Testing with Many Weak Instruments

Donald W. K. Andrews[1]
Cowles Foundation for Research in Economics
Yale University

James H. Stock
Department of Economics
Harvard University

March 2005
Revised: September 2005

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

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 power results show that the conditional LR test is more powerful asymptotically than the AR and LM tests under many weak IV asymptotics.

Keywords: Anderson-Rubin test, conditional likelihood ratio test, instrumental variables, Lagrange multiplier test, many instrumental variables, weak instruments.

JEL Classification Numbers: C12, C30.

1 Introduction

This paper presents new results for testing in the linear IV regression model under "many weak instrument (IV) asymptotics." Such asymptotics are designed for the case in which the IVs are weak and the number of IVs, k, is relatively large compared to the sample size n. Asymptotics of this type have been considered recently by Chao and Swanson (2005), Stock and Yogo (2005), Han and Phillips (2005), Anderson, Kunitomo, and Matsushita (2005), Hansen, Hausman, and Newey (2005), Newey and Windmeijer (2005), and Andrews and Stock (2005a). 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 consider the conditional likelihood ratio (CLR) test introduced in Moreira (2003), the Anderson-Rubin (1949) (AR) test, and a Lagrange multiplier (LM) test introduced in Kleibergen (2002) and Moreira (2001). We find that in the many weak IV asymptotic set-up 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. With non-normal errors, $k^3/n \to 0$ as $n \to \infty$ is required for our results to show that the tests have the correct asymptotic size.

Andrews and Stock (2005a) 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.

We conclude that the "many weak IV" asymptotic results for non-normal errors given in the present paper buttress the arguments in Andrews, Moreira, and Stock (2004) and Andrews and Stock (2005a) 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 (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 (2003). 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, Hausman, and Kuersteiner (2004), and Hansen, Hausman, and Newey (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:

$$\begin{aligned} y_1 &= y_2\beta + X\gamma_1 + u, \\ y_2 &= \widetilde{Z}\pi + X\xi_1 + v_2, \end{aligned} \tag{2.1}$$

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

$$\begin{aligned} y_1 &= \widetilde{Z}\pi\beta + X\gamma_1 + v_1 \\ y_2 &= \widetilde{Z}\pi + X\xi_1 + v_2, \text{ where} \\ v_1 &= u + v_2\beta. \end{aligned} \tag{2.2}$$

The reduced-form errors $[v_1{:}v_2]$ are iid across rows with each row having mean zero and 2×2 nonsingular 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

$$\begin{aligned} Y_i &= a\pi'\widetilde{Z}_i + \eta' X_i + V_i \text{ for } i \le n, \text{ where} \\ a &= (\beta, 1)', \text{ and } \eta = [\gamma_1 : \xi_1] \in R^{p\times 2}. \end{aligned} \tag{2.3}$$

Define

$$\begin{aligned} Z_i^* &= \widetilde{Z}_i - E\widetilde{Z}_i X_i'(EX_iX_i')^{-1}X_i \text{ and} \\ \lambda_{n,k}^* &= n\pi' EZ_i^* Z_i^{*\prime}\pi. \end{aligned} \tag{2.4}$$

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

We use the following assumptions.

Assumption 1. $\{(V_i, X_i, \widetilde{Z}_i) : i \le n\}$ are iid across i for each n and $\{(V_i, X_i) : i \le n; n \ge 1\}$ 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, $\liminf_{n\to\infty} \lambda_{\min}(EZ_i^*Z_i^{*\prime}) > 0$, and $\sup_{j\leq k;n\geq 1}(E||V_i||^4\widetilde{Z}_{ij}^4 + E||V_i||^4 + E\widetilde{Z}_{ij}^4 + E||X_i||^4) < \infty$, where $\widetilde{Z}_i = (\widetilde{Z}_{i1}, ..., \widetilde{Z}_{ik})'$.

Assumption 3. $EV_iV_i' = \Omega$, $E(V_iV_i' \otimes Z_i^*Z_i^{*\prime}) = \Omega \otimes EZ_i^*Z_i^{*\prime}$ for all $n \geq 1$, 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 $\tau \leq 1/2$; $\beta = \beta_0 + Bk^{1/2-\tau}$ when $\tau \in (1/2, 1]$; and $\beta = \beta_0 + Bk^{-\tau/2}$ when $\tau \geq 1$.

Assumption 1 states that the errors, exogenous variables, and IVs are random and iid across $i \leq n$. 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 uncorelated 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.

Assumption 5 controls the relative magnitude of the IV strength, as measured by $\lambda_{n,k}^*$, to the number of IVs k. For example, Assumptions 5 holds if $\pi = C(k^\tau/n)^{1/2}$ for some $C \in R^k$ with $||C|| = 1$ and $CEZ_i^*Z_i^{*\prime}C \to r_\tau$. The smaller is τ, the weaker are the IVs relative to k. Andrews and Stock (2005a) 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 (5.6) 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 $\tau \leq 1/2$ 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 $\tau \leq 1/2$ 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:

$$H_0 : \beta = \beta_0 \text{ and } H_1 : \beta \neq \beta_0. \tag{3.1}$$

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

$$\widehat{\Omega}_n = (n-k-p)^{-1}\widehat{V}'\widehat{V}, \text{ where } \widehat{V} = Y - P_{\widetilde{Z}}Y - P_X Y. \tag{3.2}$$

We define

$$\begin{aligned}
\widehat{S}_n &= (Z'Z)^{-1/2}Z'Yb_0 \cdot (b_0'\widehat{\Omega}_n b_0)^{-1/2}, \text{ where } Z = \widetilde{Z} - P_X\widetilde{Z} \text{ and } b_0 = (1, -\beta_0)', \\
\widehat{T}_n &= (Z'Z)^{-1/2}Z'Y\widehat{\Omega}_n^{-1}a_0 \cdot (a_0'\widehat{\Omega}_n^{-1}a_0)^{-1/2}, \text{ where } a_0 = (\beta_0, 1)', \\
\widehat{Q}_{\lambda,k,n} &= [\widehat{S}_n : \widehat{T}_n]'[\widehat{S}_n : \widehat{T}_n] = \begin{bmatrix} \widehat{S}_n'\widehat{S}_n & \widehat{S}_n'\widehat{T}_n \\ \widehat{T}_n'\widehat{S}_n & \widehat{T}_n'\widehat{T}_n \end{bmatrix} = \begin{bmatrix} \widehat{Q}_{S,n} & \widehat{Q}_{ST,n} \\ \widehat{Q}_{ST,n} & \widehat{Q}_{T,n} \end{bmatrix}, \text{ and} \\
\widehat{\widehat{Q}}_{\lambda,k,n} &= (\widehat{Q}_{\lambda,k,n} - kI_2)/k^{1/2}.
\end{aligned} \tag{3.3}$$

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

$$\begin{aligned}
\widehat{AR}_n &= \widehat{Q}_{S,n}/k, \\
\widehat{LM}_n &= \widehat{Q}_{ST,n}^2/\widehat{Q}_{T,n}, \text{ and} \\
\widehat{LR}_n &= \frac{1}{2}\left(\widehat{Q}_{S,n} - \widehat{Q}_{T,n} + \sqrt{(\widehat{Q}_{S,n} - \widehat{Q}_{T,n})^2 + 4\widehat{Q}_{ST,n}^2}\right)
\end{aligned} \tag{3.4}$$

(see Andrews, Moreira, and Stock (2004) and Andrews and Stock (2005a).)

Under H_0, $\widehat{AR}_n \to_d \chi_k^2/k$ and $\widehat{LM}_n \to_d \chi_1^2$ as $n \to \infty$ under strong and weak IV asymptotics assuming iid homoskedastic errors and k fixed for all n (e.g., see Andrews, Moreiera, and Stock (2004)). 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_1^2 critical value is used for the LM test.

The CLR test rejects the null hypothesis when

$$\widehat{LR}_n > \kappa_\alpha^{CLR}(\widehat{Q}_{T,n}), \tag{3.5}$$

where the conditional critical value function $\kappa_\alpha^{CLR}(\widehat{Q}_{T,n})$ is defined to satisfy $P_{\beta_0}(\widehat{LR}_n > \kappa_\alpha^{CLR}(q_T)|\widehat{Q}_{T,n} = q_T) = \alpha$. Moreira (2003) gives a table of $\kappa_\alpha^{CLR}(q_T)$ 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 (2005a), we do not assume that the errors are normally distributed.

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

$$
\begin{aligned}
c_\beta &= (\beta - \beta_0) \cdot (b_0'\Omega b_0)^{-1/2} \in R, \\
d_\beta &= a'\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \in R, \\
V_{3,\tau} &= \begin{cases} Diag\{2,1,2\} & \text{if } 0 < \tau \leq 1/2 \\ Diag\{2,1,0\} & \text{if } 1/2 < \tau < 1 \\ Diag\{2, 1 + d_{\beta_0}^2 r_1, 0\} & \text{if } \tau = 1 \\ Diag\{2, d_{\beta_0}^2 r_\tau, 0\} & \text{if } \tau > 1, \text{ and} \end{cases} \\
\gamma_B &= (b_0'\Omega b_0)^{-1/2} d_{\beta_0} B \in R, \qquad (4.1)
\end{aligned}
$$

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 noncentrality parameter δ.

Theorem 1 *Suppose Assumptions* 1-6 *hold. Then, the following results hold.*
(a) *If* $0 < \tau < 1/2$ *and* β *is fixed,*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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. The critical value function for the CLR test and critical values for the AR and LM tests yield the correct asymptotic significance level α under normality of the errors. Hence, the three tests also have the desired asymptotic significance level α under non-normality.

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 (2005a) 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 (2005a) 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).[1] 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) and the AR test has trivial power against these alternatives.

3. Note that the cases considered in Chao and Swanson (2005) and Han and Phillips (2005) correspond to $\tau > 1/2$. Those considered in Stock and Yogo (2005), Anderson, Kunitomo, and Matsushita (2005), Hansen, Hausman, and Newey (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.

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 level provided $k^3/n \rightarrow 0$ as $n \rightarrow \infty$ no matter how weak are the IVs. For

[1] By "essentially," we mean that exhaustive simulations show that the asymptotic power of the CLR test is on the asymptotic power envelope.

normal errors, the results of Andrews and Stock (2005a) 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, Moreira, and Stock (2004), 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. We start by stating several Lemmas, the purposes of which are discussed following Lemma 1 below. Lemma 1 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, Thm. 1) univariate CLT for degenerate U-statistics. Newey and Windmeijer (2005, Lem. A2) makes a similar use of Hall's result when establishing the asymptotic distribution of empirical likelihood estimators with many weak IVs.

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

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

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

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

Comment. In Lemma 1 (and Lemma 2 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 Lemmas 3 and 4 below.

We now summarize the purpose of Lemma 1 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 1(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 1 does not hold. Hence, Lemma 1 is extended in Lemma 2 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 1 does not hold. Hence, Lemma 2 is extended in Lemma 3 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 $Z_i^* = \widetilde{Z}_i - [E\widetilde{Z}_iX_i'(EX_iX_i')^{-1}]X_i$, so Lemma 3 is extended in Lemma 4 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 4 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 $\tau < 1$. Lemmas 5 and 6 provide results analogous to those of Lemmas 1-4 for the case in which the means are of a magnitude that corresponds to $\tau \in [1,2]$. Finally, Lemma 7 provides results on the asymptotic behavior of the sample matrices $\widetilde{Z}'\widetilde{Z}$, $X'X$, $X'\widetilde{Z}$, and $Z'Z$.

Lemma 2 *Let* $\{(\xi_{ni},\eta_{ni}) : i \leq n; n \geq 1\}$ *be a triangular array of random vectors that satisfies the assumptions of Lemma 1, but 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* $\lambda_{n\xi} = n\mu_{n\xi}'\mu_{n\xi}$, $\lambda_{n\eta} = n\mu_{n\eta}'\mu_{n\eta}$, *and* $\lambda_{n\xi\eta} = n\mu_{n\xi}'\mu_{n\eta}$. *Then,*

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

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

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

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

Lemma 3 *Let* $\{(\xi_{ni},\eta_{ni}) : i \leq n; n \geq 1\}$ *be a triangular array of random vectors that satisfies the assumptions of Lemma 1, 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* $\lambda_{n\xi}^* = n\mu_{n\xi}'\Sigma_{n\xi}^{-1}\mu_{n\xi}$, $\lambda_{n\eta}^* = n\mu_{n\eta}'\Sigma_{n\eta}^{-1}\mu_{n\eta}$, *and* $\lambda_{n\xi\eta}^* = n\mu_{n\xi}'\Sigma_{n\xi}^{-1/2}\Sigma_{n\eta}^{-1/2}\mu_{n\eta}$, *with assumption* (vii) *replaced by* (vii)$'$ $k^3/n \to 0$, *and with the addition of assumption* (viii) $\inf_{n\geq 1} \lambda_{\min}(\Sigma_{n\xi}) > 0$ *and* $\inf_{n\geq 1} \lambda_{\min}(\Sigma_{n\eta}) > 0$. *Then,*

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

Lemma 4 *Suppose* (a) $(\xi_{ni}, \eta_{ni}, \widehat{\Sigma}_{n\xi}, \widehat{\Sigma}_{n\eta})$ *satisfy the conditions of Lemma 3,* (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) $\{(\xi_{n2i},\eta_{n2i}) : i \leq n\}$ *are iid across* $i \leq n$ *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 3 holds with* $(\widehat{\xi}_{ni}, \widehat{\eta}_{ni})$ *in place of* (ξ_{ni}, η_{ni}).

The next Lemma is an extension of Lemma 1 that is needed when $\tau \geq 1$.

Lemma 5 *Let* $\{(\xi_{ni}, \eta_{ni}) : i \leq n; n \geq 1\}$ *be a triangular array of random vectors that satisfies the conditions of Lemma 1. Let* $\{h_n : n \geq 1\}$ *be a sequence of constant* k*-vectors that satisfies* $nh_n'h_n/k^\tau \to m$ *as* $n \to \infty$ *for some constants* $\tau > 0$ *and* $m \geq 0$. *Then,*

$$\begin{pmatrix} \frac{1}{nk^{1/2}} \sum\sum_{1\leq i<j\leq n} \begin{pmatrix} 2\xi_{ni}'\xi_{nj} \\ \xi_{ni}'\eta_{nj} + \xi_{nj}'\eta_{ni} \\ 2\eta_{ni}'\eta_{nj} \end{pmatrix} \\ \frac{1}{k^{\tau/2}} \sum_{i=1}^n \begin{pmatrix} h_n'\xi_{ni} \\ h_n'\eta_{ni} \end{pmatrix} \end{pmatrix} \to_d N(0, V_5), \text{ where } V_5 = Diag\{2, 1, 2, m, m\}.$$

Comment. The proof of Lemma 5 is similar to Newey and Windmeijer's (2005) proof of their Lemma A.2, which is an extension of Hall's (1984) CLT for degenerate U-statistics to cover joint convergence with sums of iid random variables. However, Newey and Windmeijer's Lemma A.2 does not cover the terms $\xi_{ni}'\xi_{ni}$ and $\eta_{ni}'\eta_{ni}$ in the Lemma 5 (because of their assumption that $EY_iZ_i = 0$).

The next result is employed when $\tau \in [1, 2]$. For a constant $\kappa_{\eta\tau} \geq 0$, define

$$V_{3,\tau}^* = \begin{cases} Diag\{2, 1 + \kappa_{\eta\tau}, 2 + 4\kappa_{\eta\tau}\} & \text{if } \tau = 1 \\ Diag\{2, \kappa_{\eta\tau}, 4\kappa_{\eta\tau}\} & \text{if } \tau > 1. \end{cases} \tag{5.1}$$

Lemma 6 *Suppose* $\tau \in [1, 2]$ *and* $(\xi_{n1i}, \xi_{n2i}, \eta_{n1i}, \eta_{n2i}, \widehat{D}_{n\xi}, \widehat{D}_{n\eta}, \widehat{\Sigma}_{n\xi}, \widehat{\Sigma}_{n\eta})$ *satisfy the conditions of Lemma 4 but with* $||E\eta_{n2i}||^2 = O(k^{1+\tau}/n)$, *rather than* $||E\eta_{n2i}||^2 = O(k^2/n)$, *and with* $\lambda_{n\xi}/k \to 0$ *and* $\lambda_{n\eta}/k^\tau \to \kappa_{\eta\tau}$ *for some constant* $\kappa_{\eta\tau} \geq 0$, *rather than* $(\lambda_{n\xi} + \lambda_{n\eta})/k \to 0$. *Then,*

$$\text{(a) } \frac{1}{n} \sum\sum_{1\leq i<j\leq n} \begin{pmatrix} k^{-1/2}[2\xi_{ni}{}'\widehat{\Sigma}_{n\xi}^{-1}\xi_{nj}] \\ k^{-\tau/2}[(\widehat{\Sigma}_{n\xi}^{-1/2}\xi_{ni})'\widehat{\Sigma}_{n\eta}^{-1/2}\eta_{nj} + (\widehat{\Sigma}_{n\xi}^{-1/2}\xi_{nj})'\widehat{\Sigma}_{n\eta}^{-1/2}\eta_{ni}] \\ k^{-\tau/2}[2\eta_{ni}'\widehat{\Sigma}_{n\eta}^{-1}\eta_{nj}] \end{pmatrix}$$

$$- \begin{pmatrix} \lambda_{n\xi}/k^{1/2} \\ \lambda_{n\xi\eta}/k^{\tau/2} \\ \lambda_{n\eta}/k^{\tau/2} \end{pmatrix} \to_d N(0, V_{3,\tau}^*),$$

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

$$\text{(c) } \begin{pmatrix} k^{-1/2}[n^{-1/2}\sum_{i=1}^n(\widehat{\Sigma}_{n\xi}^{-1/2}\xi_{ni})'n^{-1/2}\sum_{j=1}^n \widehat{\Sigma}_{n\xi}^{-1/2}\xi_{nj} - k] \\ k^{-\tau/2}[n^{-1/2}\sum_{i=1}^n(\widehat{\Sigma}_{n\xi}^{-1/2}\xi_{ni})'n^{-1/2}\sum_{j=1}^n \widehat{\Sigma}_{n\eta}^{-1/2}\eta_{nj}] \\ k^{-\tau/2}[n^{-1/2}\sum_{i=1}^n(\widehat{\Sigma}_{n\eta}^{-1/2}\eta_{ni})'n^{-1/2}\sum_{j=1}^n \widehat{\Sigma}_{n\eta}^{-1/2}\eta_{nj} - k] \end{pmatrix} - \begin{pmatrix} \lambda_{n\xi}/k^{1/2} \\ \lambda_{n\xi\eta}/k^{\tau/2} \\ \lambda_{n\eta}/k^{\tau/2} \end{pmatrix}$$

$$\to_d N(0, V_{3,\tau}^*).$$

Comment. In the proof of Lemma 6, the constraint that $\tau \leq 2$ is used only at the end of the proof when showing that an analogue of Lemma 4 holds. In addition, the proof of part (b) uses $\tau < 6$ at an earlier stage of the proof.

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) $||n^{-1}Z'Z - EZ_i^*Z_i^{*\prime}|| = o_p(k^{-1/2})$.

Proof of Theorem 1. First, we show that the results of the Theorem hold for $\tau \in (0,1)$ 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 4 with

$$\begin{aligned}
\xi_{ni} &= Z_i^* \cdot Y_i'b_*,\ \xi_{n1i} = \widetilde{Z}_i \cdot Y_i'b_*,\ \xi_{n2i} = (X_i' \cdot Y_i'b_*, 0_{k-p}')' \in R^k, \\
\widehat{\Sigma}_{n\xi} &= \widehat{\Sigma}_{n\eta} = n^{-1}Z'Z,\ D_{n\xi} = D_{n\eta} = [E\widetilde{Z}_iX_i'(EX_iX_i')^{-1} : 0_{k\times(k-p)}] \in R^{k\times k}, \\
\widehat{D}_{n\xi} &= \widehat{D}_{n\eta} = [n^{-1}\widetilde{Z}'X(n^{-1}X'X)^{-1} : 0_{k\times(k-p)}] \in R^{k\times k}, \qquad (5.2) \\
\eta_{ni} &= Z_i^* \cdot Y_i'a_*,\ \eta_{n1i} = \widetilde{Z}_i \cdot Y_i'a_*,\ \text{and } \eta_{n2i} = (X_i' \cdot Y_i'a_*, 0_{k-p}')' \in R^k.
\end{aligned}$$

Assumptions (b) and (e) of Lemma 4 follow immediately from (5.2).

Assumption (a) of Lemma 4 requires that Assumptions (i), (ii), (iii)$''$, (iv)$'$, and (v)-(viii) of Lemmas 1-3 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

$$\begin{aligned}
\Sigma_{n\xi} &= Var(\xi_{ni}) = EZ_i^*Z_i^{*\prime} \cdot E(V_i'b_0)^2(b_0'\Omega b_0)^{-1} = EZ_i^*Z_i^{*\prime}, \\
\Sigma_{n\eta} &= Var(\eta_{ni}) = EZ_i^*Z_i^{*\prime} \cdot E(V_i'\Omega^{-1}a_0)^2(a_0'\Omega^{-1}a_0)^{-1} = EZ_i^*Z_i^{*\prime}, \\
Cov(\xi_{ni}, \eta_{ni}) &= EZ_i^*Z_i^{*\prime} \cdot E(b_0'V_iV_i'\Omega^{-1}a_0)(b_0'\Omega b_0)^{-1/2}(a_0'\Omega^{-1}a_0)^{-1/2} = 0, \quad (5.3)
\end{aligned}$$

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

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

$$\begin{aligned}
&\mu_{n\xi} = E\xi_{ni} = EZ_i^*Y_i'b_* = EZ_i^*(\widetilde{Z}_i'\pi a' + X_i'\eta)b_* = EZ_i^*Z_i^{*\prime}\pi a'b_* = EZ_i^*Z_i^{*\prime}\pi c_\beta, \\
&\mu_{n\eta} = E\eta_{ni} = EZ_i^*Y_i'a_* = EZ_i^*Z_i^{*\prime}\pi a'\Omega^{-1}a_* = EZ_i^*Z_i^{*\prime}\pi d_\beta, \qquad (5.4) \\
&(\lambda_{n\xi}^*, \lambda_{n\xi\eta}^*, \lambda_{n\eta}^*) = n\pi'EZ_i^*Z_i^{*\prime}\pi \cdot (c_\beta^2, c_\beta d_\beta, d_\beta^2), \\
&(\lambda_{n\xi}^* + \lambda_{n\eta}^*)/k = (\lambda_{n,k}^*/k^\tau)k^{\tau-1}(c_\beta^2 + d_\beta^2) = O(k^{\tau-1}) = o(1),
\end{aligned}$$

where $\lambda_{n,k}^*$ is defined in (2.4) and the last equality holds because $\tau < 1$.

To show assumption (c) of Lemma 4, we write

$$||E\xi_{n2i}||^2 = ||EX_i \cdot Y_i'b_*||^2 = ||EX_i\widetilde{Z}_i'\pi||^2(a'b_*)^2 \leq ||EX_i\widetilde{Z}_i'||^2 \cdot ||\pi||^2(a'b_*)^2, \quad (5.5)$$

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

$$O(1) = \lambda_{n,k}^*/k^\tau = n\pi'EZ_i^*Z_i^{*\prime}\pi/k^\tau \geq n\pi'\pi\lambda_{\min}(EZ_i^*Z_i^{*\prime})/k^\tau. \qquad (5.6)$$

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

Assumption (d) of Lemma 4 holds because

$$\begin{aligned}
||\widehat{D}_{n\xi} - D_{n\xi}|| &= ||n^{-1}\widetilde{Z}'X(n^{-1}X'X)^{-1} - E\widetilde{Z}_iX_i'(n^{-1}X'X)^{-1} \\
&\quad + E\widetilde{Z}_iX_i'(n^{-1}X'X)^{-1} - E\widetilde{Z}_iX_i'(EX_iX_i')^{-1}|| \\
&\le ||n^{-1}\widetilde{Z}'X - E\widetilde{Z}_iX_i'|| \cdot ||(n^{-1}X'X)^{-1}|| \\
&\quad + ||E\widetilde{Z}_iX_i'|| \cdot ||(n^{-1}X'X)^{-1} - (EX_iX_i')^{-1}|| \\
&= o_p(k^{-1})O_p(1) + O(k^{1/2})O_p(n^{-1/2}) = o_p(k^{-1}), \qquad (5.7)
\end{aligned}$$

where the second equality holds by Lemma 7(b)-(d) and the fact that $(n^{-1}X'X)^{-1} = O_p(1)$ by the WLLNs, Slutsky's Theorem, and $EX_iX_i' > 0$ 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 (5.4) and Assumptions 5 and 6, we have

$$\begin{aligned}
(\lambda^*_{n\xi}/k^{1/2}, \lambda^*_{n\xi\eta}/k^{1/2}, \lambda^*_{n\eta}/k^{1/2})' &= (\lambda^*_{n,k}/k^{1/2}) \cdot (c_\beta^2, c_\beta d_\beta, d_\beta^2)' \\
&\to \begin{cases} r_{1/2}(c_\beta^2, c_\beta d_\beta, d_\beta^2)' & \text{when } \tau = 1/2 \\ (0,0,0)' & \text{when } \tau < 1/2. \end{cases} \qquad (5.8)
\end{aligned}$$

When $\tau \in (1/2, 1)$, (5.4) and Assumptions 5 and 6 lead to

$$\begin{aligned}
& c_\beta = Bk^{1/2-\tau}\overline{b}_0,\ \overline{b}_0 = (b_0'\Omega b_0)^{-1/2},\ d_\beta = d_{\beta_0}(1 + o(1)), \text{ and} \\
& (\lambda^*_{n\xi}/k^{1/2}, \lambda^*_{n\xi\eta}/k^{1/2}, \lambda^*_{n\eta}/k^\tau)' = (c_\beta^2\lambda^*_{n,k}k^{-1/2}, c_\beta d_\beta\lambda^*_{n,k}k^{-1/2}, d_\beta^2\lambda^*_{n,k}k^{-\tau})' \\
& = (B^2\overline{b}_0^2k^{1-2\tau}\lambda^*_{n,k}k^{-1/2}, B\overline{b}_0d_\beta k^{1/2-\tau}\lambda^*_{n,k}k^{-1/2}, d_\beta^2\lambda^*_{n,k}k^{-\tau})' \\
& \to (0, B\overline{b}_0d_{\beta_0}r_\tau, d_{\beta_0}^2r_\tau)' = (0, \gamma_Br_\tau, d_{\beta_0}^2r_\tau)'. \qquad (5.9)
\end{aligned}$$

Hence, when $\tau \in (1/2, 1)$, Lemma 4 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 $d_\beta^2\lambda^*_{n,k}/k^\tau \to d_{\beta_0}^2r_\tau$, gives the desired result that $(\widehat{T}_n{}'\widehat{T}_n - k)/k^\tau \to_p d_{\beta_0}^2r_\tau$ 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 Thm. 1 of Andrews and Stock (2005a).

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 (2005a) 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 (2005a).

Next, we consider the case where $\tau \in [1, 2]$. The proof is the same as for $\tau < 1$ with the following changes. First, we apply Lemma 6 instead of Lemma 4. Second, the last line of (5.4) does not hold because $\tau \ge 1$. Instead, we have $c_\beta^2 = O((\beta - \beta_0)^2) = O(k^{-\tau})$, $\lambda^*_{n\xi}/k = (\lambda^*_{n,k}/k^\tau)k^{\tau-1}c_\beta^2 = O(k^{-1}) = o(1)$ as required by Lemma 6, and $\lambda^*_{n\eta}/k^\tau = (\lambda^*_{n,k}/k^\tau)d_\beta^2 \to r_\tau d_{\beta_0}^2$, which verifies an assumption in Lemma 6 with

$\kappa_{\eta\tau} = r_\tau d_{\beta_0}^2$. Third, by (5.5) and (5.6) with ξ replaced by η, we have $||E\eta_{n2i}||^2 \leq O(||EX_i \widetilde{Z}_i'||^2 \cdot ||\pi||^2) = O(k^{1+\tau}/n)$ as is assumed in Lemma 6.

Fourth, when $\tau = [1, 2]$, (5.4) and Assumptions 5 and 6 lead to

$$\begin{aligned}
&c_\beta = Bk^{-\tau/2}\overline{b}_0,\ \overline{b}_0 = (b_0'\Omega b_0)^{-1/2},\ d_\beta = d_{\beta_0}(1+o(1)),\ \text{and} \\
&(\lambda_{n\xi}^*/k^{1/2}, \lambda_{n\xi\eta}^*/k^{\tau/2}, \lambda_{n\eta}^*/k^\tau)' = (c_\beta^2\lambda_{n,k}^* k^{-1/2}, c_\beta d_\beta \lambda_{n,k}^* k^{-\tau/2}, d_\beta^2 \lambda_{n,k}^* k^{-\tau})' \\
&= (B^2\overline{b}_0^2 k^{-\tau}\lambda_{n,k}^* k^{-1/2}, B\overline{b}_0 d_\beta k^{\tau/2}\lambda_{n,k}^* k^{-\tau/2}, d_\beta^2\lambda_{n,k}^* k^{-\tau})' \\
&\to (0, B\overline{b}_0 d_{\beta_0} r_\tau, d_{\beta_0}^2 r_\tau)' = (0, \gamma_B r_\tau, d_{\beta_0}^2 r_\tau)'. \qquad (5.10)
\end{aligned}$$

Hence, Lemma 6 shows that $((\widehat{S}_n{}'\widehat{S}_n - k)/k^{1/2}, \widehat{S}_n{}'\widehat{T}_n/k^{\tau/2}) \to_d (\overline{Q}_{S,\infty}, \overline{Q}_{ST,\infty})$. In addition, Lemma 6 implies that $(\widehat{T}_n'\widehat{T}_n - k)/k^\tau - d_\beta^2\lambda_{n,k}^*/k^\tau = o_p(1)$ because $\tau > \tau/2$. This, combined with $d_\beta^2\lambda_{n,k}^*/k^\tau \to d_{\beta_0}^2 r_\tau$, gives the desired result that $(\widehat{T}_n{}'\widehat{T}_n - k)/k^\tau \to_p d_{\beta_0}^2 r_\tau$. $\square$

Proof of Lemma 1. 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 $\alpha \neq 0$,

$$\begin{aligned}
\alpha' U_n &= \sum\sum_{1 \leq i < j \leq n} S_{nij} = \sum_{j=2}^n M_{nj} \to_d N(0, \alpha' V_3 \alpha),\ \text{where } M_{nj} = \sum_{i=1}^{j-1} S_{nij} \text{ and} \\
S_{nij} &= \frac{1}{nk^{1/2}}(2\alpha_1\xi_{ni}'\xi_{nj} + \alpha_2(\xi_{ni}'\eta_{nj} + \xi_{nj}'\eta_{ni}) + 2\alpha_3\eta_{ni}'\eta_{nj}). \qquad (5.11)
\end{aligned}$$

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

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

$$\begin{aligned}
H_n(x, x_*) &= \sum_{s=1}^3 \alpha_s H_{sn}(x, x_*),\ \text{where } x = (\xi', \eta')' \in R^{2k},\ x_* = (\xi_*', \eta_*')' \in R^{2k}, \\
H_{1n}(x, x_*) &= 2n^{-1}k^{-1/2}\xi'\xi_*,\ H_{2n}(x, x_*) = n^{-1}k^{-1/2}(\xi'\eta_* + \xi_*'\eta),\ \text{and} \\
H_{3n}(x, x_*) &= 2n^{-1}k^{-1/2}\eta'\eta_*. \qquad (5.12)
\end{aligned}$$

Note that $E(H_n(X_{n1}, X_{n2})|X_{n2}) = 0$ a.s. because X_{n1} and X_{n2} are independent with mean zero. In consequence, the U-statistic $\alpha' U_n$ in (5.11) is degenerate. Hall's Thm. 1 states that

$$\begin{aligned}
&\alpha' U_n / \left(n^2 E H_n^2(X_{n1}, X_{n2})/2\right)^{1/2} \to_d N(0,1) \text{ provided} \\
&\text{(I) } n^{-1} E H_n^4(X_{n1}, X_{n2}) / (E H_n^2(X_{n1}, X_{n2}))^2 \to 0 \text{ and} \\
&\text{(II) } E G_n^2(X_{n1}, X_{n2}) / (E H_n^2(X_{n1}, X_{n2}))^2 \to 0,\ \text{where} \\
&\quad G_n(x, x_*) = E H_n(X_{n1}, x) H_n(X_{n1}, x_*) \text{ for } x, x_* \in R^{2k}. \qquad (5.13)
\end{aligned}$$

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 (5.12). First, we have

$$n^2 EH_n^2(X_{n1}, X_{n2})/2 = \frac{1}{2k}\alpha' E \begin{pmatrix} 2\xi_{n1}'\xi_{n2} \\ \xi_{n1}'\eta_{n2} + \xi_{n2}'\eta_{n1} \\ 2\eta_{n1}'\eta_{n2} \end{pmatrix} \begin{pmatrix} 2\xi_{n1}'\xi_{n2} \\ \xi_{n1}'\eta_{n2} + \xi_{n2}'\eta_{n1} \\ 2\eta_{n1}'\eta_{n2} \end{pmatrix}' \alpha, \tag{5.14}$$

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

$$\begin{aligned}
&E(2\xi_{n1}'\xi_{n2})^2 = 4tr(E\xi_{n2}\xi_{n2}'\xi_{n1}\xi_{n1}') = 4tr(E\xi_{n2}\xi_{n2}' \cdot E\xi_{n1}\xi_{n1}') = 4k, \\
&E(\xi_{n1}'\eta_{n2} + \xi_{n2}'\eta_{n1})^2 = E(\xi_{n1}'\eta_{n2})^2 + 2E\xi_{n1}'\eta_{n2}\xi_{n2}'\eta_{n1} + E(\xi_{n2}'\eta_{n1})^2 \\
= {} & tr(E\xi_{n1}\xi_{n1}' \cdot E\eta_{n2}\eta_{n2}') + 2tr(E\eta_{n2}\xi_{n2}' \cdot E\eta_{n1}\xi_{n1}') + tr(E\xi_{n2}\xi_{n2}' \cdot E\eta_{n1}\eta_{n1}') = 2k, \\
&E(2\xi_{n1}'\xi_{n2})(\xi_{n1}'\eta_{n2} + \xi_{n2}'\eta_{n1}) = 2E\xi_{n1}'\xi_{n2}\xi_{n1}'\eta_{n2} + 2E\xi_{n1}'\xi_{n2}\xi_{n2}'\eta_{n1} \\
= {} & 2tr(E\xi_{n1}\xi_{n1}' \cdot E\xi_{n2}\eta_{n2}') + 2tr(E\xi_{n2}\xi_{n2}' \cdot E\eta_{n1}\xi_{n1}') = 0, \text{ and} \\
&E(2\xi_{n1}'\xi_{n2})(2\eta_{n1}'\eta_{n2}) = 4tr(E\xi_{n1}\eta_{n1}'\eta_{n2}\xi_{n2}') = 4tr(E\xi_{n1}\eta_{n1}' \cdot E\eta_{n2}\xi_{n2}') = 0,
\end{aligned} \tag{5.15}$$

using assumptions (ii) and (iv) of the Lemma. Likewise, we have $E(2\eta_{n1}'\eta_{n2})^2 = 4k$ and $E(2\eta_{n1}'\eta_{n2})(\xi_{n1}'\eta_{n2} + \xi_{n2}'\eta_{n1}) = 0$. Combining these results with (5.14 and (5.15) implies that

$$n^2 EH_n^2(X_{n1}, X_{n2})/2 = \alpha' V_3 \alpha > 0 \text{ for all } n, \tag{5.16}$$

which yields the asymptotic variance given in (5.11).

Now, to verify condition (I) of (5.13), we have

$$\begin{aligned}
EH_{1n}^4(X_{n1}, X_{n2}) &= \frac{16}{n^4k^2} E(\xi_{n1}'\xi_{n2})^4 = \frac{16}{n^4k^2} E(\sum_{\ell=1}^{k} \xi_{n1\ell}\xi_{n2\ell})^4 \\
&= \frac{16}{n^4k^2} \sum_{\ell_1=1}^{k}\sum_{\ell_2=1}^{k}\sum_{\ell_3=1}^{k}\sum_{\ell_4=1}^{k} E\xi_{n1\ell_1}\xi_{n2\ell_1}\xi_{n1\ell_2}\xi_{n2\ell_2}\xi_{n1\ell_3}\xi_{n2\ell_3}\xi_{n1\ell_4}\xi_{n2\ell_4} \\
&\le \frac{16k^2}{n^4} \sup_{\ell_1,\ell_2,\ell_3,\ell_4 \le 4, n\ge 1} E\xi_{n1\ell_1}\xi_{n1\ell_2}\xi_{n1\ell_3}\xi_{n1\ell_4} \cdot E\xi_{n2\ell_1}\xi_{n2\ell_2}\xi_{n2\ell_3}\xi_{n2\ell_4} = O(\frac{k^2}{n^4}),
\end{aligned} \tag{5.17}$$

where $\xi_{ni} = (\xi_{ni1}, ..., \xi_{nik})'$ 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 $EH_{sn}^4(X_{n1}, X_{n2}) = O(k^2/n^4)$ for $s = 2, 3$. These results and Minkowski's inequality then give $EH_n^4(X_{n1}, X_{n2}) = O(k^2/n^4)$. Combining this with (5.16) establishes condition (I) of (5.13).provided $n^{-1}k^2 \to 0$, which holds by assumption (vii) of the Lemma.

To verify condition (II) of (5.13), by the Cauchy-Schwartz inequality, it suffices to verify condition (II) with $EG_n^2(X_{ni}, X_{nj})$ replaced by $EG_{sn}^2(X_{ni}, X_{nj})$ for $s = 1, 2, 3$, where $G_{sn}(\cdot, \cdot)$ is defined as $G_n(\cdot, \cdot)$ is defined in (5.13), but with $H_{sn}(\cdot, \cdot)$ in place of $H_n(\cdot, \cdot)$. We have

$$\begin{aligned}
G_{1n}(x, x_*) &= EH_{1n}(X_{n1}, x)H_{1n}(X_{n1}, x_*) = \frac{4}{n^2k} E\xi_{n1}'\xi\xi_{n1}'\xi_* = \frac{4}{n^2k}\xi'\xi_*, \\
G_{2n}(x, x_*) &= EH_{2n}(X_{n1}, x)H_{2n}(X_{n1}, x_*) \\
&= \frac{1}{n^2k} E(\xi_{n1}'\eta + \xi'\eta_{n1})(\xi_{n1}'\eta_* + \xi_*'\eta_{n1}) = \frac{1}{n^2k}(\xi'\xi_* + \eta'\eta_*),
\end{aligned} \tag{5.18}$$

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

$$\begin{aligned} EG_{1n}^2(X_{n1}, X_{n2}) &= \frac{16}{n^4k^2}E(\xi_{n1}'\xi_{n2})^2 = \frac{16}{n^4k^2}tr(E\xi_{n1}\xi_{n1}' \cdot E\xi_{n2}\xi_{n2}') = \frac{16}{n^4k}, \\ EG_{2n}^2(X_{n1}, X_{n2}) &= \frac{1}{n^4k^2}E(\xi_{n1}'\xi_{n2} + \eta_{n1}'\eta_{n2})^2 = \frac{2}{n^4k}. \end{aligned} \tag{5.19}$$

Similarly, $EG_{3n}^2(X_{n1}, X_{n2}) = 16/(n^4k)$. Combining (5.16), (5.18), and (5.19) yields condition (II) of (5.13) 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

$$\begin{aligned} &E(\xi_{n1}'\xi_{n1})^2 = \sum_{\ell_1=1}^{k}\sum_{\ell_2=1}^{k} E\xi_{n1\ell_1}^2\xi_{n1\ell_2}^2 \leq k^2 \sup_{\ell \leq k, n \geq 1} E\xi_{n1\ell}^4 = O(k^2), \qquad (5.20) \\ &Var(n^{-1}k^{-1/2}\sum_{i=1}^{n}(\xi_{ni}'\xi_{ni} - k)) = n^{-1}k^{-1}Var((\xi_{ni}'\xi_{ni} - k)) = n^{-1}O(k) = o(1) \end{aligned}$$

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

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 2. To prove part (a), we write

$$\begin{aligned} &\frac{2}{nk^{1/2}} \sum\sum_{1 \leq i < j \leq n} \xi_{ni}'\xi_{nj} = A_{1n} + A_{2n} + A_{3n}, \text{ where} \\ &A_{1n} = \frac{2}{nk^{1/2}} \sum\sum_{1 \leq i < j \leq n} (\xi_{ni} - \mu_{n\xi})'(\xi_{nj} - \mu_{n\xi}), \\ &A_{2n} = \frac{2}{nk^{1/2}} \sum\sum_{1 \leq i < j \leq n} [\mu_{n\xi}'(\xi_{ni} - \mu_{n\xi}) + \mu_{n\xi}'(\xi_{nj} - \mu_{n\xi})], \text{ and} \\ &A_{3n} = \frac{2}{nk^{1/2}} \sum\sum_{1 \leq i < j \leq n} \mu_{n\xi}'\mu_{n\xi}. \end{aligned} \tag{5.21}$$

Now, some calculations yield

$$\begin{aligned} A_{3n} &= \frac{n(n-1)}{nk^{1/2}}\mu_{n\xi}'\mu_{n\xi} = \frac{1}{k^{1/2}}\lambda_{n\xi} - \frac{1}{nk^{1/2}}\lambda_{n\xi} = \frac{1}{k^{1/2}}\lambda_{n\xi} + o(1), \\ A_{2n} &= \frac{2(n-1)}{nk^{1/2}}\mu_{n\xi}'\sum_{i=1}^{n}(\xi_{ni} - \mu_{n\xi}), \ EA_{2n} = 0, \text{ and} \\ Var(A_{2n}) &= \frac{4(n-1)^2}{nk}\mu_{n\xi}'Var(\xi_{ni})\mu_{n\xi} = \frac{4(n-1)^2}{n^2k}\lambda_{n\xi} = o(1), \end{aligned} \tag{5.22}$$

using $\lambda_{n\xi}/k \to 0$ and $k^2/n \to 0$. Combining (5.21) and (5.22) gives

$$\frac{2}{nk^{1/2}} \sum\sum_{1 \le i < j \le n} \xi'_{ni}\xi_{nj} - \frac{1}{k^{1/2}}\lambda_{n\xi} = A_{1n} + o_p(1). \tag{5.23}$$

Similar calculations yield

$$\begin{aligned} &\frac{1}{nk^{1/2}} \sum\sum_{1 \le i < j \le n} (\xi'_{ni}\eta_{nj} + \xi'_{nj}\eta_{ni}) - \frac{1}{k^{1/2}}\lambda_{n\xi\eta} \qquad (5.24)\\ &= \frac{1}{nk^{1/2}} \sum\sum_{1 \le i < j \le n} [(\xi_{ni} - \mu_{n\xi})'(\eta_{nj} - \mu_{n\eta}) + (\xi_{nj} - \mu_{n\xi})'(\eta_{ni} - \mu_{n\eta})] + o_p(1) \text{ and}\\ &\frac{2}{nk^{1/2}} \sum\sum_{1 \le i < j \le n} \eta'_{ni}\eta_{nj} - \frac{1}{k^{1/2}}\lambda_{n\eta} = \frac{1}{nk^{1/2}} \sum\sum_{1 \le i < j \le n} (\eta_{ni} - \mu_{n\eta})'(\eta_{nj} - \mu_{n\eta}) + o_p(1). \end{aligned}$$

Stacking the results of (5.23) and (5.24) and applying Lemma 1(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

$$\begin{aligned} &\frac{1}{nk^{1/2}} \sum_{i=1}^{n}(\xi'_{ni}\xi_{ni} - k) = F_{1n} + F_{2n} + F_{3n}, \text{ where}\\ &F_{1n} = \frac{1}{nk^{1/2}} \sum_{i=1}^{n}[(\xi_{ni} - \mu_{n\xi})'(\xi_{ni} - \mu_{n\xi}) - k], \ F_{2n} = \frac{2}{nk^{1/2}} \sum_{i=1}^{n} \mu'_{n\xi}(\xi_{ni} - \mu_{n\xi}),\\ &F_{3n} = \frac{1}{nk^{1/2}} \sum_{i=1}^{n} \mu'_{n\xi}\mu_{n\xi}. \end{aligned} \tag{5.25}$$

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

$$\begin{aligned} F_{3n} &= \frac{1}{nk^{1/2}}\lambda_{n\xi} = \frac{k^{1/2}}{n}\frac{\lambda_{n\xi}}{k} \to 0, \ EF_{2n} = 0, \text{ and}\\ Var(F_{2n}) &= \frac{4}{nk}\mu'_{n\xi}Var(\xi_{ni})\mu_{n\xi} = \frac{4}{n^2k}\lambda_{n\xi} \to 0 \end{aligned} \tag{5.26}$$

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 3. Lemma 2(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

$$\delta_n = A'_n(\Sigma_{n\xi}^{-1} - \widehat{\Sigma}_{n\xi}^{-1})A_n = o_p(k^{1/2}), \text{ where } A_n = n^{-1/2}\sum_{i=1}^{n}\xi_{ni}, \tag{5.27}$$

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

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

$$A_n'\Sigma_{n\xi}^{-1}A_n = O_p(k) \tag{5.28}$$

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

$$\lambda_{\min}^{-1}(\widehat{\Sigma}_{n\xi}) = O_p(1) \tag{5.29}$$

because $|\lambda_{\min}(\widehat{\Sigma}_{n\xi}) - \lambda_{\min}(\Sigma_{n\xi})| \leq ||\widehat{\Sigma}_{n\xi} - \Sigma_{n\xi}|| = o_p(1)$ by assumption (iv)$'$ and $\lambda_{\min}^{-1}(\Sigma_{n\xi}) = O(1)$ 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) $||HGH|| \leq \lambda_{\max}^2(H)||G||$, (b) $||Hc|| \leq \lambda_{\max}(H)||c|| \leq ||H|| \cdot ||c||$, (c) $c'Gc \leq ||G|| \cdot ||c||^2$, 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

$$\begin{aligned}
\delta_n &= A_n'(C_n^{-2} - D_n^{-2})A_n \\
&= A_n'C_n^{-1}(I_k - C_nD_n^{-2}C_n)C_n^{-1}A_n \\
&= A_n'C_n^{-1}(C_n^{-1}D_n^2C_n^{-1} - I_k)C_n^{-1}A_n \\
&\quad -A_n'C_n^{-1}(C_n^{-1}D_n^2C_n^{-1} - I_k)C_nD_n^{-2}C_n(C_n^{-1}D_n^2C_n^{-1} - I_k)C_n^{-1}A_n \\
&\leq A_n'C_n^{-1}[C_n^{-1}(D_n^2 - C_n^2)C_n^{-1}]C_n^{-1}A_n \\
&\quad +||C_n(C_n^{-1}D_n^2C_n^{-1} - I_k)C_n^{-1}A_n||^2 \cdot \lambda_{\max}^2(D_n^{-1}) \\
&\leq ||C_n^{-1}(D_n^2 - C_n^2)C_n^{-1}|| \cdot ||C_n^{-1}A_n||^2 \\
&\quad +||(D_n^2 - C_n^2)C_n^{-1}C_n^{-1}A_n||^2 \cdot \lambda_{\min}^{-2}(D_n) \\
&\leq ||D_n^2 - C_n^2|| \cdot \lambda_{\max}^2(C_n^{-1}) \cdot ||C_n^{-1}A_n||^2 \\
&\quad\ +||D_n^2 - C_n^2||^2||C_n^{-1}C_n^{-1}A_n||^2 \cdot \lambda_{\min}^{-2}(D_n) \\
&\leq ||D_n^2 - C_n^2|| \cdot \lambda_{\min}^{-1}(C_n^2) \cdot ||C_n^{-1}A_n||^2 \\
&\quad\ +||D_n^2 - C_n^2||^2 \cdot ||C_n^{-1}A_n||^2 \cdot \lambda_{\max}^2(C_n^{-1}) \cdot \lambda_{\min}^{-2}(D_n) \\
&= o_p(k^{-1/2})O(1)O_p(k) + o_p(k^{-1})O_p(k)O(1)O_p(1) \\
&= o_p(k^{1/2}),
\end{aligned} \tag{5.30}$$

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), (5.28), and (5.29). This establishes (5.27).

The same argument holds with $(\xi_{ni}, \widehat{\Sigma}_{n\xi})$ replaced by $(\eta_{ni}, \widehat{\Sigma}_{n\eta})$. Hence, (5.27) holds and the Lemma is proved. $\square$

Proof of Lemma 4. It suffices to show $\Delta_n = o_p(k^{1/2})$ and an analogous result with (ξ_{n1i}, ξ_{n2i}) replaced by (η_{n1i}, η_{n2i}), where Δ_n is defined by

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

$$
\begin{aligned}
\Delta_n &= |(G_n + \widehat{D}_{n\xi}H_n)'\widehat{\Sigma}_{n\xi}^{-1}(G_n + \widehat{D}_{n\xi}H_n) - (G_n + D_{n\xi}H_n)'\widehat{\Sigma}_{n\xi}^{-1}(G_n + D_{n\xi}H_n)| \\
&= |H_n'(\widehat{D}_{n\xi} - D_{n\xi})'\widehat{\Sigma}_{n\xi}^{-1}(\widehat{D}_{n\xi} - D_{n\xi})H_n + 2H_n'(\widehat{D}_{n\xi} - D_{n\xi})'\widehat{\Sigma}_{n\xi}^{-1}(G_n + D_{n\xi}H_n)| \\
&\leq P_{1n} + 2P_{1n}^{1/2}P_{2n}^{1/2}, \\
P_{1n} &= ||\widehat{\Sigma}_{n\xi}^{-1/2}(\widehat{D}_{n\xi} - D_{n\xi})H_n||^2,\ P_{2n} = (G_n + D_{n\xi}H_n)'\widehat{\Sigma}_{n\xi}^{-1}(G_n + D_{n\xi}H_n), \qquad (5.31)
\end{aligned}
$$

and the inequality holds by the Cauchy-Schwarz inequality. We have $P_{2n} = O_p(k)$ by Lemma 3. Hence, the Lemma holds if $P_{1n} = o_p(1)$.

We have

$$
P_{1n} \leq \lambda_{\max}^2(\widehat{\Sigma}_{n\xi}^{-1/2}) \cdot ||(\widehat{D}_{n\xi} - D_{n\xi})H_n||^2 \leq \lambda_{\min}^{-1}(\widehat{\Sigma}_{n\xi}) \cdot ||\widehat{D}_{n\xi} - D_{n\xi}||^2||H_n||^2, \quad (5.32)
$$

where the two inequalities hold by inequality (b) stated following (5.29) above.

Next, we have: (I) $||H_n||^2 = O_p(k^2)$ because $||H_n|| \leq ||H_n - EH_n|| + ||EH_n||$, $E||H_n - EH_n||^2 = E(\xi_{n2i} - E\xi_{n2i})'(\xi_{n2i} - E\xi_{n2i}) = O(k)$, which implies that $||H_n - EH_n||^2 = O_p(k)$, and $||EH_n||^2 = ||n^{1/2}E\xi_{n2i}||^2 = O(k^2)$ by assumption (c) of the Lemma, (II) $\lambda_{\max}(\widehat{\Sigma}_{n\xi}^{-1/2}) = \lambda_{\min}^{-1/2}(\widehat{\Sigma}_{n\xi}) = O_p(1)$ by (5.29) above, and (III) $||\widehat{D}_{n\xi} - D_{n\xi}|| = o_p(k^{-1})$ by assumption (d) of the Lemma. Hence, $P_{1n} = o_p(1)$ and $\Delta_n = o_p(k^{1/2})$.

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

Proof of Lemma 5. The result is established by applying Brown's (1971) martingale CLT to a linear combination of the 5-vector in the Lemma and then applying the Cramér-Wold device. For the former, it suffices to show that for any $\alpha_* = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)' \neq 0$,

$$
\sum_{j=1}^{n}(M_{nj} + G_{nj}) \to_d N(0, \alpha_*'V_5\alpha_*), \text{ where } G_{nj} = k^{-\tau/2}h_n'(\alpha_4\eta_{nj} + \alpha_5\xi_{nj}), \quad (5.33)
$$

M_{nj} and α are defined in (5.11), and $M_{n1} = 0$. Note that $E(M_{nj} + G_{nj}|\mathcal{F}_{j-1}) = 0$ a.s., where $\mathcal{F}_{j-1} = \sigma(\xi_{nj-1}, \eta_{nj-1}, ..., \xi_{n1}, \eta_{n1})$. Brown's CLT requires the verification of (i) the convergence of $Var(\sum_{j=1}^n(M_{nj} + G_{nj}))$ to $\alpha_*'V_5\alpha_*$, (ii) a Lindeberg condition, and (iii) a conditional variance condition. For (i), we have

$$
Var(\sum_{j=1}^{n}(M_{nj}+G_{nj})) = \sum_{j=1}^{n}(EM_{nj}^2+EG_{nj}^2) \to \alpha'V_3\alpha+m(\alpha_4^2+\alpha_5^2) = \alpha_*'V_5\alpha_*, \quad (5.34)
$$

where the equality uses $EM_{nj} = EG_{nj} = EM_{nj}G_{nj} = 0$ and the convergence holds by (5.14)-(5.16) and $EG_{nj}^2 = nk^{-\tau}h_n'h_n(\alpha_4^2 + \alpha_5^2) \to m(\alpha_4^2 + \alpha_5^2)$.

The Lindeberg condition is implied by $\sum_{j=1}^n EM_{nj}^4 \to 0$ and $\sum_{j=1}^n EG_{nj}^4 \to 0$. The former holds by condition (I) of (5.13) and the proof of Hall's (1984) Thm. 1. The latter holds because

$$
\begin{aligned}
\sum_{j=1}^{n} EG_{nj}^4 &= \sum_{j=1}^{n} E[k^{-\tau/2}h_n'(\alpha_4\eta_{nj} + \alpha_5\xi_{nj})]^4 \\
&\leq n^{-1}(nh_n'h_n/k^\tau)^2E||\alpha_4\eta_{nj} + \alpha_5\xi_{nj}||^4 \to 0, \qquad (5.35)
\end{aligned}
$$

using the Cauchy-Schwarz inequality, $nk^{-\tau}h_n'h_n \to m$, and $E||\eta_{nj}||^4 + E||\xi_{nj}||^4 < \infty$.

The conditional variance condition holds if

$$\sum_{j=1}^{n} E((M_{nj} + G_{nj})^2|\mathcal{F}_{j-1}) \to_p \alpha_*'V_5\alpha_*. \tag{5.36}$$

We have $\sum_{j=1}^{n} E(M_{nj}^2|\mathcal{F}_{j-1}) \to_p \alpha'V_3\alpha$ by condition (II) of (5.13) and the proof of Hall's (1984) Thm. 1. In addition, $E(G_{nj}^2|\mathcal{F}_{j-1}) = EG_{nj}^2$ a.s. and $\sum_{j=1}^{n} EG_{nj}^2 \to m(\alpha_4^2 + \alpha_5^2)$, as above. Hence, it remains to show that $\sum_{j=1}^{n} E(M_{nj}G_{nj}|\mathcal{F}_{j-1}) \to_p 0$. We have

$$\begin{aligned}
&E(M_{nj}k^{-\tau/2}h_n'\eta_{nj}|\mathcal{F}_{j-1})\\
&= n^{-1}k^{-1/2-\tau/2}E\Big(\sum_{i=1}^{j-1}(2\alpha_1\xi_{ni}'\xi_{nj} + \alpha_2(\xi_{ni}'\eta_{nj} + \xi_{nj}'\eta_{ni}) + 2\alpha_3\eta_{ni}'\eta_{nj})h_n'\eta_{nj}|\mathcal{F}_{j-1}\Big)\\
&= n^{-1}k^{-1/2-\tau/2}\sum_{i=1}^{j-1}(\alpha_2\xi_{ni}'h_n + 2\alpha_3\eta_{ni}'h_n).
\end{aligned} \tag{5.37}$$

Next, we have

$$\begin{aligned}
&E\left(\sum_{j=1}^{n} n^{-1}k^{-1/2-\tau/2}\sum_{i=1}^{j-1}\xi_{ni}'h_n\right)^2 = n^{-2}k^{-1-\tau}E\left(\sum_{i=1}^{n}(n-i)\xi_{ni}'h_n\right)^2\\
&= n^{-3}k^{-1}(nh_n'h_n/k^\tau)\left(\sum_{i=1}^{n}(n-i)\right)^2 = O(n^{-1}k^{-1}) = o(1).
\end{aligned} \tag{5.38}$$

The same holds with ξ_{ni} replaced by η_{ni}. These results, combined with (5.37) and Markov's inequality, gives $\sum_{j=1}^{n} E(M_{nj}k^{-\tau/2}h_n'\eta_{nj}|\mathcal{F}_{j-1}) \to_p 0$. Analogously, $\sum_{j=1}^{n} E(M_{nj}k^{-\tau/2}h_n'\xi_{nj}|\mathcal{F}_{j-1}) \to_p 0$. In consequence, $\sum_{j=1}^{n} E(M_{nj}G_{nj}|\mathcal{F}_{j-1}) \to_p 0$ and the proof is complete. □

Proof of Lemma 6. First, we consider the case where $\widehat{\Sigma}_{n\xi} = \Sigma_{n\xi} = I_k$, $\widehat{\Sigma}_{n\eta} = \Sigma_{n\eta} = I_k$, $\widehat{D}_{n\xi} = D_{n\xi}$, and $\widehat{D}_{n\eta} = D_{n\eta}$, which is analogous to the situation in Lemma 2. We start with the proof of part (a). Because $\lambda_{n\xi}/k \to 0$, just as in Lemma 2, (5.23) still holds. Next, we write

$$\begin{aligned}
&\frac{1}{nk^{\tau/2}}\sum\sum_{1\le i<j\le n}[\xi_{ni}'\eta_{nj} + \xi_{nj}'\eta_{ni}] = B_{1n} + B_{2n} + B_{3n} + B_{4n}, \text{ where}\\
&B_{1n} = \frac{1}{nk^{\tau/2}}\sum\sum_{1\le i<j\le n}[(\xi_{ni} - \mu_{n\xi})'(\eta_{nj} - \mu_{n\eta}) + (\xi_{nj} - \mu_{n\xi})'(\eta_{ni} - \mu_{n\eta})],\\
&B_{2n} = \frac{1}{nk^{\tau/2}}\sum\sum_{1\le i<j\le n}[\mu_{n\eta}'(\xi_{ni} - \mu_{n\xi}) + \mu_{n\eta}'(\xi_{nj} - \mu_{n\xi})] = \frac{n-1}{nk^{\tau/2}}\sum_{i=1}^{n}\mu_{n\eta}'(\xi_{ni} - \mu_{n\xi}),
\end{aligned}$$

$$B_{3n} = \frac{1}{nk^{\tau/2}} \sum\sum_{1 \le i < j \le n} [\mu'_{n\xi}(\eta_{ni} - \mu_{n\eta}) + \mu'_{n\xi}(\eta_{nj} - \mu_{n\eta})] = \frac{n-1}{nk^{\tau/2}} \sum_{i=1}^{n} \mu'_{n\xi}(\eta_{ni} - \mu_{n\eta}),$$

$$B_{4n} = \frac{2}{nk^{\tau/2}} \sum\sum_{1 \le i < j \le n} \mu'_{n\xi}\mu_{n\eta} = \frac{n-1}{n}\lambda_{n\xi\eta}/k^{\tau/2}. \tag{5.39}$$

We have: $EB_{3n} = 0$, $Var(B_{3n}) = ((n-1)/n)^2\lambda_{n\xi}/k^{\tau} \to 0$, and hence $B_{3n} = o_p(1)$. Thus,

$$\frac{1}{nk^{1/2}} \sum\sum_{1 \le i < j \le n} [\xi'_{ni}\eta_{nj} + \xi'_{nj}\eta_{ni}] - \lambda_{n\xi\eta}/k^{\tau/2} = B_{1n} + B_{2n} + o_p(1). \tag{5.40}$$

When $\tau > 1$, $B_{1n} = o_p(1)$, because Lemma 1(a) implies that $k^{\tau/2-1/2}B_{1n} = O_p(1)$.

As in (5.21)-(5.23), but with $(\eta, k^{\tau/2})$ in place of $(\xi, k^{1/2})$, we can write

$$\begin{aligned}
&\frac{2}{nk^{\tau/2}} \sum\sum_{1 \le i < j \le n} \eta'_{ni}\eta_{nj} - \frac{1}{k^{\tau/2}}\lambda_{n\xi} = C_{1n} + C_{2n} + o_p(1), \text{ where} \\
&C_{1n} = \frac{2}{nk^{\tau/2}} \sum\sum_{1 \le i < j \le n} (\eta_{ni} - \mu_{n\eta})'(\eta_{nj} - \mu_{n\eta}) \text{ and} \\
&C_{2n} = \frac{2(n-1)}{nk^{\tau/2}}\mu'_{n\eta} \sum_{i=1}^{n} (\eta_{ni} - \mu_{n\eta}).
\end{aligned} \tag{5.41}$$

In the present case C_{2n} is not $o_p(1)$ because $Var(C_{2n}) = 4((n-1)/n)^2 k^{-\tau}\lambda_{n\eta} \to 4\kappa_{\eta\tau}$. When $\tau > 1$, $C_{1n} = o_p(1)$, because Lemma 1(a) implies that $k^{\tau/2-1/2}C_{1n} = O_p(1)$.

Now, we stack the results of (5.23), (5.40), and (5.41) and apply Lemma 5 to the rhs of these stacked equations. When $\tau = 1$, the rhs satisfies $(A_{1n}, B_{1n} + B_{2n}, C_{1n} + C_{2n})' + o_p(1)$ and we apply Lemma 5 with (ξ_{ni}, η_{ni}) equal to $(\xi_{ni} - \mu_{n\xi}, \eta_{ni} - \mu_{n\eta})$ of the present Lemma, with $h_n = ((n-1)/n)\mu_{n\eta}$, and with $m = \kappa_{\eta 1}$. Then, Lemma 5 gives $(A_{1n}, B_{1n} + B_{2n}, C_{1n} + C_{2n})' \to_d N(0, Diag\{2, 1 + \kappa_{\eta 1}, 2 + 4\kappa_{\eta 1}\}) = N(0, V^*_{3,1})$, which is the result of part (a). When $\tau > 1$, the rhs of the stacked equations satisfies $(A_{1n}, B_{2n}, C_{2n})' + o_p(1)$ and we apply Lemma 5 in the same way as above to obtain $(A_{1n}, B_{2n}, C_{2n})' \to_d N(0, Diag\{2, \kappa_{\eta\tau}, 4\kappa_{\eta\tau}\}) = N(0, V^*_{3,\tau})$, as desired. This completes the proof of part (a) for the case where $\widehat{\Sigma}_{n\xi} = \Sigma_{n\xi}$, etc.

Next, we prove parts (b) and (c) for the case where $\widehat{\Sigma}_{n\xi} = \Sigma_{n\xi} = I_k$, etc. The proof that $(nk^{1/2})^{-1}\sum_{i=1}^{n} \xi'_{ni}\xi_{ni} = o_p(1)$ is exactly the same as in (5.25)-(5.26) because $\lambda_{n\xi}/k \to 0$. The proof that $(nk^{\tau/2})^{-1}\sum_{i=1}^{n}(\eta'_{ni}\eta_{ni} - k) = o_p(1)$ is analogous to that given in (5.25)-(5.26) with ξ replaced by η, with $(nk^{\tau/2})^{-1}$ in place of $(nk^{1/2})^{-1}$, and with $\lambda_{n\eta}/k^{\tau} = O(1)$ rather than $\lambda_{n\xi}/k = o(1)$. In consequence, $F_{3n} = (nk^{\tau/2})^{-1}\lambda_{n\eta} = (k^{\tau/2}/n)(\lambda_{n\eta}/k^{\tau}) \to 0$ provided $k^3/n \to 0$ and $\tau < 6$. We have $EF_{2n} = 0$ and $Var(F_{2n}) = 4n^{-2}(\lambda_{n\eta}/k^{\tau}) \to 0$. In addition, $F_{1n} \to_p 0$ by Lemma 1(b). Hence, $(nk^{\tau/2})^{-1}\sum_{i=1}^{n}(\eta'_{ni}\eta_{ni} - k) \to_p 0$. Similar calculations, using the fact that $\lambda^2_{n\xi\eta} \le \lambda_{n\xi}\lambda_{n\eta}$ by the Cauchy-Schwarz inequality and hence $\lambda_{n\xi\eta}/k^{(1+\tau)/2} \le (\lambda_{n\xi}/k)^{1/2}(\lambda_{n\eta}/k^{\tau})^{1/2} = o(1)$, show that $(nk^{\tau/2})^{-1}\sum_{i=1}^{n} \xi'_{ni}\eta_{ni} = o_p(1)$, 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).

We complete the proof by showing that the same asymptotic distributions arise when $\widehat{\Sigma}_{n\xi} \neq \Sigma_{n\xi}$, $\Sigma_{n\xi} \neq I_k$, $\widehat{D}_{n\xi} \neq D_{n\xi}$, etc. as when $\widehat{\Sigma}_{n\xi} = \Sigma_{n\xi}$, etc. using the arguments in the proofs of Lemmas 3 and 4. In these proofs, exactly the same arguments hold under the assumptions of the current Lemma when δ_n and Δ_n are based on ξ_{ni}, because the normalization is by $k^{-1/2}$ and $||E\xi_{n2i}||^2 = O(k^2/n)$. When δ_n and Δ_n are based on η_{ni}, the arguments need to be altered because the normalization is by $k^{-\tau/2}$, not $k^{-1/2}$, and $||E\eta_{n2i}||^2 = O(k^{1+\tau}/n)$, not $||E\eta_{n2i}||^2 = O(k^2/n)$, in the second and third elements of the 3-vector in part (c) of the present Lemma. In particular, in the proof of Lemma 3, it suffices to show that $\delta_n = o_p(k^{\tau/2})$ when δ_n is defined with η in place of ξ. Nevertheless, the proof of Lemma 3 goes through without change to show that $\delta_n = o_p(k^{1/2})$, which is stronger than necessary. (This relies on the fact that $A_n'\Sigma_{n\eta}^{-1}A_n = ||C_n^{-1}A_n||^2 = O_p(k)$, due to the centering by k, by the result proved above for the case $\widehat{\Sigma}_{n\xi} = \Sigma_{n\xi}$, etc.) In the proof of Lemma 4, it suffices to show that $\Delta_n = o_p(k^{\tau/2})$ when δ_n is defined with η in place of ξ. When $\tau = 1$, the same proof goes through without change. But, when $\tau > 1$, the proof needs to be altered. In the present case, we still have $P_{2n} = O_p(k)$ (due to centering at k), but $||EH_n||^2 = ||n^{1/2}E\eta_{n2i}||^2 = O(k^{1+\tau})$ by the assumption of the present Lemma, rather than $O(k^2)$. In consequence, using (5.32), $P_{1n} = o_p(k^{-2})O_p(k^{1+\tau}) = o_p(k^{\tau-1})$. Using $P_{2n} = O_p(k)$, this leads to $P_{1n}^{1/2}P_{2n}^{1/2} = o_p(k^{\tau/2})$. Also, if $\tau \leq 2$, this gives $P_{1n} = o_p(k^{\tau/2})$. These results and (5.31) combine to give $\Delta_n = o_p(k^{\tau/2})$ if $\tau \leq 2$. $\square$

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

$$\begin{aligned}
&P(k||n^{-1}\widetilde{Z}'\widetilde{Z} - E\widetilde{Z}_i\widetilde{Z}_i'||^2 > \varepsilon) \\
&\leq kEtr\left(\left(n^{-1}\sum_{i=1}^{n}\widetilde{Z}_i\widetilde{Z}_i' - E\widetilde{Z}_1\widetilde{Z}_1'\right)\left(n^{-1}\sum_{j=1}^{n}\widetilde{Z}_j\widetilde{Z}_j' - E\widetilde{Z}_1\widetilde{Z}_1'\right)\right)/\varepsilon \\
&= k \cdot tr\left(n^{-1}E\left(\widetilde{Z}_2\widetilde{Z}_2' - E\widetilde{Z}_1\widetilde{Z}_1'\right)\left(\widetilde{Z}_2\widetilde{Z}_2' - E\widetilde{Z}_1\widetilde{Z}_1'\right)\right)/\varepsilon \\
&= kn^{-1}\left(E(\widetilde{Z}_2'\widetilde{Z}_2)^2 - 2E(\widetilde{Z}_2'\widetilde{Z}_1)^2 + tr\left([E\widetilde{Z}_1\widetilde{Z}_1']E\widetilde{Z}_1\widetilde{Z}_1'\right)\right)/\varepsilon \\
&\leq O(k^3/n) = o(1),
\end{aligned} \tag{5.42}$$

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 $\sup_{j\leq k;n\geq 1} E\widetilde{Z}_{ij}^4 < \infty$ by Assumption 2, and the third equality holds by Assumption 4.

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

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

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

$$\begin{aligned}
&P(k^2||n^{-1}X'\widetilde{Z} - EX_i\widetilde{Z}_i'||^2 > \varepsilon) \\
&\leq k^2Etr((n^{-1}\sum_{i=1}^{n}X_i\widetilde{Z}_i' - EX_i\widetilde{Z}_i')'n^{-1}\sum_{j=1}^{n}X_j\widetilde{Z}_j' - EX_j\widetilde{Z}_j'))/\varepsilon
\end{aligned}$$

$$= (k^2/n)tr((E(X_i\widetilde{Z}_i' - EX_i\widetilde{Z}_i')'(X_i\widetilde{Z}_i' - EX_i\widetilde{Z}_i'))/\varepsilon$$
$$\leq (k^3/n)p \sup_{j\leq k, n\geq 1} E||X_i\widetilde{Z}_{ij}||^2 = o(1), \tag{5.43}$$

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

$$\begin{aligned} n^{-1}Z'Z &= n^{-1}\widetilde{Z}'\widetilde{Z} - n^{-1}\widetilde{Z}'X(X'X)^{-1}X'\widetilde{Z} \text{ and} \\ EZ_i^*Z_i^{*\prime} &= E\widetilde{Z}_i\widetilde{Z}_i' - E\widetilde{Z}_iX_i'(EX_iX_i')^{-1}EX_i\widetilde{Z}_i'. \end{aligned} \tag{5.44}$$

By the triangle inequality, we have

$$\begin{aligned} &||n^{-1}\widetilde{Z}'X(X'X)^{-1}X'\widetilde{Z} - E\widetilde{Z}_iX_i'(EX_iX_i')^{-1}EX_i\widetilde{Z}_i'|| \leq L_{n1} + L_{n2} + L_{n3}, \text{ where} \\ &L_{n1} = ||n^{-1}\widetilde{Z}'X(n^{-1}X'X)^{-1}(n^{-1}X'\widetilde{Z} - EX_i\widetilde{Z}_i')||, \\ &L_{n2} = ||n^{-1}\widetilde{Z}'X[(n^{-1}X'X)^{-1} - (EX_iX_i')^{-1}]EX_i\widetilde{Z}_i'||, \text{ and} \\ &L_{n3} = ||(n^{-1}\widetilde{Z}'X - E\widetilde{Z}_iX_i')(EX_iX_i')^{-1}EX_i\widetilde{Z}_i'||. \end{aligned} \tag{5.45}$$

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

$$||n^{-1}\widetilde{Z}'X|| \leq ||n^{-1}\widetilde{Z}'X - E\widetilde{Z}_iX_i'|| + ||E\widetilde{Z}_iX_i'|| = o_p(k^{-1}) + O(k^{1/2}) = O_p(k^{1/2}). \tag{5.46}$$

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

$$\begin{aligned} &L_{n1} \leq ||n^{-1}\widetilde{Z}'X|| \cdot ||(n^{-1}X'X)^{-1}|| \cdot ||n^{-1}X'\widetilde{Z} - EX_i\widetilde{Z}_i'|| \\ &= O_p(k^{1/2})O_p(1)o_p(k^{-1}) = o_p(k^{-1/2}). \end{aligned} \tag{5.47}$$

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

Using the results of (5.46) and parts (b) and (c), we have

$$\begin{aligned} L_{n2} &\leq ||n^{-1}\widetilde{Z}'X|| \cdot ||(n^{-1}X'X)^{-1} - (EX_iX_i')^{-1}|| \cdot ||EX_i\widetilde{Z}_i'|| \\ &= O_p(k^{1/2})O_p(n^{-1/2})O(k^{1/2}) = O_p((k^3/n)^{1/2}k^{-1/2}) = o_p(k^{-1/2}). \end{aligned} \tag{5.48}$$

Hence, the left-hand side in (5.45) is $o_p(k^{-1/2})$. This, (5.44), and part (a) combine to establish part (e). $\square$

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,* Ser. B, 38, 1-36.

Anderson, T. W., N. Kunitomo, and Y. Matsushita (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.

Anderson, T. W. and H. Rubin (1949): "Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations," *Annals of Mathematical Statistics,* 21, 570-582.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2004): "Optimal Invariant Similar Tests for Instrumental Variables Regression with Weak Instruments," Cowles Foundation Discussion Paper No. 1476, Yale University. Available at http://cowles.econ.yale.edu.

Andrews, D. W. K. and J. H. Stock (2005a): "Inference with Weak Instruments," World Congress of the Econometric Society invited paper, Cowles Foundation, Yale University.

——— (2005b): "Inference with Weak Instruments," Cowles Foundation Discussion Paper No. 1530, Yale University.

Bekker, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica,* 62, 657-681.

Brown, B. M. (1971): "Martingale Central Limit Theorems," *Annals of Mathematical Statistics,* 42, 59-66.

Chamberlain, G. and G. Imbens (2004): "Random Effects Estimators with Many Instrumental Variables, *Econometrica,* 72, 295-306.

Chao, J. C. and N. R. Swanson (2003): "Alternative Approximations of the Bias and MSE of the IV Estimator under Weak Identification with an Application to Bias Correction," Cowles Foundation Discussion Paper No. 1418, Yale University.

——— (2005): "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica,* 73, 1673-1692.

Donald, S. G. and W. K. Newey (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., J. Hausman, and G. Kuersteiner (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. and P. C. B. Phillips (2005): "GMM with Many Moment Conditions," *Econometrica,* 73, forthcoming.

Hansen, C., J. Hausman, and W. K. Newey (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, Dept. of Economics, University of California, Berkeley.

——— (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. and F. Windmeijer (2005): "Many Weak Moment Asymptotics for Generalized Empirical Likelihood Estimators," unpublished manuscript, Department of Economics, M.I.T.

Staiger, D. and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica,* 65, 557-586.

Stock, J. H. and M. Yogo (2005): "Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments," in *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas J. Rothenberg,* ed. by D. W. K. Andrews and J. H. Stock. Cambridge, UK: Cambridge University Press.