_{ni}', \eta_{ni}')'$ and his $H_n(x, y)$ equal to

$$H_n(x, x_*) = \sum_{s=1}^3 \alpha_s H_{sn}(x, x_*), \quad \text{where } x = (\xi', \eta')' \in R^{2k}, \;\; x_* = (\xi_*', \eta_*')' \in R^{2k},$$

$$H_{1n}(x, x_*) = 2n^{-1}k^{-1/2}\xi'\xi_*, \quad H_{2n}(x, x_*) = n^{-1}k^{-1/2}(\xi'\eta_* + \xi_*'\eta), \quad \text{and}$$

$$H_{3n}(x, x_*) = 2n^{-1}k^{-1/2}\eta'\eta_*. \tag{20}$$

Note that $\mathrm{E}(H_n(X_{n1}, X_{n2})|X_{n2}) = 0$ a.s. because X_{n1} and X_{n2} are independent with mean zero. In consequence, the U-statistic $\alpha' U_n$ in (19) is degenerate. Hall's Theorem 1 states that

$$\begin{aligned}
&\alpha' U_n/(n^2 \mathrm{E} H_n^2(X_{n1}, X_{n2})/2)^{1/2} \rightarrow_d \mathrm{N}(0,1) \text{ provided}\\
&\text{(I) } n^{-1}\mathrm{E}H_n^4(X_{n1}, X_{n2})/(\mathrm{E}H_n^2(X_{n1}, X_{n2}))^2 \to 0 \text{ and}\\
&\text{(II) } \mathrm{E}G_n^2(X_{n1}, X_{n2})/(\mathrm{E}H_n^2(X_{n1}, X_{n2}))^2 \to 0, \quad \text{where}\\
&G_n(x, x_*) = \mathrm{E}H_n(X_{n1}, x)H_n(X_{n1}, x_*) \quad \text{for } x, x_* \in R^{2k}. \qquad (21)
\end{aligned}$$

Conditions (I) and (II) suffice for the Lindeberg condition and the conditional variance condition, respectively, required in Brown's martingale CLT. We verify (I) and (II) for $H_n(x, x_*)$ defined in (20). First, we have

$$n^2 \mathrm{E} H_n^2(X_{n1}, X_{n2})/2 = \frac{1}{2k}\alpha' \mathrm{E}\begin{pmatrix} 2\xi_{n1}'\xi_{n2} \\ \xi_{n1}'\eta_{n2} + \xi_{n2}'\eta_{n1} \\ 2\eta_{n1}'\eta_{n2}\end{pmatrix}\begin{pmatrix} 2\xi_{n1}'\xi_{n2} \\ \xi_{n1}'\eta_{n2} + \xi_{n2}'\eta_{n1} \\ 2\eta_{n1}'\eta_{n2}\end{pmatrix}'\alpha, \qquad (22)$$

where $X_{ni} = (\xi_{ni}', \eta_{ni}')'$. Next, we have

$$\mathrm{E}(2\xi_{n1}'\xi_{n2})^2 = 4\,tr(\mathrm{E}\xi_{n2}\xi_{n2}'\xi_{n1}\xi_{n1}') = 4\,tr(\mathrm{E}\xi_{n2}\xi_{n2}' \cdot \mathrm{E}\xi_{n1}\xi_{n1}') = 4k,$$

$$\begin{aligned}
\mathrm{E}(\xi_{n1}'\eta_{n2} + \xi_{n2}'\eta_{n1})^2 &= \mathrm{E}(\xi_{n1}'\eta_{n2})^2 + 2\mathrm{E}\xi_{n1}'\eta_{n2}\xi_{n2}'\eta_{n1} + \mathrm{E}(\xi_{n2}'\eta_{n1})^2\\
&= tr(\mathrm{E}\xi_{n1}\xi_{n1}' \cdot \mathrm{E}\eta_{n2}\eta_{n2}') + 2\,tr(\mathrm{E}\eta_{n2}\xi_{n2}' \cdot \mathrm{E}\eta_{n1}\xi_{n1}')\\
&\quad + tr(\mathrm{E}\xi_{n2}\xi_{n2}' \cdot \mathrm{E}\eta_{n1}\eta_{n1}') = 2k,\\
\mathrm{E}(2\xi_{n1}'\xi_{n2})(\xi_{n1}'\eta_{n2} + \xi_{n2}'\eta_{n1}) &= 2\mathrm{E}\xi_{n1}'\xi_{n2}\xi_{n1}'\eta_{n2} + 2\mathrm{E}\xi_{n1}'\xi_{n2}\xi_{n2}'\eta_{n1}\\
&= 2\,tr(\mathrm{E}\xi_{n1}\xi_{n1}' \cdot \mathrm{E}\xi_{n2}\eta_{n2}') + 2\,tr(\mathrm{E}\xi_{n2}\xi_{n2}' \cdot \mathrm{E}\eta_{n1}\xi_{n1}') = 0,
\end{aligned}$$

and

$$\mathrm{E}(2\xi_{n1}'\xi_{n2})(2\eta_{n1}'\eta_{n2}) = 4\,tr(\mathrm{E}\xi_{n1}\eta_{n1}'\eta_{n2}\xi_{n2}') = 4\,tr(\mathrm{E}\xi_{n1}\eta_{n1}' \cdot \mathrm{E}\eta_{n2}\xi_{n2}') = 0, \qquad (23)$$

using Assumptions (ii) and (iv) of the Lemma. Likewise, we have $\mathrm{E}(2\eta_{n1}'\eta_{n2})^2 = 4k$ and $\mathrm{E}(2\eta_{n1}'\eta_{n2})(\xi_{n1}'\eta_{n2} + \xi_{n2}'\eta_{n1}) = 0$. Combining these results with (22) and (23) implies that

$$n^2 \mathrm{E} H_n^2(X_{n1}, X_{n2})/2 = \alpha' V_3 \alpha > 0 \quad \text{for all } n, \qquad (24)$$

which yields the asymptotic variance given in (19).

Now, to verify condition (I) of (21), we have

$$\begin{aligned}
\mathrm{E}H_{1n}^4(X_{n1}, X_{n2}) &= \frac{16}{n^4 k^2}\,\mathrm{E}(\xi_{n1}'\xi_{n2})^4 = \frac{16}{n^4k^2}\mathrm{E}\left(\sum_{\ell=1}^k \xi_{n1\ell}\xi_{n2\ell}\right)^4\\
&= \frac{16}{n^4k^2}\sum_{\ell_1=1}^k\sum_{\ell_2=1}^k\sum_{\ell_3=1}^k\sum_{\ell_4=1}^k \mathrm{E}\xi_{n1\ell_1}\xi_{n2\ell_1}\xi_{n1\ell_2}\xi_{n2\ell_2}\xi_{n1\ell_3}\xi_{n2\ell_3}\xi_{n1\ell_4}\xi_{n2\ell_4}\\
&\leqslant \frac{16k^2}{n^4}\sup_{\ell_1,\ell_2,\ell_3,\ell_4 \leqslant k, n\geqslant 1} \mathrm{E}\xi_{n1\ell_1}\xi_{n1\ell_2}\xi_{n1\ell_3}\xi_{n1\ell_4} \cdot \mathrm{E}\xi_{n2\ell_1}\xi_{n2\ell_2}\xi_{n2\ell_3}\xi_{n2\ell_4}\\
&= \mathrm{O}\left(\frac{k^2}{n^4}\right), \qquad (25)
\end{aligned}$$

where $\xi_{ni} = (\xi_{ni1}, \ldots, \xi_{nik})'$ and the last equality holds by Assumption (v) of the Lemma and the Cauchy–Schwartz inequality. Similar calculations and the use of Minkowski's inequality yields $\mathrm{E}H_{sn}^4(X_{n1},X_{n2})=\mathrm{O}(k^2/n^4)$ for $s=2,3$. These results and Minkowski's inequality then give $\mathrm{E}H_n^4(X_{n1},X_{n2})=\mathrm{O}(k^2/n^4)$. Combining this with (24) establishes condition (I) of (21) provided $n^{-1}k^2\to 0$, which holds by Assumption (vii) of the Lemma.

To verify condition (II) of (21), by the Cauchy–Schwartz inequality, it suffices to verify condition (II) with $\mathrm{E}G_n^2(X_{ni},X_{nj})$ replaced by $\mathrm{E}G_{sn}^2(X_{ni},X_{nj})$ for $s=1,2,3$, where $G_{sn}(\cdot,\cdot)$ is defined as $G_n(\cdot,\cdot)$ is defined in (21), but with $H_{sn}(\cdot,\cdot)$ in place of $H_n(\cdot,\cdot)$. We have

$$G_{1n}(x,x_*)=\mathrm{E}H_{1n}(X_{n1},x)H_{1n}(X_{n1},x_*)=\frac{4}{n^2k}\,\mathrm{E}\xi_{n1}'\xi\xi_{n1}'\xi_*=\frac{4}{n^2k}\,\xi'\xi_*,$$

$$\begin{aligned}G_{2n}(x,x_*)&=\mathrm{E}H_{2n}(X_{n1},x)H_{2n}(X_{n1},x_*)\\&=\frac{1}{n^2k}\,\mathrm{E}(\xi_{n1}'\eta+\xi'\eta_{n1})(\xi_{n1}'\eta_*+\xi_*'\eta_{n1})=\frac{1}{n^2k}(\xi'\xi_*+\eta'\eta_*),\end{aligned}\tag{26}$$

where $x=(\xi,\eta)'$ and $x_*=(\xi_*,\eta_*)'$. Hence,

$$\begin{aligned}\mathrm{E}G_{1n}^2(X_{n1},X_{n2})&=\frac{16}{n^4k^2}\,\mathrm{E}(\xi_{n1}'\xi_{n2})^2=\frac{16}{n^4k^2}\,tr(\mathrm{E}\xi_{n1}\xi_{n1}'\cdot\mathrm{E}\xi_{n2}\xi_{n2}')=\frac{16}{n^4k},\\ \mathrm{E}G_{2n}^2(X_{n1},X_{n2})&=\frac{1}{n^4k^2}\mathrm{E}(\xi_{n1}'\xi_{n2}+\eta_{n1}'\eta_{n2})^2=\frac{2}{n^4k}.\end{aligned}\tag{27}$$

Similarly, $\mathrm{E}G_{3n}^2(X_{n1},X_{n2})=16/(n^4k)$. Combining (24), (26), and (27) yields condition (II) of (21) provided $k\to\infty$, which holds by Assumption (vi) of the Lemma.

Part (b) of the Lemma holds because the left-hand side in part (b) has mean zero and variance that is $o(1)$. The latter holds because

$$\begin{aligned}&\mathrm{E}(\xi_{n1}'\xi_{n1})^2=\sum_{\ell_1=1}^{k}\sum_{\ell_2=1}^{k}\mathrm{E}\xi_{n1\ell_1}^2\xi_{n1\ell_2}^2\leqslant k^2\sup_{\ell\leqslant k,n\geqslant 1}\mathrm{E}\xi_{n1\ell}^4=\mathrm{O}(k^2),\\ &Var(n^{-1}k^{-1/2}\sum_{i=1}^{n}(\xi_{ni}'\xi_{ni}-k))=n^{-1}k^{-1}\,Var((\xi_{ni}'\xi_{ni}-k))=n^{-1}\mathrm{O}(k)=\mathrm{o}(1)\end{aligned}\tag{28}$$

using Assumptions (v) and (vii) of the Lemma. Similarly, $\mathrm{E}(\xi_{ni}'\eta_{ni})^2=\mathrm{O}(k^2)$ and $\mathrm{E}(\eta_{n1}'\eta_{n1})^2=\mathrm{O}(k^2)$ yield $Var(n^{-1}k^{-1/2}\sum_{i=1}^n\xi_{ni}'\eta_{ni})=\mathrm{o}(1)$ and $Var(n^{-1}k^{-1/2}\sum_{i=1}^n(\eta_{ni}'\eta_{ni}-k)=\mathrm{o}(1)$.

Part (c) follows from parts (a) and (b) because the lhs of part (c) equals the sum of the lhs of parts (a) and (b). $\square$

Proof of Lemma 4. To prove part (a), we write

$$\begin{aligned}&\frac{2}{nk^{1/2}}\sum\sum_{1\leqslant i<j\leqslant n}\xi_{ni}'\xi_{nj}=A_{1n}+A_{2n}+A_{3n},\ \text{where}\\ &A_{1n}=\frac{2}{nk^{1/2}}\sum\sum_{1\leqslant i<j\leqslant n}(\xi_{ni}-\mu_{n\xi})'(\xi_{nj}-\mu_{n\xi}),\\ &A_{2n}=\frac{2}{nk^{1/2}}\sum\sum_{1\leqslant i<j\leqslant n}[\mu_{n\xi}'(\xi_{ni}-\mu_{n\xi})+\mu_{n\xi}'(\xi_{nj}-\mu_{n\xi})],\ \text{and}\\ &A_{3n}=\frac{2}{nk^{1/2}}\sum\sum_{1\leqslant i<j\leqslant n}\mu_{n\xi}'\mu_{n\xi}.\end{aligned}\tag{29}$$

Now, some calculations yield

$$A_{3n} = \frac{n(n-1)}{nk^{1/2}}\mu'_{n\xi}\mu_{n\xi} = \frac{1}{k^{1/2}}\lambda_{n\xi} - \frac{1}{nk^{1/2}}\lambda_{n\xi} = \frac{1}{k^{1/2}}\lambda_{n\xi} + o(1),$$
$$A_{2n} = \frac{2(n-1)}{nk^{1/2}}\mu'_{n\xi}\sum_{i=1}^{n}(\xi_{ni} - \mu_{n\xi}), \quad EA_{2n} = 0,$$

and

$$Var(A_{2n}) = \frac{4(n-1)^2}{nk}\,\mu'_{n\xi}\,Var(\xi_{ni})\mu_{n\xi} = \frac{4(n-1)^2}{n^2k}\lambda_{n\xi} = o(1), \tag{30}$$

using $\lambda_{n\xi}/k \to 0$ and $k^2/n \to 0$. Combining (29) and (30) gives

$$\frac{2}{nk^{1/2}}\sum\sum_{1\leqslant i<j\leqslant n}\xi'_{ni}\xi_{nj} - \frac{1}{k^{1/2}}\lambda_{n\xi} = A_{1n} + o_p(1). \tag{31}$$

Similar calculations yield

$$\frac{1}{nk^{1/2}}\sum\sum_{1\leqslant i<j\leqslant n}(\xi'_{ni}\eta_{nj} + \xi'_{nj}\eta_{ni}) - \frac{1}{k^{1/2}}\,\lambda_{n\xi\eta}$$
$$= \frac{1}{nk^{1/2}}\sum\sum_{1\leqslant i<j\leqslant n}[(\xi_{ni} - \mu_{n\xi})'(\eta_{nj} - \mu_{n\eta}) + (\xi_{nj} - \mu_{n\xi})'(\eta_{ni} - \mu_{n\eta})] + o_p(1)$$

and

$$\frac{2}{nk^{1/2}}\sum\sum_{1\leqslant i<j\leqslant n}\eta'_{ni}\eta_{nj} - \frac{1}{k^{1/2}}\lambda_{n\eta} = \frac{1}{nk^{1/2}}\sum\sum_{1\leqslant i<j\leqslant n}(\eta_{ni} - \mu_{n\eta})'(\eta_{nj} - \mu_{n\eta}) + o_p(1). \tag{32}$$

Stacking the results of (31) and (32) and applying Lemma 3(a) to the rhs of these stacked equations yields convergence in distribution to $N(0, V_3)$, which is the result of part (a).

To show part (b), we write

$$\frac{1}{nk^{1/2}}\sum_{i=1}^{n}(\xi'_{ni}\xi_{ni} - k) = F_{1n} + F_{2n} + F_{3n}, \text{ where}$$
$$F_{1n} = \frac{1}{nk^{1/2}}\sum_{i=1}^{n}[(\xi_{ni} - \mu_{n\xi})'(\xi_{ni} - \mu_{n\xi}) - k], \quad F_{2n} = \frac{2}{nk^{1/2}}\sum_{i=1}^{n}\mu'_{n\xi}(\xi_{ni} - \mu_{n\xi}),$$
$$F_{3n} = \frac{1}{nk^{1/2}}\sum_{i=1}^{n}\mu'_{n\xi}\mu_{n\xi}. \tag{33}$$

We have $F_{1n} \to_p 0$ by Lemma 3(b). In addition,

$$F_{3n} = \frac{1}{nk^{1/2}}\lambda_{n\xi} = \frac{k^{1/2}}{n}\frac{\lambda_{n\xi}}{k} \to 0, \quad EF_{2n} = 0, \quad \text{and}$$
$$Var(F_{2n}) = \frac{4}{nk}\mu'_{n\xi}Var(\xi_{ni})\mu_{n\xi} = \frac{4}{n^2k}\lambda_{n\xi} \to 0 \tag{34}$$

using Assumptions (vii) and (iii)$'$. These results combine to show that $F_{1n} + F_{2n} + F_{3n}$ is $o_p(1)$. Similar calculations show that $n^{-1}k^{-1/2}\sum_{i=1}^{n}(\eta'_{ni}\eta_{ni} - k) = o_p(1)$ and $n^{-1}k^{-1/2}\sum_{i=1}^{n}\xi'_{ni}\eta_{ni} = o_p(1)$, which completes the proof of part (b).

Part (c) follows from parts (a) and (b) because the lhs of part (c) equals the sum of the lhs of parts (a) and (b). $\square$

Proof of Lemma 5. Lemma 4(c) with (ξ_{ni}, η_{ni}) of that Lemma set equal to $(\Sigma_{n\xi}^{-1/2}\xi_{ni}, \Sigma_{n\eta}^{-1/2}\eta_{ni})$ of the present Lemma gives the desired result but with $(\Sigma_{n\xi}, \Sigma_{n\eta})$ in place of $(\widehat{\Sigma}_{n\xi}, \widehat{\Sigma}_{n\eta})$. Hence, it suffices to show

$$\delta_n = A_n'(\Sigma_{n\xi}^{-1} - \widehat{\Sigma}_{n\xi}^{-1})A_n = \mathrm{o_p}(k^{1/2}), \quad \text{where } A_n = n^{-1/2}\sum_{i=1}^{n}\xi_{ni}, \tag{35}$$

and likewise with $(\xi_{ni}, \Sigma_{n\xi})$ replaced by $(\eta_{ni}, \Sigma_{n\eta})$.

Lemma 4(c) applied to $(\Sigma_{n\xi}^{-1/2}\xi_{ni}, \Sigma_{n\eta}^{-1/2}\eta_{ni})$ also gives

$$A_n'\Sigma_{n\xi}^{-1}A_n = \mathrm{O_p}(k) \tag{36}$$

(due to the centering at kI_2). In addition, we have

$$\lambda_{\min}^{-1}(\widehat{\Sigma}_{n\xi}) = \mathrm{O_p}(1) \tag{37}$$

because $|\lambda_{\min}(\widehat{\Sigma}_{n\xi}) - \lambda_{\min}(\Sigma_{n\xi})| \leqslant \|\widehat{\Sigma}_{n\xi} - \Sigma_{n\xi}\| = \mathrm{o_p}(1)$ by Assumption (iv)$'$ and $\lambda_{\min}^{-1}(\Sigma_{n\xi}) = \mathrm{O}(1)$ by Assumption (viii).

The following are standard or hold by algebra: if H is a symmetric psd $k \times k$ matrix, G is a $k \times k$ matrix, and c is a k-vector, then (a) $\|HGH\| \leqslant \lambda_{\max}^2(H)\|G\|$, (b) $\|Hc\| \leqslant \lambda_{\max}(H)\|c\| \leqslant \|H\| \cdot \|c\|$, (c) $c'Gc \leqslant \|G\| \cdot \|c\|^2$, and (d) $I_k - H^{-1} = H - I_k - (H - I_k)'\, H^{-1}(H - I_k)$.

Let $C_n = \Sigma_{n\xi}^{1/2}$ and $D_n = \widehat{\Sigma}_{n\xi}^{1/2}$. Then, we have

$$\begin{aligned}
\delta_n &= A_n'(C_n^{-2} - D_n^{-2})A_n \\
&= A_n'C_n^{-1}(I_k - C_nD_n^{-2}C_n)C_n^{-1}A_n \\
&= A_n'C_n^{-1}(C_n^{-1}D_n^2C_n^{-1} - I_k)C_n^{-1}A_n \\
&\quad - A_n'C_n^{-1}(C_n^{-1}D_n^2C_n^{-1} - I_k)C_nD_n^{-2}C_n(C_n^{-1}D_n^2C_n^{-1} - I_k)C_n^{-1}A_n \\
&\leqslant A_n'C_n^{-1}[C_n^{-1}(D_n^2 - C_n^2)C_n^{-1}]C_n^{-1}A_n \\
&\quad + \|C_n(C_n^{-1}D_n^2C_n^{-1} - I_k)C_n^{-1}A_n\|^2 \cdot \lambda_{\max}^2(D_n^{-1}) \\
&\leqslant \|C_n^{-1}(D_n^2 - C_n^2)C_n^{-1}\| \cdot \|C_n^{-1}A_n\|^2 \\
&\quad + \|(D_n^2 - C_n^2)C_n^{-1}C_n^{-1}A_n\|^2 \cdot \lambda_{\min}^{-2}(D_n) \\
&\leqslant \|D_n^2 - C_n^2\| \cdot \lambda_{\max}^2(C_n^{-1}) \cdot \|C_n^{-1}A_n\|^2 \\
&\quad + \|D_n^2 - C_n^2\|^2\|C_n^{-1}C_n^{-1}A_n\|^2 \cdot \lambda_{\min}^{-2}(D_n) \\
&\leqslant \|D_n^2 - C_n^2\| \cdot \lambda_{\min}^{-1}(C_n^2) \cdot \|C_n^{-1}A_n\|^2 \\
&\quad + \|D_n^2 - C_n^2\|^2 \cdot \|C_n^{-1}A_n\|^2 \cdot \lambda_{\max}^2(C_n^{-1}) \cdot \lambda_{\min}^{-2}(D_n) \\
&= \mathrm{o_p}(k^{-1/2})\mathrm{O}(1)\mathrm{O_p}(k) + \mathrm{o_p}(k^{-1})\mathrm{O_p}(k)\mathrm{O}(1)\mathrm{O_p}(1) \\
&= \mathrm{o_p}(k^{1/2}),
\end{aligned} \tag{38}$$

where the third equality uses (d) with $H = C_n^{-1}D_n^2C_n^{-1}$, the first inequality holds by the triangle inequality and (b), the second inequality holds by (c), the third inequality holds by (a) and (b), the fourth inequality holds by (b), and the second last equality holds by Assumptions (iv)$'$ and (viii), (36), and (37). This establishes (35).

The same argument holds with $(\xi_{ni}, \widehat{\Sigma}_{n\xi})$ replaced by $(\eta_{ni}, \widehat{\Sigma}_{n\eta})$. Hence, (35) holds and the Lemma is proved. $\square$

Proof of Lemma 6. It suffices to show $\varDelta_n = \mathrm{o_p}(k^{1/2})$ and an analogous result with (ξ_{n1i}, ξ_{n2i}) replaced by (η_{n1i}, η_{n2i}), where $\varDelta_n$ is defined by

$$G_n = n^{-1/2}\sum_{i=1}^{n} \xi_{n1i}, \quad H_n = n^{-1/2}\sum_{i=1}^{n} \xi_{n2i},$$

$$\begin{aligned}\varDelta_n &= |(G_n + \widehat{D}_{n\xi}H_n)'\widehat{\Sigma}_{n\xi}^{-1}(G_n + \widehat{D}_{n\xi}H_n) - (G_n + D_{n\xi}H_n)'\widehat{\Sigma}_{n\xi}^{-1}(G_n + D_{n\xi}H_n)| \\ &= |H_n'(\widehat{D}_{n\xi} - D_{n\xi})'\widehat{\Sigma}_{n\xi}^{-1}(\widehat{D}_{n\xi} - D_{n\xi})H_n + 2H_n'(\widehat{D}_{n\xi} - D_{n\xi})'\widehat{\Sigma}_{n\xi}^{-1}(G_n + D_{n\xi}H_n)| \\ &\leqslant P_{1n} + 2P_{1n}^{1/2}P_{2n}^{1/2},\end{aligned}$$

$$P_{1n} = \|\widehat{\Sigma}_{n\xi}^{-1/2}(\widehat{D}_{n\xi} - D_{n\xi})H_n\|^2, \quad P_{2n} = (G_n + D_{n\xi}H_n)'\widehat{\Sigma}_{n\xi}^{-1}(G_n + D_{n\xi}H_n), \tag{39}$$

and the inequality holds by the Cauchy–Schwarz inequality. We have $P_{2n} = \mathrm{O_p}(k)$ by Lemma 5. Hence, the Lemma holds if $P_{1n} = \mathrm{o_p}(1)$.

We have

$$P_{1n} \leqslant \lambda_{\max}^2(\widehat{\Sigma}_{n\xi}^{-1/2}) \cdot \|(\widehat{D}_{n\xi} - D_{n\xi})H_n\|^2 \leqslant \lambda_{\min}^{-1}(\widehat{\Sigma}_{n\xi}) \cdot \|\widehat{D}_{n\xi} - D_{n\xi}\|^2\|H_n\|^2, \tag{40}$$

where the two inequalities hold by inequality (b) stated following (37) above.

Next, we have: (I) $\|H_n\|^2 = \mathrm{O_p}(k^2)$ because $\|H_n\| \leqslant \|H_n - \mathrm{E}H_n\| + \|\mathrm{E}H_n\|$, $\mathrm{E}\|H_n - \mathrm{E}H_n\|^2 = \mathrm{E}(\xi_{n2i} - \mathrm{E}\xi_{n2i})'(\xi_{n2i} - \mathrm{E}\xi_{n2i}) = \mathrm{O}(k)$, which implies that $\|H_n - \mathrm{E}H_n\|^2 = \mathrm{O_p}(k)$, and $\|\mathrm{E}H_n\|^2 = \|n^{1/2}\mathrm{E}\xi_{n2i}\|^2 = \mathrm{O}(k^2)$ by Assumption (c) of the Lemma, (II) $\lambda_{\max}(\widehat{\Sigma}_{n\xi}^{-1/2}) = \lambda_{\min}^{-1/2}(\widehat{\Sigma}_{n\xi}) = \mathrm{O_p}(1)$ by (37) above, and (III) $\|\widehat{D}_{n\xi} - D_{n\xi}\| = \mathrm{o_p}(k^{-1})$ by Assumption (d) of the Lemma. Hence, $P_{1n} = \mathrm{o_p}(1)$ and $\varDelta_n = \mathrm{o_p}(k^{1/2})$.

An analogous result holds with (η_{n1i}, η_{n2i}) in place of (ξ_{n1i}, ξ_{n2i}), which completes the proof. $\square$

Proof of Lemma 7. Part (a) holds because for all $\varepsilon > 0$

$$\begin{aligned}&P(k\|n^{-1}\widetilde{Z}'\widetilde{Z} - \mathrm{E}\widetilde{Z}_i\widetilde{Z}_i'\|^2 > \varepsilon) \\ &\quad\leqslant k\mathrm{E}\, tr\left(\left(n^{-1}\sum_{i=1}^{n}\widetilde{Z}_i\widetilde{Z}_i' - \mathrm{E}\widetilde{Z}_1\widetilde{Z}_1'\right)\left(n^{-1}\sum_{j=1}^{n}\widetilde{Z}_j\widetilde{Z}_j' - \mathrm{E}\widetilde{Z}_1\widetilde{Z}_1'\right)\right)\Big/\varepsilon \\ &\quad= k \cdot tr(n^{-1}\mathrm{E}(\widetilde{Z}_2\widetilde{Z}_2' - \mathrm{E}\widetilde{Z}_1\widetilde{Z}_1')(\widetilde{Z}_2\widetilde{Z}_2' - \mathrm{E}\widetilde{Z}_1\widetilde{Z}_1'))/\varepsilon \\ &\quad= kn^{-1}(\mathrm{E}(\widetilde{Z}_2'\widetilde{Z}_2)^2 - 2\mathrm{E}(\widetilde{Z}_2'\widetilde{Z}_1)^2 + tr([\mathrm{E}\widetilde{Z}_1\widetilde{Z}_1']\mathrm{E}\widetilde{Z}_1\widetilde{Z}_1'))/\varepsilon \\ &\quad\leqslant \mathrm{O}(k^3/n) = \mathrm{o}(1),\end{aligned} \tag{41}$$

where the first inequality holds by Markov's inequality, the first equality holds because the expectation of terms with $i \neq j$ is zero by independence, the second equality holds by algebra, the second inequality holds because $\sup_{j\leqslant k;n\geqslant 1}\mathrm{E}\widetilde{Z}_{ij}^4 < \infty$ by Assumption 2, and the third equality holds by Assumption 4.

Part (b) holds by the CLT and the delta method because $\mathrm{E}\|X_i\|^4 < \infty$, $\mathrm{E}X_iX_i'$ is pd, and the dimension p of X_i is fixed for all n.

Part (c) holds because $\|\mathrm{E}X_i\widetilde{Z}_i'\| \leqslant k^{1/2}p^{1/2}\sup_{j\leqslant k, n\geqslant 1}(\mathrm{E}\|X_i\widetilde{Z}_{ij}\|^2)^{1/2} = \mathrm{O_p}(k^{1/2})$ using the fact that p is fixed for all n.

Part (d) is established as follows. By Markov's inequality, for all $\varepsilon > 0$,

$$\begin{aligned}
&P(k^2\|n^{-1}X'\widetilde{Z} - \mathrm{E}X_i\widetilde{Z}_i'\|^2 > \varepsilon) \\
&\leqslant k^2 \mathrm{E}\, tr\left(\left(n^{-1}\sum_{i=1}^n X_i\widetilde{Z}_i' - \mathrm{E}X_i\widetilde{Z}_i'\right)'\left(n^{-1}\sum_{j=1}^n X_j\widetilde{Z}_j' - \mathrm{E}X_j\widetilde{Z}_j'\right)\right)\bigg/\varepsilon \\
&= (k^2/n)\, tr((\mathrm{E}(X_i\widetilde{Z}_i' - \mathrm{E}X_i\widetilde{Z}_i')'(X_i\widetilde{Z}_i' - \mathrm{E}X_i\widetilde{Z}_i'))/\varepsilon \\
&\leqslant (k^3/n)p \sup_{j\leqslant k, n\geqslant 1} \mathrm{E}\|X_i\widetilde{Z}_{ij}\|^2 = \mathrm{o}(1),
\end{aligned} \tag{42}$$

where the first equality holds by the iid assumption, the second inequality uses the fact that the dimensions of X_i and $\widetilde{Z}_i$ are p and k, and the second equality uses Assumption 4.

To prove part (e), we write

$$\begin{aligned}
&n^{-1}Z'Z = n^{-1}\widetilde{Z}'\widetilde{Z} - n^{-1}\widetilde{Z}'X(X'X)^{-1}X'\widetilde{Z} \quad \text{and} \\
&\mathrm{E}Z_i^*Z_i^{*'} = \mathrm{E}\widetilde{Z}_i\widetilde{Z}_i' - \mathrm{E}\widetilde{Z}_iX_i'(\mathrm{E}X_iX_i')^{-1}\mathrm{E}X_i\widetilde{Z}_i'.
\end{aligned} \tag{43}$$

By the triangle inequality, we have

$$\begin{aligned}
&\|n^{-1}\widetilde{Z}'X(X'X)^{-1}X'\widetilde{Z} - \mathrm{E}\widetilde{Z}_iX_i'(\mathrm{E}X_iX_i')^{-1}\mathrm{E}X_i\widetilde{Z}_i'\| \leqslant L_{n1} + L_{n2} + L_{n3}, \text{ where} \\
&L_{n1} = \|n^{-1}\widetilde{Z}'X(n^{-1}X'X)^{-1}(n^{-1}X'\widetilde{Z} - \mathrm{E}X_i\widetilde{Z}_i')\|, \\
&L_{n2} = \|n^{-1}\widetilde{Z}'X[(n^{-1}X'X)^{-1} - (\mathrm{E}X_iX_i')^{-1}]\mathrm{E}X_i\widetilde{Z}_i'\|, \quad \text{and} \\
&L_{n3} = \|(n^{-1}\widetilde{Z}'X - \mathrm{E}\widetilde{Z}_iX_i')(\mathrm{E}X_iX_i')^{-1}\mathrm{E}X_i\widetilde{Z}_i'\|.
\end{aligned} \tag{44}$$

Using parts (c) and (d), we have

$$\|n^{-1}\widetilde{Z}'X\| \leqslant \|n^{-1}\widetilde{Z}'X - \mathrm{E}\widetilde{Z}_iX_i'\| + \|\mathrm{E}\widetilde{Z}_iX_i'\| = \mathrm{o_p}(k^{-1}) + \mathrm{O}(k^{1/2}) = \mathrm{O_p}(k^{1/2}). \tag{45}$$

In addition, $\|(n^{-1}X'X)^{-1}\| = \mathrm{O_p}(1)$ by the LLN, Slutsky's Theorem, and the fact $\mathrm{E}X_iX_i'$ is pd. These results, the result of part (d), and $\|AB\| \leqslant \|A\| \cdot \|B\|$ give

$$\begin{aligned}
L_{n1} &\leqslant \|n^{-1}\widetilde{Z}'X\| \cdot \|(n^{-1}X'X)^{-1}\| \cdot \|n^{-1}X'\widetilde{Z} - \mathrm{E}X_i\widetilde{Z}_i'\| \\
&= \mathrm{O_p}(k^{1/2})\mathrm{O_p}(1)\mathrm{o_p}(k^{-1}) = \mathrm{o_p}(k^{-1/2}).
\end{aligned} \tag{46}$$

By similar calculations, $L_{n3} = \mathrm{o_p}(k^{-1/2})$.

Using the results of (45) and parts (b) and (c), we have

$$\begin{aligned}
L_{n2} &\leqslant \|n^{-1}\widetilde{Z}'X\| \cdot \|(n^{-1}X'X)^{-1} - (\mathrm{E}X_iX_i')^{-1}\| \cdot \|\mathrm{E}X_i\widetilde{Z}_i'\| \\
&= \mathrm{O_p}(k^{1/2})\mathrm{O_p}(n^{-1/2})\mathrm{O}(k^{1/2}) = \mathrm{O_p}((k^3/n)^{1/2}k^{-1/2}) = \mathrm{o_p}(k^{-1/2}).
\end{aligned} \tag{47}$$

Hence, the left-hand side in (44) is $\mathrm{o_p}(k^{-1/2})$. This, (43), and part (a) combine to establish part (e). $\square$

Proof of Theorem 2. We prove part (a) for the case of $\tau = 1/2$. The proofs for other values of τ are analogous but use different scale factors. Define

$$\begin{aligned}&\widetilde{Q}_{S,n,k} = (\widehat{Q}_{S,n} - k)/k^{1/2}, \quad \widetilde{Q}_{T,n,k} = (\widehat{Q}_{T,n} - k)/k^{1/2},\\ &\widetilde{\kappa}_{LR,\alpha}(\widetilde{q}_T; k) = \kappa_{LR,\alpha}(\widetilde{q}_T k^{1/2} + k)/k^{1/2}, \quad \text{and } \widetilde{q}_T = (q_T - k)/k^{1/2}.\end{aligned} \tag{48}$$

Straightforward calculations give

$$P(\widehat{LR}_n > \kappa_{LR,\alpha}(q_T)|\widehat{Q}_{T,n} = q_T) = P(\widehat{LR}_n/k^{1/2} > \widetilde{\kappa}_{LR,\alpha}(\widetilde{q}_T; k)|\widetilde{Q}_{T,n} = \widetilde{q}_T). \tag{49}$$

Using (8), we can write

$$\widehat{LR}_n/k^{1/2} = \frac{1}{2}\left(\widetilde{Q}_{S,n,k} - \widetilde{Q}_{T,n,k} + \sqrt{(\widetilde{Q}_{S,n,k} - \widetilde{Q}_{T,n,k})^2 + 4\widehat{Q}_{ST,n}^2/k}\right). \tag{50}$$

Now, consider the case where $\widehat{LR}_n$ is defined with the true Ω rather than $\widehat{\Omega}_n$, the reduced-form errors are normal, and the null hypothesis holds. Call this "case N." In case N, $\widetilde{Q}_{S,n,k}$, $\widehat{Q}_{ST,n}/(\widehat{Q}_{S,n}\widehat{Q}_{T,n})^{1/2}$, and $\widetilde{Q}_{T,n,k}$ are independent, see Andrews et al. (2006a, Lemma 3). Also, in case N, the probability in (49) equals α for all $k \geqslant 1$ by the definition of $\kappa_{LR,\alpha}(\cdot)$.

Until stated otherwise below, we assume case N holds. Conditional on $\widetilde{Q}_{T,n,k} = \widetilde{q}_{T,n,k}$, we have

$$\frac{\widehat{Q}_{ST,n}}{k^{1/2}} = \frac{1}{k^{1/2}}\left(\frac{\widehat{Q}_{ST,n}}{(\widehat{Q}_{S,n}\widehat{Q}_{T,n})^{1/2}}\right)\widehat{Q}_{S,n}^{1/2}(k^{1/2}(\widetilde{Q}_{T,n,k} + k^{1/2}))^{1/2}. \tag{51}$$

By independence of the ratio in the parentheses from $\widehat{Q}_{S,n}^{1/2}(k^{1/2}(\widetilde{Q}_{T,n} + k^{1/2}))^{1/2}$, the conditional distribution of $\widehat{Q}_{ST,n}/k^{1/2}$ given $\widetilde{Q}_{T,n,k} = \widetilde{q}_{T,n,k}$ equals the unconditional distribution of

$$\begin{aligned}&\frac{1}{k^{1/2}}\left(\frac{\widehat{Q}_{ST,n}}{(\widehat{Q}_{S,n}\widehat{Q}_{T,n})^{1/2}}\right)\widehat{Q}_{S,n}^{1/2}(k^{1/2}(\widetilde{q}_{T,n,k} + k^{1/2}))^{1/2}\\ &\quad = \frac{\widehat{Q}_{ST,n}}{k^{1/2}}(\widehat{Q}_{T,n}/k)^{-1/2}(\widetilde{q}_{T,n,k}k^{-1/2} + 1)^{1/2} = \frac{\widehat{Q}_{ST,n}}{k^{1/2}}(1 + o_p(1)),\end{aligned} \tag{52}$$

where the second equality holds because $\widehat{Q}_{T,n}/k = 1 + o_p(1)$ (unconditionally) by Theorem 1. Eq. (52) implies that the conditional asymptotic distribution of $\widehat{Q}_{ST,n}/k^{1/2}$ as $n \to \infty$ given $\widetilde{Q}_{T,n,k} = \widetilde{q}_{T,n,k}$ equals the unconditional asymptotic distribution of $\widehat{Q}_{ST,n}/k^{1/2}$, which is the distribution of $\overline{Q}_{ST,\infty}$ by Theorem 1.

In consequence, using (50), for any sequence $\{\widetilde{q}_{T,n,k} : n \geqslant 1\}$ such that $\widetilde{q}_{T,n,k} \to \widetilde{q}_T$ for some $\widetilde{q}_T > 0$, the conditional distribution of $\widehat{LR}_n/k^{1/2}$ given $\widetilde{Q}_{T,n,k} = \widetilde{q}_{T,n,k}$ satisfies

$$\begin{aligned}\widehat{LR}_n/k^{1/2} &= \frac{1}{2}\left(\widetilde{Q}_{S,n,k} - \widetilde{q}_{T,n,k} + \sqrt{(\widetilde{Q}_{S,n,k} - \widetilde{q}_{T,n,k})^2 + 4\widehat{Q}_{ST,n}^2/k}\right)\\ &\to_d \frac{1}{2}(\overline{Q}_{S,\infty} - \widetilde{q}_T + \sqrt{(\overline{Q}_{S,\infty} - \widetilde{q}_T)^2 + 4\overline{Q}_{ST,\infty}^2}) = \overline{LR}_\infty(\widetilde{q}_T),\end{aligned} \tag{53}$$

where the last equality defines $\overline{LR}_\infty(\widetilde{q}_T)$.

Define $\widetilde{\kappa}_{LR,\alpha}(\widetilde{q}_T;\infty)$ by

$$P(\overline{LR}_\infty(\widetilde{q}_T) > \widetilde{\kappa}_{LR,\alpha}(\widetilde{q}_T;\infty)) = \alpha. \tag{54}$$

Given (53), some calculations show that for a sequence of constants $\{x_n : n \geqslant 1\}$ we have

$$\lim_{n\to\infty} P(\widehat{LR}_n/k^{1/2} > x_n | \widetilde{Q}_{T,n,k} = \widetilde{q}_{T,n,k}) = P(\overline{LR}_\infty(\widetilde{q}_T) > \widetilde{\kappa}_{LR,\alpha}(\widetilde{q}_T;\infty)) = \alpha \tag{55}$$

only if $x_n \to \widetilde{\kappa}_{LR,\alpha}(\widetilde{q}_T;\infty)$ as $n \to \infty$ because $\overline{LR}_\infty(\widetilde{q}_T)$ is absolutely continuous with strictly increasing distribution function. By (49) with $\widetilde{q}_T$ replaced by $\widetilde{q}_{T,n,k}$ and the fact that the probability in (49) equals α for all $k \geqslant 1$ and all $\widetilde{q}_T$, we have

$$\lim_{n\to\infty} P(\widehat{LR}_n/k^{1/2} > \widetilde{\kappa}_{LR,\alpha}(\widetilde{q}_{T,n,k};k) | \widetilde{Q}_{T,n,k} = \widetilde{q}_{T,n,k}) = \alpha. \tag{56}$$

This and (55) imply that

$$\widetilde{\kappa}_{LR,\alpha}(\widetilde{q}_{T,n,k};k) \to \widetilde{\kappa}_{LR,\alpha}(\widetilde{q}_T;\infty) \text{ as } n \to \infty \tag{57}$$

for any sequence $\widetilde{q}_{T,n,k} \to \widetilde{q}_T$ as $n \to \infty$.

We no longer assume case N, but we assume the null hypothesis holds. Eq. (57) and Theorem 1 gives

$$\widetilde{\kappa}_{LR,\alpha}(\widetilde{Q}_{T,n,k};k) \to \widetilde{\kappa}_{LR,\alpha}(\overline{Q}_{T,\infty};\infty) \quad \text{as } n \to \infty. \tag{58}$$

Note that the equality in (53) and Theorem 1 imply that $\widehat{LR}_n/k^{1/2} \to_d \overline{LR}_\infty(\overline{Q}_{T,\infty})$ as $n \to \infty$ (unconditionally and jointly with the convergence in (58)). Hence,

$$P(\widehat{LR}_n/k^{1/2} > \widetilde{\kappa}_{LR,\alpha}(\widetilde{Q}_{T,n,k};k)) \to P(\overline{LR}_\infty(\overline{Q}_{T,\infty}) > \widetilde{\kappa}_{LR,\alpha}(\overline{Q}_{T,\infty};\infty)) = \alpha, \tag{59}$$

where the equality holds by (54) using iterated expectations. By the definition of $\widetilde{\kappa}_{LR,\alpha}(\widetilde{Q}_{T,n,k};k)$, the left-hand side in (59) equals $P(\widehat{LR}_n > \kappa_{LR,\alpha}(\widehat{Q}_{T,n}))$, similarly to (49). Hence, part (a) is proved.

Part (b) is an immediate consequence of Theorem 1 because $\widehat{LM}_n$ has a $\chi^2_1(0)$ distribution for all $\tau \in (0,2]$.

Next, we prove part (c). Under case N, we have

$$\alpha = P(\widehat{AR}_n > F_{k,n-k-p}) = P((\widehat{AR}_n - 1)k^{1/2}/\sqrt{2} > (F_{k,n-k-p} - 1)k^{1/2}/\sqrt{2}). \tag{60}$$

By Theorem 1, $(\widehat{AR}_n - 1)k^{1/2}/\sqrt{2} \to_d N(0,1)$ in case N. Hence, $(F_{k,n-k-p} - 1)k^{1/2}/\sqrt{2}) \to z_{1-\alpha}$ as $n \to \infty$. Thus, when case N does not necessarily hold, but the null hypothesis holds, we have

$$\begin{aligned} P(\widehat{AR}_n > F_{k,n-k-p}) &= P((\widehat{AR}_n - 1)k^{1/2}/\sqrt{2} > (F_{k,n-k-p} - 1)k^{1/2}/\sqrt{2}) \\ &\to P(Z > z_{1-\alpha}) = \alpha, \end{aligned} \tag{61}$$

where $Z \sim \mathrm{N}(0,1)$ and the convergence uses Theorem 1. $\square$

Acknowledgments

The contents of this paper appeared first in Andrews and Stock (2005). Andrews and Stock gratefully acknowledge the research support of the National Science Foundation via Grant numbers SES-0417911 and SBR-0214131, respectively.

References

Anderson, T.W., 1976. Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics. Journal of the Royal Statistical Society, Series B 38, 1–36.

Anderson, T.W., Rubin, H., 1949. Estimators of the parameters of a single equation in a complete set of stochastic equations. Annals of Mathematical Statistics 21, 570–582.

Anderson, T.W., Kunitomo, N., Matsushita, Y., 2005. A new light from old wisdoms: alternative estimation methods of simultaneous equations and microeconometric models. Unpublished Manuscript, Graduate School of Economics, University of Tokyo.

Andrews, D.W.K., J.H. Stock, 2005. Inference with weak instruments. Cowles Foundation Discussion Paper No. 1530, Yale University.

Andrews, D.W.K., Stock, J.H., 2006. Inference with weak instruments. In: Blundell, R., Newey, W.K., Persson, T. (Eds.), Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society. Cambridge University Press, Cambridge, UK, forthcoming.

Andrews, D.W.K., Moreira, M.J., Stock, J.H., 2006a. Optimal two-sided invariant similar tests for instrumental variables regression with weak instruments. Econometrica 74, 715–752.

Andrews, D.W.K., Moreira, M.J., Stock, J.H., 2006b. Supplement to Optimal two-sided invariant similar tests for instrumental variables regression. Available on the Econometric Society website at ⟨www.econometricsociety.org⟩ under Supplemental Material and at ⟨http://ksghome.harvard.edu/~.JStock.Academic .Ksg/ams/websupp/index.htm⟩.

Angrist, J.D., Krueger, A.B., 1991. Does compulsory school attendance affect schooling and earnings? Quarterly Journal of Economics 106, 979–1014.

Bekker, P.A., 1994. Alternative approximations to the distributions of instrumental variable estimators. Econometrica 62, 657–681.

Bekker, P.A., Kleibergen, F., 2003. Finite-sample instrumental variables inference using an asymptotically pivotal statistic. Econometric Theory 19, 744–753.

Brown, B.M., 1971. Martingale central limit theorems. Annals of Mathematical Statistics 42, 59–66.

Chamberlain, G., Imbens, G., 2004. Random effects estimators with many instrumental variables. Econometrica 72, 295–306.

Chao, J.C., Swanson, N.R., 2005. Consistent estimation with a large number of weak instruments. Econometrica 73, 1673–1692.

Chao, J.C., Swanson, N.R., 2006. Alternative approximations of the bias and MSE of the IV estimator under weak identification with an application to bias correction. Journal of Econometrics, forthcoming.

Donald, S.G., Newey, W.K., 2001. Choosing the number of instruments. Econometrica 69, 1161–1191.

Hahn, J., 2002. Optimal inference with many instruments. Econometric Theory 18, 140–168.

Hahn, J., Hausman, J., Kuersteiner, G., 2004. Estimation with weak instruments: accuracy of higher-order bias and MSE approximations. Econometrics Journal 7, 272–306.

Hall, P., 1984. Central limit theorem for integrated square error of multivariate nonparametric density estimators. Journal of Multivariate Analysis 14, 1–16.

Han, C., Phillips, P.C.B., 2006. GMM with many moment conditions. Econometrica 74, 147–192.

Hansen, C., Hausman, J., Newey, W.K., 2005. Estimation with many instrumental variables. Unpublished Manuscript, Department of Economics, M.I.T.

Kleibergen, F., 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. Econometrica 70, 1781–1803.

Kunitomo, N., 1980. Asymptotic expansions of the distribution of estimators in a linear functional relationship and simultaneous equations. Journal of the American Statistical Association 75, 693–700.

Moreira, M.J., 2001. Tests with correct size when instruments can be arbitrarily weak. Unpublished Manuscript, Department of Economics, University of California, Berkeley.

Moreira, M.J., 2003. A conditional likelihood ratio test for structural models. Econometrica 71, 1027–1048.
Morimune, K., 1983. Approximate distributions of the k-class estimators when the degree of overidentifiability is large compared with the sample size. Econometrica 51, 821–841.
Newey, W.K., Windmeijer, F., 2005. Many weak moment asymptotics for generalized empirical likelihood estimators. Unpublished Manuscript, Department of Economics, M.I.T.
Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.
Stock, J.H., Yogo, M., 2005. Asymptotic distributions of instrumental variables statistics with many instruments. In: Andrews, D.W.K., Stock, J.H. (Eds.), Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas J. Rothenberg. Cambridge University Press, Cambridge, UK, pp. 109–120.