Pythagorean Generalization of Testing the Equality of Two Symmetric Positive Definite Matrices

JIN SEO CHO

School of Economics, Yonsei University &

Department of Economics

Hong Kong University of Science and Technology

Email: jinseocho@yonsei.ac.kr

PETER C.B. PHILLIPS

Yale University

University of Auckland

Singapore Management University &

University of Southampton

May 9, 2017

Abstract

We provide a new test for equality of two symmetric positive-definite matrices that leads to a convenient mechanism for testing specification using the information matrix equality or the sandwich asymptotic covariance matrix of the GMM estimator. The test relies on a new characterization of equality between two k dimensional symmetric positive-definite matrices A and B: the traces of AB^{-1} and BA^{-1} are equal to k if and only if $A = B$. Using this simple criterion, we introduce a class of omnibus test statistics for equality and examine their null and local alternative approximations under some mild regularity conditions. A preferred test in the class with good omni-directional power is recommended for practical work. Monte Carlo experiments are conducted to explore performance characteristics under the null and local as well as fixed alternatives. The test is applicable in many settings, including GMM estimation, SVAR models and high dimensional variance matrix settings.

Key Words: Matrix equality; Trace; Determinant; Arithmetic mean; Geometric mean; Harmonic mean; Sandwich covariance matrix; Eigenvalues.

Subject Classification: C01, C12, C52

1   Introduction

Comparing covariance matrices and testing the equivalence of two symmetric positive-definite matrices have attracted substantial past attention in both multivariate analysis and econometrics. For example, the asymptotic distribution of the maximum likelihood (ML) estimator is characterized by the usual information matrix equality. On the other hand, the information matrix equality does not hold for the quasi-ML (QML) estimator. As another example, least squares (LS) and generalized method of moments (GMM) estimators have relatively simple covariance matrix structures except when heteroskedasticity, model misspecification, or autocorrelation is present. The simple covariance matrix structure is then delivered by the proportional equality of two symmetric positive-definite matrices (viz., $X'X$ and $X'\Sigma X$ in the usual regression notation).

These material econometric concerns have led to much literature on covariance matrix equality testing, with special attention being given to the information matrix equality (*e.g.*, White, 1982; Hall, 1987; Orme, 1988; Chesher and Spady, 1991; Horowitz 1994; Dhaene and Hoorelbeke, 2004; and Golden, Henley, White, and Kashner, 2013), although work is not limited to that setting alone (*e.g.*, Bera and Hall, 1991). Much of this past work arises from the desire for an omnibus test without level distortion and with high power. The problem in size control is simply stated. For two general $k \times k$ symmetric positive-definite matrices A and B say, testing *every* pair of corresponding elements in A and B generates enormous level distortions for the tests even with moderately sized k.

The primary goals of the current study are to develop simple and straightforward omnibus tests for the equality of two symmetric positive-definite matrices and to broaden their implications for applied research. The approach that we use here has an antecedent in Cho and White (2014; CW, henceforth). CW provided omnibus tests of matrix equality by using the fact that the conditions $\text{tr}[BA^{-1}] = k$ and $\det[BA^{-1}] = 1$ are necessary and sufficient for $A = B$. Our starting point is to replace this condition with another, even simpler, characterization of equality that enables a new class of omnibus tests for equality that have little size distortion and comparable powers to other tests. The tests given in CW then become special cases of our approach. We also seek to clarify the interrelationships among the many tests that are now available and those that are developed in the current study. The paper therefore contributes by (i) introducing a class of easily implemented new tests that have good size and power properties, and (ii) providing a comprehensive study of the properties and performance characteristics of an extensive range of tests for covariance matrix equality.

These goals are achieved by evolving the CW approach using a new necessary and sufficient condition for the matrix equality. First, we examine a number of omnibus test statistics for equality of two symmetric

positive-definite matrices by the Pythagorean characterization. We show that the simple dual conditions $\text{tr}[AB^{-1}] = k$ and $\text{tr}[BA^{-1}] = k$ are also necessary and sufficient for $A = B$. This characterization is made by noting that $k^{-1}\text{tr}[BA^{-1}]$ and $k\text{tr}[AB^{-1}]^{-1}$ are the arithmetic and harmonic means of the eigenvalues of BA^{-1}, respectively, and that these means are equal if and only if all eigenvalues are identical. Under the given conditions, all eigenvalues are unity, implying that $BA^{-1} = I$. Note that this characterization is additional to that given in CW, viz. $\text{tr}[AB^{-1}] = k$ and $\det[BA^{-1}] = 1$ if and only if $A = B$, and a number of new testing factors can be obtained from this new characterization, which are additional to those in CW. Further, an even wider range of test factors can be obtained by pairing the arithmetic and geometric means or combining all of the Pythagorean means. More tests again are obtained by reversing the roles of A and B in the relations $k^{-1}\text{tr}[AB^{-1}]$, $\det[AB^{-1}]^{1/k}$, and $k\text{tr}[BA^{-1}]^{-1}$. All these test factors form a class, and we can apply continuous distance functions to the class to yield omnibus test statistics. Within this general framework, the tests given in CW become special cases of those developed here.

Second, we examine the performance characteristics of the various tests under the null, local, and fixed alternative hypotheses, and derive their asymptotic approximations. This examination broadens the analysis commenced in CW as it transpires that the null and alternative approximations given in CW continue to apply for our test statistics. All test statistics given in the current study are asymptotically governed by a single distribution under the null and local alternative, and they exhibit different behaviors under the fixed alternative. We single out the factors leading to test consistency and analyze the power relationships of the tests under the fixed alternative. When the fixed alternative hypothesis is partitioned into a number of explicit regions, the tests developed here are designed to estimate the dominant characteristic of each region, and estimating this factor delivers a test with respectable omni-directional power. This process of selection reduces the class size of the omnibus test statistics by considering only test statistics with distinctive power features under fixed alternatives.

Finally, the practical applicability of our test extends to a wide range of estimation methodologies. CW specifically focus on the application of their tests in the context of QML estimation, and for practical implementation they use Horowitz's (1994) parametric bootstrap to test information matrix equality. Our approach achieves wider applicability through implementation in the GMM context and by using the null limit distribution provided for the test and the residual bootstrap. Specifically, we show that our test statistic is useful in practical work for testing the optimal weight matrix condition in GMM estimation. Monte Carlo simulations are conducted, and we evaluate the performance of our test by the null limit distribution and residual bootstrapping. The simulations assist in confirming the relevance of the asymptotic theory and asymptotic comparisons in finite samples.

There are many potential uses of our approach in empirical applications beyond the work mentioned above for motivation. First, many structural models are specified using covariance matrix properties. For example, full information maximum likelihood (FIML) estimation is popular in structural vector autoregressions (SVARs), and model over-identification is often tested by the likelihood ratio (LR) test on the covariance matrix condition. As detailed below, the LR test for model over-identification turns out to be equivalent to the test statistics we consider here under the local alternative. As another example, multifactor models are often specified for portfolio risk analysis, and hypotheses on factors are also often tested by the LR test using the conditions on the covariance matrix. The LR test can be analyzed similar to the SVAR analysis provided here. Second, there is vast, growing literature on testing high-dimensional covariance matrix conditions, and our test approach can also be adapted to this case. As the dimension of the covariance matrix increases, the typical testing problem becomes nonstandard because the number of estimates also increases. Our approach can be easily adapted to testing a large covariance matrix as we illustrate below using the example in Bai, Ziang, Yao, and Zheng (2009). Third, Hong, Linton, and and Zhang (2017) recently examine a multivariate version of the variance ratio test. The test statistics they consider can be easily adapted to the test bases we are using, thereby broadening the relevance of their methodology. In addition to these examples, there are many other aspects of covariance matrix equivalence that are well-suited to the test procedure developed here.

The plan of this paper is as follows. Section 2 provides a fundamental result characterizing the equality between two symmetric positive-definite matrices. Section 3 motivates and defines the test statistics employed, and develops asymptotic theory under the null, alternative, and local alternative hypotheses. Section 4 examines the application of our tests to FIML estimation in the SVAR context. Simulation results are reported in Section 5. Concluding remarks are provided in Section 6. Mathematical proofs are collected in the online Supplement to this paper.

Before proceeding, we provide some notation. A function mapping $f : \mathcal{X} \mapsto \mathcal{Y}$ is denoted by $f(\cdot)$, evaluated derivatives such as $f'(x)|_{x=x_*}$ are written simply as $f'(x_*)$, and $\partial_x f(x) := (\partial/\partial x) f(x)$, $\partial^2_{x,y} f(x, y)$ $:= (\partial^2/\partial x \partial y) f(x, y)$.

2 A Basic Lemma and Its Testing Implications

Our starting point is the following fundamental lemma that characterizes the equality of two symmetric positive-definite matrices.

Lemma 1. *Let A and B be real symmetric positive-definite $k \times k$ matrices with $k \in \mathbb{N}$. Then, $A = B$ if and*

only if (i) $\text{tr}[D] = \text{tr}[D^{-1}] = k$, *where* $D := BA^{-1}$*; or (ii)* $\det[D] = 1$ *and* $\text{tr}[D^{-1}] = k$.   □

To our knowledge and somewhat surprisingly given its simplicity, Lemma 1 is new to the literature and is proved in the Supplement. Briefly, part (*i*) follows because the arithmetic mean of positive numbers is identical to their harmonic mean, if and only if all of the positive numbers are identical. Since $k^{-1}\text{tr}[D]$ is the arithmetic mean of the eigenvalues of D, and $k^{-1}\text{tr}[D^{-1}]$ is the inverse of the harmonic mean of the same eigenvalues, we have $D = I$, if and only if all the eigenvalues are identical to unity, which implies that $A = B$. Notably, and most conveniently for practical work, the criteria in (*i*) and (*ii*) involve only the leading elementary symmetric functions of the matrices D and D^{-1}.

As pointed out by a reviewer, the characterization in Lemma 1(*i*) can be generalized by associating the eigenvalues of D with a strictly Schur-convex function of them (*e.g.*, Marshall, Olkin, and Arnold, 1979). Thus, if $f(\cdot)$ and λ are a Schur-convex function and the vector of the eigenvalues of D, respectively, we have the equivalence $D = I$ if and only if $\text{tr}[D] = k$ and $f(\lambda) = f(\iota)$, where ι is the vector of ones. Note that $\text{tr}[D^{-1}] = \sum_{i=1}^{k} \lambda_i^{-1}$, and $1/x$ is a Schur-convex function, so that $\text{tr}[D^{-1}]$ is a convex function of the eigenvalues. This proves Lemma 1(*i*). In addition, we can apply D^{-1} to the criterion instead of D: if we let ρ be the vector of the eigenvalues of D^{-1}, $\text{tr}[D^{-1}] = k$ and $\det[D] = \exp(-\sum_{i=1}^{k} \log(\rho_i))$. Here, we note that $-\log(x)$ is a Schur-convex function, so that $\det[D]$ is a strictly Schur-convex function of ρ. Therefore, $D^{-1} = I$ if and only if $\text{tr}[D^{-1}] = k$ and $\det[D] = 1$, proving Lemma 1(*ii*). In another way, the characterization in Lemma 1(*i*) can also be associated with a convexity property of the trace operator. Note that $\phi(\cdot) := \text{tr}[(\cdot)^{-1}] + \text{tr}[\cdot]$ is a convex function on the space of $k \times k$ symmetric positive-definite matrices (*e.g.*, Bernstein, 2005, p. 283) and is also bounded from below by $2k$ (*e.g.*, Abadir and Magnus, 2005, p.338). The lower bound is achieved if and only if the argument of $\phi(\cdot)$ is I.

The characterization in Lemma 1 is different from that used in CW, in which the equality of two equal symmetric positive-definite matrices is characterized by both $\det[D]$ and $\text{tr}[D]$. Note that $\det[D]^{1/k}$ is the geometric mean of the eigenvalues of D. Furthermore, the geometric mean of positive numbers is identical to the arithmetic mean, if and only if the positive numbers are identical. Using this simple fact, CW characterized two equal symmetric positive-definite matrices by the condition that $\det[D] = 1$ and $\text{tr}[D] = k$. Lemma 1(*ii*) is then a corollary of Lemma 1(*i*) and the CW characterization.

Both Lemma 1 and the characterization in CW rely on fundamental properties of the Pythagorean (harmonic, geometric, and arithmetic) means of positive numbers: Harmonic mean $\leq$ Geometric mean $\leq$ Arithmetic mean. All three means are identical if the positive numbers are identical. Lemma 1(*i*) is obtained by interrelating the harmonic mean with the arithmetic mean, and CW links the geometric mean to the arithmetic mean for their characterization. Lemma 1(*ii*) also associates the harmonic and geometric means for

the equality.

There are solid grounds to use the trace and determinant-based test statistics for the equality of two symmetric positive-definite matrices. First, as mentioned above, these invariant polynomials are the leading elementary symmetric functions of the positive semi-definite matrices, and are simple and straightforward for practical implementation. Our Monte Carlo simulations also show that the test statistics defined below exhibit quality finite sample performance. Second, the theory of model selection information criteria has been developed by replacing Akaike's penalty term with the trace and/or determinant of the asymptotic covariance of an estimator (*e.g.*, Takeuchi, 1976; Bozdogan 2000), particularly when models are possibly misspecified. This motivates testing the equal covariance matrix hypothesis using the trace and determinant. Third, if the eigenvalues are explicitly involved in the test statistics (as distinct from only an implicit involvement via the elementary symmetric functions), it is challenging to obtain the null limit approximations of the test statistics. If we let $T_n := f(\widehat{\lambda}_n)$ for testing $D = I$, where $\widehat{\lambda}_n = \lambda(\widehat{D}_n)$ and $\widehat{D}_n$ is a consistent estimator for D, it is necessary to approximate T_n using the differential of $\lambda(\cdot)$ around I, which does not exist, *e.g.* Magnus (1985). Under the null $D = I$, $\lambda(D)$ is not simple, making it challenging to obtain the null limit approximation of T_n particularly when $\widehat{D}_n$ involves parameter estimation. James (1964) and Onatski, Moreira, and Hallin (2013) provide distributional properties of $\widehat{\lambda}_n$ for normally associated samples, although not for a general case that involves parameter estimation. Hence, testing equality of two symmetric positive definite matrices without explicitly involving the eigenvalues leads directly to the use of the elementary symmetric functions, thereby motivating the choice of $\mathrm{tr}[\widehat{D}_n]$ and $\det[\widehat{D}_n]$ as vehicles for testing $D = I$. Finally, as will be demonstrated below, our proposed tests are asymptotically equivalent to the likelihood ratio test statistic under a local alternative in a prototypical structural model context, implying that these statistics are locally optimal and can therefore be expected to have good power properties.

We now exploit Lemma 1 to test the equality of two symmetric positive-definite matrices. Lemma 1(*i*) is our first focus. Let $\tau := k^{-1}\mathrm{tr}[D]-1, \eta := k\mathrm{tr}[D^{-1}]^{-1}-1$, and $\xi := k^{-1}\mathrm{tr}[D]-k\mathrm{tr}[D^{-1}]^{-1}$ for notational simplicity. Note that if any two of τ, η, and ξ equal zero, the remaining one is also zero. Therefore, Lemma 1(*i*) holds if and only if any two of τ, η, and ξ equal zero. This implies that the equality of two symmetric positive-definite matrices can be tested by testing one of the following base hypotheses:

$$\mathcal{H}_0^{(1)}: \quad \tau = 0 \ \text{ and } \ \eta = 0 \quad \text{vs.} \quad \mathcal{H}_1^{(1)}: \tau \neq 0 \ \text{ or } \ \eta \neq 0;$$
$$\mathcal{H}_0^{(2)}: \quad \tau = 0 \ \text{ and } \ \xi = 0 \quad \text{vs.} \quad \mathcal{H}_1^{(2)}: \tau \neq 0 \ \text{ or } \ \xi \neq 0;$$
$$\mathcal{H}_0^{(3)}: \quad \eta = 0 \ \text{ and } \ \xi = 0 \quad \text{vs.} \quad \mathcal{H}_1^{(3)}: \eta \neq 0 \ \text{ or } \ \xi \neq 0.$$

Similarly, we can exploit Lemma 1(*ii*) and for this, let $\delta := \det[D]^{1/k}-1$ and $\gamma := \det[D]^{1/k}-k\mathrm{tr}[D^{-1}]^{-1}$.

If any two of δ, η, and γ are zero, the remaining one is zero, so that Lemma 1(ii) holds if and only if any two of them are zero. Hence, we construct the corresponding hypotheses as

$$\begin{aligned}
\mathcal{H}_0^{(4)} : \quad & \delta = 0 \ \text{ and } \ \eta = 0 \quad \text{vs.} \quad \mathcal{H}_1^{(4)} : \delta \neq 0 \ \text{ or } \ \eta \neq 0; \\
\mathcal{H}_0^{(5)} : \quad & \delta = 0 \ \text{ and } \ \gamma = 0 \quad \text{vs.} \quad \mathcal{H}_1^{(5)} : \delta \neq 0 \ \text{ or } \ \gamma \neq 0; \\
\mathcal{H}_0^{(6)} : \quad & \eta = 0 \ \text{ and } \ \gamma = 0 \quad \text{vs.} \quad \mathcal{H}_1^{(6)} : \eta \neq 0 \ \text{ or } \ \gamma \neq 0.
\end{aligned}$$

The following hypotheses correspond to those considered in CW. They let $\sigma := k^{-1}\mathrm{tr}[D] - \det[D]^{1/k}$ and test whether any two of τ, δ, and σ are zero by considering the following hypotheses:

$$\begin{aligned}
\mathcal{H}_0^{(7)} : \quad & \tau = 0 \ \text{ and } \ \delta = 0 \quad \text{vs.} \quad \mathcal{H}_1^{(7)} : \tau \neq 0 \ \text{ or } \ \delta \neq 0; \\
\mathcal{H}_0^{(8)} : \quad & \tau = 0 \ \text{ and } \ \sigma = 0 \quad \text{vs.} \quad \mathcal{H}_1^{(8)} : \tau \neq 0 \ \text{ or } \ \sigma \neq 0; \\
\mathcal{H}_0^{(9)} : \quad & \delta = 0 \ \text{ and } \ \sigma = 0 \quad \text{vs.} \quad \mathcal{H}_1^{(9)} : \delta \neq 0 \ \text{ or } \ \sigma \neq 0.
\end{aligned}$$

Furthermore, the prior hypotheses can be extended to involve τ, δ, and η at the same time, yielding

$$\begin{aligned}
\mathcal{H}_0^{(10)} : \quad & \tau = 0 \ \text{ and } \ \delta - \eta = 0 \quad \text{vs.} \quad \mathcal{H}_1^{(10)} : \tau \neq 0 \ \text{ or } \ \delta - \eta \neq 0; \\
\mathcal{H}_0^{(11)} : \quad & \delta = 0 \ \text{ and } \ \tau - \eta = 0 \quad \text{vs.} \quad \mathcal{H}_1^{(11)} : \delta \neq 0 \ \text{ or } \ \tau - \eta \neq 0; \\
\mathcal{H}_0^{(12)} : \quad & \eta = 0 \ \text{ and } \ \tau - \delta = 0 \quad \text{vs.} \quad \mathcal{H}_1^{(12)} : \eta \neq 0 \ \text{ or } \ \tau - \delta \neq 0.
\end{aligned}$$

All these 12 hypothesis systems are equivalent systems of hypotheses to the simple null $\mathcal{H}_0 : A = B$ versus the alternative $\mathcal{H}_1 : A \neq B$.

Several remarks are warranted regarding this testing methodology. First, note that the testing factors τ, δ, and η are invariant to linear transformations of the null $\mathcal{H}_0 : A = B$. Thus, for any invertible matrix H, if $\tau^\dagger$, $\delta^\dagger$, and $\eta^\dagger$ are computed using AH and BH, it easily follows that $\tau^\dagger = \tau$, $\delta^\dagger = \delta$, and $\eta^\dagger = \eta$ because $BH(AH)^{-1} = D$. Therefore, τ, δ, and η are invariant to linear transformations. The other factors ξ, γ, and σ share the same property. Second, the roles of A and B can be reversed when computing the testing factors τ, δ and η.

We introduce testing environments by supposing that the previously defined matrices A and B are in fact parameterized as $A \equiv A(\theta_*)$ and $B \equiv B(\theta_*)$, respectively, where both $A(\cdot)$ and $B(\cdot)$ are defined on $\Theta \in \mathbb{R}^\ell$, and $\theta_* \in \Theta$ is an unknown parameter. This dependence is motivated by the fact that most covariance estimators are obtained as second-stage outputs after estimating the unknown parameter. For example, the asymptotic covariance matrix of the (Q)ML estimator is the sandwich covariance matrix that is often consistently obtained through the vehicle of heteroskedastic and autocorrelation consistent estimation

in the typical limit theory $\sqrt{n}(\widehat{\theta}_n - \theta_*) \overset{A}{\sim} N(0, A(\theta_*)^{-1} B(\theta_*) A(\theta_*)^{-1})$. If $A(\theta_*) = B(\theta_*)$, the asymptotic covariance of $\widehat{\theta}_n$ becomes identical to that of ML estimator, so that testing $A(\theta_*) = B(\theta_*)$ can be associated with model specification testing. When proceeding with this association, we further suppose that $A_n := A_n(\theta_*)$ and $B_n := B_n(\theta_*)$ estimate $A(\theta_*)$ and $B(\theta_*)$ consistently, where $A_n(\cdot)$ and $B_n(\cdot)$ are consistent for $A(\cdot)$ and $B(\cdot)$ uniformly on Θ and are uniformly positive definite almost surely on Θ for large enough n. That is, for any $x \in \mathbb{R}^k \setminus \{0\}$, $\mathbb{P}(\lim_{n\to\infty} \inf_{\theta\in\Theta} x' A_n(\theta) x > 0) = 1$ and $\mathbb{P}(\lim_{n\to\infty} \inf_{\theta\in\Theta} x' B_n(\theta) x > 0) = 1$. Therefore, $D_n := B_n A_n^{-1}$ and D_n^{-1} consistently estimate D and D^{-1}, respectively. Here, D is estimated by a two-step estimation procedure. Specifically, the unknown parameter θ_* is consistently estimated by an estimator $\widehat{\theta}_n$, so that $\widehat{A}_n := A_n(\widehat{\theta}_n)$ and $\widehat{B}_n := B_n(\widehat{\theta}_n)$ are consistent for $A(\theta_*)$ and $B(\theta_*)$, respectively. Therefore, $\widehat{D}_n := \widehat{B}_n \widehat{A}_n^{-1}$ and $\widehat{D}_n^{-1} = \widehat{A}_n \widehat{B}_n^{-1}$ are also consistent for D and D^{-1}, respectively. To reduce notational clutter, we simply indicate the influence of θ_* on these matrices by letting $A_* := A(\theta_*)$ $B_* := B(\theta_*)$, $D_* := B_* A_*^{-1}$. Similarly, let $\tau_* := k^{-1}\mathrm{tr}[D_*] - 1$, $\eta_* := k/\mathrm{tr}[D_*^{-1}] - 1$, $\delta_* := \det[D_*]^{1/k} - 1$, $\xi_* := \tau_* - \eta_*$, $\gamma_* := \delta_* - \eta_*$, and $\sigma_* := \tau_* - \delta_*$. When these matrices are estimated using $\widehat{A}_n$ and $\widehat{B}_n$, we denote the resulting statistics as $\widehat{\tau}_n := k^{-1}\mathrm{tr}[\widehat{D}_n] - 1$, $\widehat{\eta}_n := k/\mathrm{tr}[\widehat{D}_n^{-1}] - 1$, $\widehat{\delta}_n := \det[\widehat{D}_n]^{1/k} - 1$, $\widehat{\xi}_n := \widehat{\tau}_n - \widehat{\eta}_n$, $\widehat{\gamma}_n := \widehat{\delta}_n - \widehat{\eta}_n$, and $\widehat{\sigma}_n := \widehat{\tau}_n - \widehat{\delta}_n$. All these statistics, which form the base elements of the tests given below, are dependent upon $\widehat{\theta}_n$. For notational simplicity, we also let $\widetilde{D}_* := A_* B_*^{-1}$ and $\widetilde{D}_n := \widehat{A}_n \widehat{B}_n^{-1}$. Therefore, $\widetilde{D}_* = D_*^{-1}$ and $\widetilde{D}_n := \widehat{D}_n^{-1}$.

With this discussion in hand, we now define the first group of tests

$$\widehat{\mathfrak{B}}_n^{(1)} := \frac{nk}{2}\left(\widehat{\tau}_n^2 + 2\widehat{\sigma}_n\right) \quad \text{and} \quad \widehat{\mathfrak{B}}_n^{(2)} := \frac{nk}{2}\left(\widehat{\delta}_n^2 + 2\widehat{\sigma}_n\right),$$

which modify the tests in CW. These tests exploit the discriminatory properties of the statistics $\widehat{\tau}_n$ and $\widehat{\delta}_n$, which embody elements of the Wald (1943) test principle and correspond with $\mathcal{H}^{(8)}$ and $\mathcal{H}^{(9)}$, respectively. The coefficients of these statistics differ from those in CW by the scale factor $1/2$. As detailed below, this modification is useful to achieve a direct comparison of the leading terms that are obtained as approximations of the tests under the alternative.

We define a second group of tests as follows:

$$\widehat{\mathfrak{D}}_n^{(1)} := \frac{nk}{2}\left(\widehat{\tau}_n^2 + \widehat{\xi}_n\right), \quad \widehat{\mathfrak{D}}_n^{(2)} := \frac{nk}{2}\left(\widehat{\eta}_n^2 + \widehat{\xi}_n\right), \quad \widehat{\mathfrak{S}}_n^{(1)} := \frac{nk}{2}\left(\widehat{\delta}_n^2 + 2\widehat{\gamma}_n\right), \quad \widehat{\mathfrak{S}}_n^{(2)} := \frac{nk}{2}\left(\widehat{\eta}_n^2 + 2\widehat{\gamma}_n\right).$$

Note that $\widehat{\mathfrak{D}}_n^{(1)}$ and $\widehat{\mathfrak{D}}_n^{(2)}$ are defined by associating the arithmetic mean with the harmonic mean, whereas $\widehat{\mathfrak{S}}_n^{(1)}$ and $\widehat{\mathfrak{S}}_n^{(2)}$ are defined by associating the geometric mean with the harmonic mean. As before, $\widehat{\tau}_n$, $\widehat{\delta}_n$, and $\widehat{\eta}_n$ are empowered with discriminatory capability.

In addition, other omnibus test statistics may be defined by reversing the roles of A_* and B_* or by combining all the test base elements. If we let $\widetilde{\tau}_* := k^{-1}\mathrm{tr}[\widetilde{D}_*] - 1$, $\widetilde{\eta}_* := k/\mathrm{tr}[\widetilde{D}_*^{-1}] - 1$, and $\widetilde{\delta}_* := \det[\widetilde{D}_*]^{1/k} - 1$, the only difference between $\widetilde{D}_*$ and D_* lies in the fact that the roles of A_* and B_* are reversed. Using these testing factors, we may be able to define additional test statistics as before. Nevertheless, note that $\widetilde{\tau}_* + 1 = 1/(1 + \eta_*)$, $\widetilde{\delta}_* + 1 = 1/(1 + \delta_*)$ and $\widetilde{\eta}_* + 1 = 1/(1 + \tau_*)$. That is, the testing factors obtained from $\widetilde{D}_*$ are inversely associated with the testing factors using D_*, so that the test procedure via $\widetilde{D}_*$ can be constructed equivalently to that using D_*. We therefore eliminate them from consideration. Next, we also consider the following test statistics:

$$\widehat{\mathfrak{E}}_n^{(1)} := \frac{nk}{2}\left(\widehat{\tau}_n^2 + 2\widehat{\gamma}_n\right), \quad \widehat{\mathfrak{E}}_n^{(2)} := \frac{nk}{2}\left(\widehat{\eta}_n^2 + 2\widehat{\sigma}_n\right), \quad \text{and} \quad \widehat{\mathfrak{E}}_n^{(3)} := \frac{nk}{2}\left(\widehat{\delta}_n^2 + \widehat{\xi}_n\right)$$

to test $\mathcal{H}_0^{(10)}$, $\mathcal{H}_0^{(11)}$, and $\mathcal{H}_0^{(12)}$, respectively. The motivations of these tests are the same as for the earlier test statistics. Indeed, the test statistics are linear combinations of the previous six test statistics, viz.,

$$\widehat{\mathfrak{E}}_n^{(1)} \equiv 2\widehat{\mathfrak{D}}_n^{(1)} - \widehat{\mathfrak{B}}_n^{(1)}, \qquad \widehat{\mathfrak{E}}_n^{(2)} \equiv 2\widehat{\mathfrak{D}}_n^{(2)} - \widehat{\mathfrak{S}}_n^{(2)}, \qquad \text{and} \qquad \widehat{\mathfrak{E}}_n^{(3)} \equiv \frac{1}{2}\widehat{\mathfrak{B}}_n^{(2)} + \frac{1}{2}\widehat{\mathfrak{S}}_n^{(1)}, \qquad (1)$$

which implies that the asymptotic behaviors of $\widehat{\mathfrak{E}}_n^{(1)}$, $\widehat{\mathfrak{E}}_n^{(2)}$, and $\widehat{\mathfrak{E}}_n^{(3)}$ are determined by those of $\widehat{\mathfrak{B}}_n^{(i)}$, $\widehat{\mathfrak{D}}_n^{(i)}$, and $\widehat{\mathfrak{S}}_n^{(i)}$ ($i = 1, 2$). Below we show that these additional test statistics are equivalent to $(\widehat{\mathfrak{B}}_n^{(i)}, \widehat{\mathfrak{D}}_n^{(i)}, \widehat{\mathfrak{S}}_n^{(i)})$ ($i = 1, 2$) under the null and local alternative.

Finally, we consider the maximal statistic

$$\widehat{\mathfrak{M}}_n := \max_{j \in \{1, 2\}}[\widehat{\mathfrak{B}}_n^{(j)}, \widehat{\mathfrak{S}}_n^{(j)}, \widehat{\mathfrak{E}}_n^{(j)}].$$

As we show in Section 3, the fixed alternative is partitioned in a way that one of $\widehat{\mathfrak{B}}_n^{(i)}$, $\widehat{\mathfrak{S}}_n^{(i)}$, and $\widehat{\mathfrak{E}}_n^{(i)}$ ($i = 1, 2$) has a dominant leading term over those of other tests that diverge as n tends to infinity, implying that asymptotic power of a test is warranted by consistently estimating the dominant leading terms. Converting this idea into a test statistic gives rise to $\widehat{\mathfrak{M}}_n$ as our final test statistic. Note that a number of distance functions can be used to define tests, and the uniform norm is popularly applied to continuous Gaussian processes in the field of specification testing for the conditional mean (*e.g.*, Cho and White, 2011; Baek, Cho, and Phillips, 2015) to detect the worst test base. By the same motivation, we can also justify use of the uniform norm for $\widehat{\mathfrak{M}}_n$. As it turns out, all tests constituting $\widehat{\mathfrak{M}}_n$ are asymptotically equivalent under the null and local alternative, so that $\widehat{\mathfrak{M}}_n$ also turns out to have the same null limit distribution as the other test statistics, implying that the asymptotic power of $\widehat{\mathfrak{M}}_n$ is greater than the other tests. On the other hand,

by virtue of its construction as a maximal statistic $\widehat{\mathfrak{M}}_n$ is expected to possess finite sample level distortion bigger than the other tests. Finite sample size distortion may be corrected by bootstrapping and below we examine the relative performance of the tests by simulation.

Before moving to the next section, we add some remarks about these tests. First, the nine test statistics $\widehat{\mathfrak{B}}_n^{(i)}$, $\widehat{\mathfrak{D}}_n^{(i)}$, $\widehat{\mathfrak{S}}_n^{(i)}$ ($i = 1, 2$) and $\widehat{\mathfrak{E}}_n^{(j)}$ ($j = 1, 2, 3$) do not specifically test any of the base hypotheses $\mathcal{H}_0^{(1)}$, $\mathcal{H}_0^{(4)}$, and $\mathcal{H}_0^{(7)}$. Tests involving the sum of any two of τ_*^2, δ_*^2, and η_*^2 are inferior to the nine test statistics mainly because their leading terms are asymptotically dominated by the maximum leading term of the nine test statistics considered here. Hence, we focus on the above nine test statistics as the primary tests leading to a single preferred maximal statistic. Second, as we show more specifically below, the space of the alternative hypothesis can be partitioned into 6 regions such that on each region, the dominant term estimated by one of the six test statistics that form $\widehat{\mathfrak{M}}_n$ is bigger than or equal to the dominant terms estimated by the other test statistics. Correspondingly, $\widehat{\mathfrak{M}}_n$ is defined without using $\widehat{\mathfrak{D}}_n^{(i)}$ ($i = 1, 2$) and $\widehat{\mathfrak{E}}_n^{(3)}$. Third, Hong, Linton, and Zhang (2017) recently examined a multivariate version of the variance ratio test. Their matrix normalized multivariate variance ratio statistic is designed to estimate a quantity of the form $A_*^{-1/2} B_* A_*^{-1/2}$ in our notation. In that case, testing $A_*^{-1/2} B_* A_*^{-1/2} = I$ corresponds to the same hypothesis as that considered in Lemma 1. Our test bases can then be easily adapted to this specific testing environment.

3 Asymptotic Expansions of the Test Statistics

This section examines asymptotic expansions of the test statistics under the null, local, and fixed alternative hypotheses. We also supplement the test statistics considered in CW. Before examining these asymptotic approximations, we provide the following regularity conditions, some of which have already been mentioned.

Assumption A (Cho and White, 2014). *(i)* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a complete probability space;*

(ii) $\Theta \subset \mathbb{R}^\ell$ *is a compact convex set with non-empty interior and* $\ell \in \mathbb{N}$*;*

(iii) a sequence of measurable mappings $\{\widehat{\theta}_n : \Omega \mapsto \Theta\}$ *is consistent for a unique* $\theta_* \in \text{int}(\Theta)$*;*

(iv) $A : \Theta \mapsto \mathbb{R}^{k \times k}$ *and* $B : \Theta \mapsto \mathbb{R}^{k \times k}$ *are in* $\mathcal{C}^{(2)}(\Theta)$*, and* A_* *and* B_* *are positive definite;*

(v) $A_n(\cdot)$ *and* $B_n(\cdot)$ *are consistent for* $A(\cdot)$ *and* $B(\cdot)$*, respectively, uniformly on* Θ*, viz.,* $\sup_{\theta \in \Theta} \|A_n(\theta) - A(\theta)\|_\infty = o_{\mathbb{P}}(1)$ *and* $\sup_{\theta \in \Theta} \|B_n(\theta) - B(\theta)\|_\infty = o_{\mathbb{P}}(1)$*, where* $\|\cdot\|_\infty$ *is the matrix maximum norm;*

(vi) $\sqrt{n}[(\widehat{\theta}_n - \theta_*)', \text{vech}[A_n - A_*]', \text{vech}[B_n - B_*]']' = O_{\mathbb{P}}(1)$*;*

(vii) for $j = 1, \ldots, \ell$*,* $\partial_j A_n(\cdot)$ *and* $\partial_j B_n(\cdot)$ *are consistent for* $\partial_j A(\cdot)$ *and* $\partial_j B(\cdot)$*, uniformly on* Θ*; and*

(viii) for $j = 1, \ldots, \ell$*,* $H_{j,n} = O_{\mathbb{P}}(n^{-1/2})$ *and* $G_{j,n} = O_{\mathbb{P}}(n^{-1/2})$*, where* $H_{j,n} := A_*^{-1} \partial_j (A_n - A_*)$ *and* $G_{j,n} := B_*^{-1} \partial_j (B_n - B_*)$*.* □

These conditions hold for most standard estimators based on (Q)MLE, LS, or (G)MM procedures when applied in standard environments. The same framework was employed in CW and facilitates comparison of our tests and findings with theirs under the same conditions.

Our omnibus tests are motivated by testing whether the critical quantities τ_*, δ_*, η_*, σ_* ξ_*, and γ_*, which we call the test base elements, equal zero. The stochastic asymptotic representations of consistent estimates of these quantities are functions of $L_n := P_n + \sum_{j=1}^{\ell} (\widehat{\theta}_{j,n} - \theta_{j,*}) R_{j,*}$, where $P_n := W_n - U_n := B_*^{-1}(B_n - B_*) - A_*^{-1}(A_n - A_*)$, and for $j = 1, 2, \ldots, \ell$, $R_{j,*} := B_*^{-1}\partial_j B_* - A_*^{-1}\partial_j A_*$. Note that under Assumption A this is a consequence of the fact that L_n is an asymptotic approximation of $\widehat{D}_n$ around D_* such that P_n is an approximation of D_n around D_*, and the remainder on the right side denotes the uncertainty from parameter estimation. We therefore have $L_n = O_{\mathbb{P}}(n^{-1/2})$, $P_n = O_{\mathbb{P}}(n^{-1/2})$, and for $j = 1, 2, \ldots, \ell$, $R_{j,*} = O(1)$. These correspond with the definitions in CW.

We now develop null approximations for each of the tests. The null approximations are obtained by higher order approximations of the tests. As noticed by CW in the case of $\widehat{\tau}_n$ and $\widehat{\delta}_n$, $\widehat{\tau}_n$, $\widehat{\delta}_n$, and $\widehat{\eta}_n$ are asymptotically equivalent under the null, so that $\widehat{\sigma}_n$, $\widehat{\xi}_n$, and $\widehat{\gamma}_n$ have a convergence rate n^{-1} that is faster than $\widehat{\tau}_n$, $\widehat{\delta}_n$ and $\widehat{\eta}_n$. In consequence, the desired asymptotic null approximations involve the study of higher order approximants for $\widehat{\tau}_n$, $\widehat{\delta}_n$, and $\widehat{\eta}_n$, which are provided in the Supplement. From these approximations, we now straightforwardly deliver the asymptotic null approximations of the tests. We collect the results together in the following theorem which characterizes the relationships between the test statistics. For notational simplicity, let $K_n := A_*^{-1}\{B_n - A_n + \sum_{j=1}^{\ell} \partial_j(B_* - A_*)(\widehat{\theta}_{j,n} - \theta_{j,*})\}$, which follows by imposing the null $A_* = B_*$ on the linearization L_n.

Theorem 1. *Given Assumption A and $\mathcal{H}_0$, all of the test statistics $\widehat{\mathfrak{B}}_n^{(i)}$, $\widehat{\mathfrak{D}}_n^{(i)}$, $\widehat{\mathfrak{S}}_n^{(i)}$, $\widehat{\mathfrak{E}}_n^{(j)}$, and $\widehat{\mathfrak{M}}_n$ have the asymptotic form $\frac{n}{2}\mathrm{tr}[K_n^2] + o_{\mathbb{P}}(1)$ ($i = 1, 2$; $j = 1, 2, 3$).* $\quad\square$

Remarks

(a) Theorem 1 is proved sequentially . The approach is to demonstrate initially that $\widehat{\mathfrak{B}}_n^{(1)}$, $\widehat{\mathfrak{B}}_n^{(2)}$, $\widehat{\mathfrak{D}}_n^{(1)}$, $\widehat{\mathfrak{D}}_n^{(2)}$, $\widehat{\mathfrak{S}}_n^{(1)}$, and $\widehat{\mathfrak{S}}_n^{(2)}$ are asymptotically equivalent under $\mathcal{H}_0$, implying that the asymptotic approximations of $\widehat{\mathfrak{E}}_n^{(1)}$, $\widehat{\mathfrak{E}}_n^{(2)}$, and $\widehat{\mathfrak{E}}_n^{(3)}$ are obtained as $\frac{n}{2}\mathrm{tr}[K_n^2] + o_{\mathbb{P}}(1)$ under $\mathcal{H}_0$ by (1), so that all of these are asymptotically equivalent under $\mathcal{H}_0$. It then follows that $\widehat{\mathfrak{M}}_n = \frac{n}{2}\mathrm{tr}[K_n^2] + o_{\mathbb{P}}(1)$ under $\mathcal{H}_0$ since all components that constitute the maximal statistic are asymptotically equivalent under $\mathcal{H}_0$.

(b) The test base L_n is not new to the literature. It reduces to K_n under $\mathcal{H}_0$ and is the same as the test base used for the information matrix equality test in White (1982).

(c) The null limit distribution of the test follows from the limit distribution of $\sqrt{n}K_n$. Given Assumption A(vi), the limit distribution of $(\widehat{\theta}_n, A_n, B_n)$ is often characterized by the central limit theorem, so that if we suppose that $\sqrt{n}[(\widehat{\theta}_n - \theta_*), A_n - A_*, B_n - B_*] \Rightarrow (Z_\theta, Z_A, Z_B)$ such that $[Z_\theta', \mathrm{vech}(Z_A)', \mathrm{vech}(Z_B)']' \sim N(0, \Sigma_*)$, then we can expect that $\sqrt{n}K_n \Rightarrow A_*^{-1}(Z_B - Z_A + (Z_\theta' \otimes I_k)Q_*)$, where $Q_* := [\partial_1(B_* - A_*)', \ldots, \partial_\ell(B_* - A_*)']'$, so that the null limit distribution of the tests in Theorem 1 is obtained as $\frac{1}{2}\mathrm{tr}[A_*^{-2}(Z_B - Z_A + (Z_\theta' \otimes I)Q_*)^2]$ using the fact that K_n is symmetric. We further note that this null limit distribution is a weighted sum of chi-squared distributions as reported in the following result.

Corollary 1. *Given Assumption A and $\mathcal{H}_0$, if it further holds that for some positive-semi definite Σ_*,*
$\sqrt{n}[(\widehat{\theta}_n - \theta_*), A_n - A_*, B_n - B_*] \Rightarrow (Z_\theta, Z_A, Z_B)$ *such that* $[Z_\theta', \mathrm{vech}(Z_A)', \mathrm{vech}(Z_B)']' \sim N(0, \Sigma_*)$, *all of* $\widehat{\mathfrak{B}}_n^{(i)}, \widehat{\mathfrak{D}}_n^{(i)}, \widehat{\mathfrak{S}}_n^{(i)}, \widehat{\mathfrak{E}}_n^{(j)}$, *and* $\widehat{\mathfrak{M}}_n$ *($i = 1, 2$; $j = 1, 2, 3$) are asymptotically distributed as* $\mathcal{Z}'\Omega_*\mathcal{Z}$, *where*

$$
\mathcal{Z} := \begin{bmatrix} \mathrm{vec}(Z_B - Z_A) \\ \mathrm{vec}(Z_\theta' \otimes I_k) \end{bmatrix} \quad and \quad \Omega_* := \frac{1}{2}\begin{bmatrix} A_*^{-1} \otimes A_*^{-1} & A_*^{-1}Q_*' \otimes A_*^{-1} \\ A_*^{-1} \otimes Q_*A_*^{-1} & A_*^{-1}Q_*' \otimes Q_*A_*^{-1} \end{bmatrix}. \qquad \square
$$

The null limit critical values of the tests can be obtained by estimating Σ_* and Ω_* consistently.

Next we examine asymptotic approximations of the tests under sequences of local alternatives. For some symmetric positive-definite $\bar{A}_*$ and $\bar{B}_*$ with $\bar{A}_* \neq \bar{B}_*$, we consider the local alternatives

$$
\mathcal{H}_\ell : A_{*,n} = A_* + n^{-1/2}\bar{A}_*, \quad B_{*,n} = B_* + n^{-1/2}\bar{B}_*, \quad and \quad A_* = B_*.
$$

As the sample size $n \to \infty$, $A_{*,n}$ and $B_{*,n}$ converge to A_* and B_* at the rate $n^{-1/2}$. $\mathcal{H}_\ell$ reduces to $\mathcal{H}_0$ if $\bar{A}_* = \bar{B}_*$ and the local alternatives differ from the null by requiring that $\bar{A}_* \neq \bar{B}_*$. This local alternative generalizes the one used in CW where it is assumed that $\bar{A}_* = 0$.

The following separate conditions are imposed for the local approximations.

Assumption B (Local Alternative). *(i) $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space;*
(ii) $\Theta \subset \mathbb{R}^\ell$ is a compact convex set with non-empty interior and $k \in \mathbb{N}$;
(iii) a sequence of measurable mappings $\{\widehat{\theta}_n : \Omega \mapsto \Theta\}$ is consistent for a unique $\theta_* \in \mathrm{int}(\Theta)$;
(iv) $A : \Theta \mapsto \mathbb{R}^{k \times k}$ and $B : \Theta \mapsto \mathbb{R}^{k \times k}$ are in $\mathcal{C}^{(2)}(\Theta)$, and A_* and B_* are symmetric and positive definite;
(v) $\bar{A} : \Theta \mapsto \mathbb{R}^{k \times k}$ and $\bar{B} : \Theta \mapsto \mathbb{R}^{k \times k}$ are in $\mathcal{C}^{(1)}(\Theta)$ and such that $\bar{A}_* := \bar{A}(\theta_*)$ and $\bar{B}_* := \bar{B}(\theta_*)$ are symmetric and positive definite, and $\bar{A}_* \neq \bar{B}_*$;
(vi) $A_n(\cdot)$ and $B_n(\cdot)$ are consistent for $A(\cdot)$ and $B(\cdot)$, respectively, uniformly on Θ;
(vii) $\sqrt{n}[(\widehat{\theta}_n - \theta_*)', \mathrm{vech}[A_n - A_{*,n}]', \mathrm{vech}[B_n - B_{*,n}]']' = O_\mathbb{P}(1)$;

11

(viii) for $j = 1, \ldots, \ell$, $\partial_j A_n(\cdot)$ and $\partial_j B_n(\cdot)$ are consistent for $\partial_j A(\cdot)$ and $\partial_j B(\cdot)$, uniformly on Θ; and

(ix) for $j = 1, \ldots, \ell$, $H_{j,o,n} = O_{\mathbb{P}}(n^{-1/2})$ and $G_{j,o,n} = O_{\mathbb{P}}(n^{-1/2})$, where $H_{j,o,n} := A_*^{-1}\partial_j(A_n - A_{*,n})$ and $G_{j,o,n} := B_*^{-1}\partial_j(B_n - B_{*,n})$. □

The major differences between Assumptions A and B are in B(*v*, *vii*, and *ix*). The localizing coefficient matrices $\bar{A}_*$ and $\bar{B}_*$ are formally introduced in Assumption B(*v*), and the other two conditions modify the corresponding conditions in Assumption A to accommodate the presence of the localizing parameters. We define $M_{o,n} := B_*^{-1}(B_n - A_n - B_{*,n} + A_{*,n})$ and $K_{o,n} := M_{o,n} + \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})S_{j,*}$, where $S_{j*} := A_*^{-1}(\partial_j B_* - \partial_j A_*)$. These statistics are defined to highlight the asymptotic roles of the localizing parameters. Using this notation, we provide the asymptotic approximations under $\mathcal{H}_\ell$ in the following theorem.

Theorem 2. *Given Assumption B and $\mathcal{H}_\ell$, all of the statistics $\widehat{\mathfrak{B}}_n^{(i)}$, $\widehat{\mathfrak{D}}_n^{(i)}$, $\widehat{\mathfrak{S}}_n^{(i)}$, $\widehat{\mathfrak{E}}_n^{(j)}$, and $\widehat{\mathfrak{M}}_n$ have the asymptotic form $\frac{1}{2}\text{tr}[(V_* + \sqrt{n}K_{o,n})^2] + o_{\mathbb{P}}(1)$ ($i = 1, 2$; $j = 1, 2, 3$), where $V_* := F_* - C_*$, $F_* := B_*^{-1}\bar{B}_*$, and $C_* := A_*^{-1}\bar{A}_*$.* □

Remarks

(a) Theorem 2 naturally extends Theorem 1. Thus, if $V_* = 0$, the implications given below Theorem 1 apply, which implies that the asymptotic approximations of the tests are obtained by shifting the location parameter of $\text{tr}[K_{o,n}]$ by $n^{-1/2}\text{tr}[V_*]$, from which the local power of the tests derives.

(b) If $\text{tr}[V_*^2] = 0$, then $\widehat{\mathfrak{B}}_n^{(1)}$, $\widehat{\mathfrak{B}}_n^{(2)}$, $\widehat{\mathfrak{D}}_n^{(1)}$, $\widehat{\mathfrak{D}}_n^{(2)}$, $\widehat{\mathfrak{S}}_n^{(1)}$, and $\widehat{\mathfrak{S}}_n^{(2)}$ all have location parameters that are the same as those under the null hypothesis, and local powers are correspondingly affected. Thus, $\text{tr}[V_*^2] \neq 0$ is necessary for these tests to have non-trivial local powers.

(c) Note that even when $A_{n,*} = A_* + n^{-1/2}\bar{A}_* + o(n^{-1/2})$ or $B_{n,*} = B_* + n^{-1/2}\bar{B}_* + o(n^{-1/2})$, the results stated in Theorem 2 still hold. For brevity, the $o(n^{-1/2})$ remainders from $A_{n,*}$ and $B_{n,*}$ in the local alternative hypothesis are omitted.

(d) The asymptotic expansion of $\widehat{\mathfrak{M}}_n$ is also equal to $\frac{1}{2}\text{tr}[(V_* + \sqrt{n}K_{o,n})^2] + o_{\mathbb{P}}(1)$ under $\mathcal{H}_\ell$ due to the fact that all components constituting the maximal test are equivalent under $\mathcal{H}_\ell$.

(e) If $\sqrt{n}K_{o,n}$ obeys a central limit theorem as in Corollary 1, the local limit distributions of all tests in Theorem 2 turn out to be identical to a weighted sum of chi-squared distributions with zero locality parameter.

We now examine asymptotic approximations of the tests under a fixed alternative. The nine equivalent test statistics $\widehat{\mathfrak{B}}_n^{(i)}, \widehat{\mathfrak{D}}_n^{(i)}, \widehat{\mathfrak{S}}_n^{(i)}$ $(i = 1, 2)$ and $\widehat{\mathfrak{E}}_n^{(j)}$ $(j = 1, 2, 3)$ under Assumption B motivate to examine them under a fixed alternative. As it turns out, the nine test statistics are approximated by the sum of a leading term estimated by the test statistic, a linear combination of $\mathrm{tr}[D_*'L_n]$, $\mathrm{tr}[\widetilde{D}_*'L_n]$, and $\mathrm{tr}[L_n]$, and a remainder that is $O_{\mathbb{P}}(1)$. As an example, $\widehat{\mathfrak{B}}_n^{(1)} = \frac{nk}{2}(\tau_*^2 + 2\sigma_*) + n\{(\tau_* + 1)\mathrm{tr}[D_*'L_n] - (\delta_* + 1)\mathrm{tr}[L_n]\} + O_{\mathbb{P}}(1)$. Note that the leading term of $\widehat{\mathfrak{B}}_n^{(1)}$, given as $\frac{nk}{2}(\tau_*^2 + 2\sigma_*)$, is what $\widehat{\mathfrak{B}}_n^{(1)}$ estimates. If this term differs from zero, the test has consistent power, and omnibus power is derived from this property and Lemma 1, although the discriminatory powers of the test statistics are also affected by the second-order terms given by $\mathrm{tr}[D_*'L_n]$, $\mathrm{tr}[\widetilde{D}_*'L_n]$ and $\mathrm{tr}[L_n]$. Expansions for the other statistics are similar, are obtained in the same way, and are provided in the Supplement.

As each test has a different leading term, its power pattern differs from the other tests. The following theorem states the interrelationships among the leading terms in $\widehat{\mathfrak{B}}_n^{(i)}, \widehat{\mathfrak{D}}_n^{(i)}, \widehat{\mathfrak{S}}_n^{(i)}$, and $\widehat{\mathfrak{E}}_n^{(j)}$ $(i = 1, 2; j = 1, 2, 3)$. Here, we let μ_* be the maximum of the leading terms in $\widehat{\mathfrak{B}}_n^{(i)}, \widehat{\mathfrak{D}}_n^{(i)}, \widehat{\mathfrak{S}}_n^{(i)}$, and $\widehat{\mathfrak{E}}_n^{(j)}$ $(i = 1, 2; j = 1, 2, 3)$. The results are collected in the following theorem.

Theorem 3. *Given Assumption A and $\mathcal{H}_1$,*

(i) if for all $d > 0$, $B_* \neq dA_*$,

    (i.a) the leading term of $\widehat{\mathfrak{B}}_n^{(1)}$ is equal to μ_* if and only if $\tau_*^2 > \max[\delta_*^2, \eta_*^2]$ and $\sigma_* > \gamma_*$;

    (i.b) the leading term of $\widehat{\mathfrak{B}}_n^{(2)}$ is equal to μ_* if and only if $\delta_*^2 > \max[\tau_*^2, \eta_*^2]$ and $\sigma_* > \gamma_*$;

    (i.c) if the leading term of $\widehat{\mathfrak{D}}_n^{(1)}$ is equal to μ_*, $\tau_*^2 > \max[\delta_*^2, \eta_*^2]$ and $\sigma_* = \gamma_*$, and if $\tau_*^2 > \max[\delta_*^2, \eta_*^2]$ and $\sigma_* = \gamma_*$, the leading terms of $\widehat{\mathfrak{B}}_n^{(1)}, \widehat{\mathfrak{D}}_n^{(1)}$, and $\widehat{\mathfrak{E}}_n^{(1)}$ are equal to μ_*;

    (i.d) if the leading term of $\widehat{\mathfrak{D}}_n^{(2)}$ is equal to μ_*, $\eta_*^2 > \max[\delta_*^2, \tau_*^2]$ and $\sigma_* = \gamma_*$, and if $\eta_*^2 > \max[\delta_*^2, \tau_*^2]$ and $\sigma_* = \gamma_*$, the leading terms of $\widehat{\mathfrak{D}}_n^{(2)}, \widehat{\mathfrak{S}}_n^{(2)}$, and $\widehat{\mathfrak{E}}_n^{(2)}$ are equal to μ_*;

    (i.e) the leading term of $\widehat{\mathfrak{S}}_n^{(1)}$ is equal to μ_* if and only if $\delta_*^2 > \max[\eta_*^2, \tau_*^2]$ and $\gamma_* > \sigma_*$;

    (i.f) the leading term of $\widehat{\mathfrak{S}}_n^{(2)}$ is equal to μ_* if and only if $\eta_*^2 > \max[\tau_*^2, \delta_*^2]$ and $\gamma_* > \sigma_*$;

    (i.g) the leading term of $\widehat{\mathfrak{E}}_n^{(1)}$ is equal to μ_* if and only if $\tau_*^2 > \max[\delta_*^2, \eta_*^2]$ and $\gamma_* > \sigma_*$;

    (i.h) if the leading term of $\widehat{\mathfrak{E}}_n^{(3)}$ is equal to μ_*, $\delta_*^2 > \max[\tau_*^2, \eta_*^2]$ and $\sigma_* = \gamma_*$, and if $\delta_*^2 > \max[\tau_*^2, \eta_*^2]$ and $\sigma_* = \gamma_*$, the leading terms of $\widehat{\mathfrak{B}}_n^{(2)}, \widehat{\mathfrak{E}}_n^{(3)}$, and $\widehat{\mathfrak{S}}_n^{(1)}$ are equal to μ_*; and

    (i.i) the leading term of $\widehat{\mathfrak{E}}_n^{(2)}$ is equal to μ_* if and only if $\eta_*^2 > \max[\tau_*^2, \delta_*^2]$ and $\sigma_* > \gamma_*$;

(ii) if for some $d_* > 0$, $B_* = d_*A_*$, for $i = 1, 2$ and $j = 1, 2, 3$, then $\widehat{\mathfrak{B}}_n^{(i)}, \widehat{\mathfrak{D}}_n^{(i)}, \widehat{\mathfrak{S}}_n^{(i)}, \widehat{\mathfrak{E}}_n^{(j)}$, and $\widehat{\mathfrak{M}}_n$ are equal to $\frac{nk}{2}(d_* - 1)^2 + nd_*(d_* - 1)\mathrm{tr}[L_n] + O_{\mathbb{P}}(1)$. $\square$

Remarks

(a) The conditions in Theorem 3 are summarized into two dimensional conditions, viz. the maximum condition between σ_* and γ_* and that among τ_*^2, δ_*^2, and η_*^2. Table 1 describes these relations in tabular format and provides the test statistics with the maximum leading term under each condition. According to Table 1, $\widehat{\mathfrak{D}}_n^{(1)}$, $\widehat{\mathfrak{D}}_n^{(2)}$, and $\widehat{\mathfrak{E}}_n^{(3)}$ cannot have the maximum leading term alone, although the other tests can do so, as indicated in Theorem 3. We note that each cell of Table 1 is occupied by at least one of the nine test statistics $\widehat{\mathfrak{B}}_n^{(i)}$, $\widehat{\mathfrak{D}}_n^{(i)}$, $\widehat{\mathfrak{S}}_n^{(i)}$ ($i = 1, 2$) and $\widehat{\mathfrak{E}}_n^{(j)}$ ($j = 1, 2, 3$), implying that the nine test statistics have the greatest leading term under each region, whose union is the alternative parameter space.

(b) If $k = 2$, it follows that $\sigma_* \geq \gamma_*$ from the fact that $(\delta_* + 1)^2 = (\tau_* + 1)(\eta_* + 1)$. Therefore, the leading terms of $\widehat{\mathfrak{S}}_n^{(1)}$, $\widehat{\mathfrak{S}}_n^{(2)}$, and $\widehat{\mathfrak{E}}_n^{(3)}$ cannot produce the maximum leading term.

(c) Theorem 3(*ii*) implies that all the tests are asymptotically equivalent even under $\mathcal{H}_1$ if for some $d_* > 0$,
$$B_* = d_* A_*.$$

(d) The conditions in Theorems 3(*i.a–i.i*) can be consistently selected by estimating τ_*, δ_*, σ_* and by comparing the conditions in Theorem 3. For example, if $\widehat{\tau}_n^2 > \widehat{\delta}_n^2$, $\widehat{\tau}_n^2 > \widehat{\eta}_n^2$, and $\widehat{\sigma}_n \geq \widehat{\gamma}_n$ and the sample size is reasonably large, then testing the hypotheses by relying on $\widehat{\mathfrak{B}}_n^{(1)}$ should give higher power than the other tests. This feature motivates the use of a maximal test statistic as a desirable test in terms of power. But rather than take the maximum of the nine test statistics, we construct the maximal test statistic $\widehat{\mathfrak{M}}_n$ based on the six statistics $\widehat{\mathfrak{B}}_n^{(i)}$, $\widehat{\mathfrak{S}}_n^{(i)}$, and $\widehat{\mathfrak{E}}_n^{(i)}$ ($i = 1, 2$) for the reason given above, that their leading terms are always dominated by the leading terms of the tests in Theorem 3. Finally, we do not need to test $\mathcal{H}_0^{(1)}$, $\mathcal{H}_0^{(4)}$, and $\mathcal{H}_0^{(7)}$ as mentioned earlier.

Additional Remarks

(e) Mauchly (1940), John (1971), Muirhead (1982), and Anderson (2003) test the sphericity condition: for some d_*, $B_* = d_* A_*$, and theorem 7 of CW implies that the LR test that is obtained under the same distributional condition as above is locally equivalent to the tests defined by $\widehat{\tau}_n$, $\widehat{\delta}_n$ and $\widehat{\eta}_n$. In particular, John (1971) shows that the test statistic defined by the test base $n\mathrm{tr}[(k\widehat{\Sigma}_n/\mathrm{tr}(\widehat{\Sigma}_n) - I_k)^2]$ is locally optimal, where $\widehat{\Sigma}_n$ is the k-dimensional sample covariance matrix of normally distributed data. Our results in Theorem 2 can be viewed as an extension of this. Note that if $k\widehat{\Sigma}_n/\mathrm{tr}(\widehat{\Sigma}_n)$ of the test base is viewed as a particular form of $\widehat{B}_n \widehat{A}_n^{-1}$, then from the fact that $n\mathrm{tr}[(\widehat{B}_n \widehat{A}_n^{-1} - I_k)^2] = \mathrm{tr}[(V_* + \sqrt{n}K_{o,n})^2] + o_\mathbb{P}(1)$ under $\mathcal{H}_\ell$, the test statistics here become locally optimal under the conditions in John (1971) by Theorem 2. We extend this discussion further below under the structural

14

model condition in Section 4 and also examine specific examples of testing the sphericity condition in Section 5.

(f) The local asymptotic approximations are equivalent to that of the likelihood ratio test under certain conditions. Nagao (1967), Nagarsenker and Pillai (1973), Muirhead (1982), and Anderson (2003) examine the LR statistic that tests equality of a covariance matrix to a certain matrix. We compare our test statistic by simulation in Section 5.

(g) The same test bases used here can be adapted even when high-dimensional matrices are estimated. When deriving Theorem 1, the dimension k is assumed invariant to n. If this condition is relaxed so that the dimension k_n increases such that $c_n := k_n/n \to c \in (0, 1)$, we obtain a result different from the earlier case. For example, for normally and independently distributed k_n-dimensional data, if we let $\widehat{B}_n$ and $\widehat{A}_n$ be the covariance matrix estimator and I_{k_n}, respectively, it follows that

$$n\widehat{T}_n - \mu_c \overset{\mathrm{A}}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2/c & 2/c \\ 2/c & -2\log(1-c)/c^2 \end{bmatrix}\right)$$

under the null hypothesis that the population covariance matrix is I_{k_n}, where $T_n := [\widehat{\tau}_n, (\log(\widehat{\delta}_n + 1) - m_n)]'$, $\mu_c := (0, \log(1-c)/(2c))'$, and $m_n := ((1 - 1/c_n)\log(1 - c_n) - 1)$. Thus, if we let $\widehat{\mathfrak{W}}_n := (n\widehat{T}_n - \widehat{\mu}_{c_n})'\widehat{\Sigma}_n^{-1}(n\widehat{T}_n - \widehat{\mu}_{c_n})$ to test $\tau_* = 0$ and $\delta_* = 0$, then it follows from theorem 3.1 of Bai, Jiang, Yao, and Zheng (2009) that $\widehat{\mathfrak{W}}_n \overset{\mathrm{A}}{\sim} \mathcal{X}_2^2$ under the null, where

$$\widehat{\mu}_{c_n} := \begin{pmatrix} 0 \\ \log(1 - c_n)/(2c_n) \end{pmatrix}, \quad \text{and} \quad \widehat{\Sigma}_n := \begin{pmatrix} 2/c_n & 2/c_n \\ 2/c_n & -2\log(1 - c_n)/c_n^2 \end{pmatrix}.$$

Although the test statistic $\widehat{\mathfrak{W}}_n$ is defined in a form different from the standard tests, it is defined using the same test factors $\widehat{\tau}_n$ and $\widehat{\delta}_n$ as before. This shows that our test bases can be adapted to estimating and testing high-dimensional matrices. The corrected version of the likelihood ratio test defined by Bai, Jiang, Yao, and Zheng (2009) is also indeed a linear function of $\widehat{T}_n$.

4 Application to SVAR Model Estimation

In this section, we examine the LR test statistic using FIML estimation in the SVAR model context where tests for over-identification serve as a prototypical example of testing two equal symmetric positive-definite matrices. The example is motivated by the fact that the LR test statistic turns out to be a locally most pow-

erful test since FIML estimation is conducted using a correctly specified distributional model assumption. Hence, we can compare the local power properties of the tests in the current study with that of the LR test statistic.

Specifically, we work with the AB-SVAR model in Amisano and Giannini (1997) which is a synthetic generalization of popular SVAR models: for $k \times k$ invertible matrices $H_{n,*}$ and $N_{n,*}$, $H_{n,*}\phi_*(L)Y_{n,t} = H_{n,*}U_{n,t}$ and $H_{n,*}U_{n,t} = N_{n,*}W_{n,t}$ such that $E[W_{n,t}] = 0$ and $E[W_{n,t}W'_{n,t}] = I$, where $U_{n,t} \sim$ IID $N(0, B_{n,*})$, $\phi_*(L) := I - \phi_{1,*}L - \phi_{2,*}L^2 - \ldots - \phi_{p,*}L^p$, and L is the lag operator. The structural parameter matrix $H_{n,*}$ and $N_{n,*}$ are estimated by maximizing the log-likelihood function

$$\mathcal{L}_n(A) := -\frac{nk}{2}\log(2\pi) - \frac{n}{2}\log(\det(A)) - \frac{n}{2}\mathrm{tr}(A^{-1}\widehat{B}_n)$$

with respect to H and N such that $A = H^{-1}NN'H'^{-1}$, where $\widehat{B}_n := \frac{1}{n}\sum_{t=1}^{n}\widehat{U}_n\widehat{U}'_t$, $\widehat{U}_t := \widehat{\phi}_n(L)Y_t$, and $\widehat{\phi}_n(L)$ is the LS estimator obtained by regressing Y_t on $(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})$. Note that $\mathcal{L}_n(\cdot)$ is under-identified: there are $2k^2$ unknowns in $H_{n,*}$ and $N_{n,*}$, whereas there are $k(k+1)/2$ first-order equations, so that there are $2k^2 - k(k+1)/2$ free parameters. Thus, the structural parameters $H_{n,*}$ and $N_{n,*}$ are estimated by re-parameterizing them as functions of another parameter ψ that satisfies order and rank conditions for identification (*e.g.*, Sargan, 1988), so that $(H_{n,*}, N_{n,*})$ can be represented as $(H_n(\psi_*), N_n(\psi_*))$ for some $H_n(\cdot)$ and $N_n(\cdot)$, and ψ_* is instead estimated by maximizing the log-likelihood function with respect to ψ. We denote the SVAR estimator obtained in this way as $(\widehat{H}_n, \widehat{N}_n) := (H_n(\widehat{\psi}_n), N_n(\widehat{\psi}_n))$ and $\widehat{A}_n := \widehat{H}_n^{-1}\widehat{N}_n\widehat{N}'_n\widehat{H}_n'^{-1}$ to estimate $A_{n,*} := H_{n,*}^{-1}N_{n,*}N'_{n,*}H_{n,*}'^{-1}$, where $\widehat{\psi}_n$ is the argument maximizing the log-likelihood function with respect to ψ.

Model over-identification is often tested using the LR test statistic. If the model is exactly identified, the log-likelihood is

$$\mathcal{L}_n(\widehat{B}_n) = -\frac{nk}{2}\log(2\pi) - \frac{n}{2}\log(\det(\widehat{B}_n)) - \frac{nk}{2},$$

so that the LR test statistic for over-identification is obtained as $\mathfrak{LR}_n := 2(\mathcal{L}_n(\widehat{B}_n) - \mathcal{L}_n(\widehat{A}_n)) = nk(\widehat{\tau}_n - \widehat{\lambda}_n)$, where $\widehat{\lambda}_n := \frac{1}{k}\log(\det[\widehat{D}_n])$. That is, the LR test statistic measures the distance between $\widehat{\tau}_n$ and $\widehat{\lambda}_n$, which is a function of $\det(\widehat{D}_n)$, were $\widehat{D}_n = \widehat{B}_n\widehat{A}_n^{-1}$ as before, and thereby statistically tests equality between $A_{n,*}$ and $B_{n,*}$.

We provide the following set of assumptions to formalize conditions.

Assumption C (SVAR). *(i) For every $n \in \mathbb{N}$, $\phi_*(L)Y_{n,t} = U_{n,t} \sim$ IID $N(0, B_{n,*})$ such that the roots of $\det(\phi_*(L)) = 0$ lie outside of the unit circle, and $B_{n,*} = B_* + n^{-1/2}\bar{B}_*$, where B_* and $\bar{B}_*$ are symmetric positive-definite matrices in $\mathbb{R}^{k \times k}$;*

16

(ii) For every $n \in \mathbb{N}$, $H_{n,*}$ and $N_{n,*}$ are invertible such that $A_{n,*} := H_{n,*}^{-1} N_{n,*} N_{n,*}' H_{n,*}'^{-1} = A_* + n^{-1/2} \bar{A}_*$ with $\bar{A}_* \neq \bar{B}_*$, where A_* and $\bar{A}_*$ are symmetric and positive definite;

(iii) $A : \Theta \mapsto \mathbb{R}^{k \times k}$ is in $\mathcal{C}^{(2)}(\Theta)$ and $A_* := A(\theta_*)$, where $\Theta(\in \mathbb{R}^\ell)$ is a compact and convex parameter space of $\theta := (\psi', \mathrm{vec}(\phi_1, \ldots, \phi_p)')'$ that contains $\theta_* := (\psi_*', \mathrm{vec}(\phi_{1*}, \ldots, \phi_{p*})')'$ as an interior element;

(iv) $\bar{A} : \Theta \mapsto \mathbb{R}^{k \times k}$ and $\bar{B} : \Theta \mapsto \mathbb{R}^{k \times k}$ are in $\mathcal{C}^{(1)}(\Theta)$, $\bar{A}_* := \bar{A}(\theta_*)$, and $\bar{B}_* := \bar{B}(\theta_*)$;

(v) $H_n : \Theta \mapsto \mathbb{R}^{k \times k}$ and $N_n : \Theta \mapsto \mathbb{R}^{k \times k}$ are in $\mathcal{C}^{(2)}(\Theta)$, $A_n(\cdot) := H_n(\cdot)^{-1} N_n(\cdot) N_n(\cdot)' H_n(\cdot)'^{-1}$ is consistent for $A(\cdot)$ uniformly on Θ, $H_{n,*} = H_n(\theta_*)$, and $N_{n,*} = N_n(\theta_*)$;

(vi) $\sqrt{n}[(\widehat{\theta}_n - \theta_*)', \mathrm{vech}[A_n - A_{*,n}]']' = O_{\mathbb{P}}(1)$;

(vii) for $j = 1, \ldots, \ell$, $\partial_j A_n(\cdot)$ is consistent for $\partial_j A(\cdot)$ uniformly on Θ; and

(viii) for $j = 1, \ldots, \ell$, $H_{j,o,n} := A_*^{-1} \partial_j (A_n - A_{*,n}) = O_{\mathbb{P}}(n^{-1/2})$. □

Remarks

(a) Assumption C is obtained from Assumption B by accommodating the SVAR features, so that the definitions of $\widehat{\phi}_n(L)$, $B_{n,*}$, and $\widehat{B}_n$ correspond with this and the conditions in Assumption B for $\widehat{\phi}_n(L)$, $B_{n,*}$, and $\widehat{B}_n$ are easily affirmed by these definitions. On the other hand, $A_{n,*}$ and $\widehat{A}_n$ cannot be uniformly defined. They are differently defined, depending on the order and rank conditions. Hence, Assumption C focuses on the regularity conditions for $\widehat{A}_n(\cdot)$ and $A_{n,*}$.

(b) The roots of $\det(\phi_*(L)) = 0$ lying outside the unit circle ensures stationarity of Y_t.

(c) Even when $A_{n,*} = A_* + n^{-1/2} \bar{A}_* + o(n^{-1/2})$ or $B_{n,*} = B_* + n^{-1/2} \bar{B}_* + o(n^{-1/2})$, the results stated below still hold. For brevity, we omit the $o(n^{-1/2})$ remainders from $A_{n,*}$ and $B_{n,*}$.

(d) The specific roles of the order and rank conditions given implicitly in $H_n(\cdot)$ and $N_n(\cdot)$ are not highlighted as our interests relate to testing for over-identification using $A_{n,*}$. Instead, we directly impose the regularity conditions for $A(\cdot)$ and $\bar{A}(\cdot)$ in Assumption C(*iii and iv*).

(e) Although $A(\cdot)$ and $\bar{A}(\cdot)$ are functions of only ψ, we treat them as functions of θ to comply with the theory of the previous section. For the same reason, $\bar{B}(\cdot)$ is also treated as a function of θ, although it is a function of only $\mathrm{vec}(\phi_1, \ldots, \phi_p)$.

(f) Most SVAR models suppose the model is locally identified, and linear re-parameterizations of ψ with deterministic coefficients are assumed for $H_n(\cdot)$ and $N_n(\cdot)$. For such cases, Assumptions C(*v and vi*) trivially hold, although $H_n(\cdot)$ and $N_n(\cdot)$ do not necessarily have to be linear transformations of ψ.

Finally, Assumption C is obtained by modifying Assumption B to fit the features of the SVAR model, so that the consequences in Theorem 2 are also valid for $\widehat{A}_n$ and $\widehat{B}_n$ of this section.

(g) The given likelihood function can be related to contiguity between the distributions generated by $B_{n,*}$ and $A_{n,*}$. If we let $\widehat{B}_n := n^{-1}\sum_{t=1}^n U_{n,t}U'_{n,t}$ and $A_* = B_*$, $\mathcal{L}_n(B_{n,*}) - \mathcal{L}_n(A_{n,*}) \overset{A}{\sim} N(-\frac{1}{4}\mathrm{tr}[V_*^2], \frac{1}{2}\mathrm{tr}[V_*^2])$ under $U_{n,t} \sim$ IID $N(0, A_{n,*})$, implying that the probability distributions generated by $B_{n,*}$ and $A_{n,*}$ are mutually contiguous and $\mathcal{L}_n(B_{n,*}) - \mathcal{L}_n(A_{n,*}) \overset{A}{\sim} N(\frac{1}{4}\mathrm{tr}[V_*^2], \frac{1}{2}\mathrm{tr}[V_*^2])$ under $U_{n,t} \sim$ IID $N(0, B_{n,*})$ by Le Cam's first and third lemmas. We show this property in the Supplement. If we further suppose that $\bar{A}_* = 0$ and test $B_* = A_*$ at a 5% significance level using two-sided testing, the power envelope $\mathfrak{P}_e$ is obtained as $\mathbb{P}(Z > 1.96 - \sqrt{\mathrm{tr}[F_*^2]/2}) + \mathbb{P}(Z < -1.96 - \sqrt{\mathrm{tr}[F_*^2]/2})$, where $Z \sim N(0,1)$. Below, we below conduct Monte Carlo simulations using this feature to explore the power envelope.

The following theorem gives the local limit approximation of the LR test statistic.

Theorem 4. *Given Assumptions C and $\mathcal{H}_\ell$, $\mathfrak{LR}_n = \frac{1}{2}\mathrm{tr}[(V_* + \sqrt{n}K_{o,n})^2] + o_\mathbb{P}(1)$.* $\qquad\square$

Remarks

(a) The asymptotic result in Theorem 4 corresponds to that of Theorem 2, so that the LR test statistic is asymptotically equivalent to $\widehat{\mathfrak{B}}_n^{(i)}$, $\widehat{\mathfrak{D}}_n^{(i)}$, $\widehat{\mathfrak{S}}_n^{(i)}$, $\widehat{\mathfrak{E}}_n^{(i)}$ $(i = 1, 2)$, and $\widehat{\mathfrak{M}}_n$ under the local alternative. Thus, if the LR test statistic turns out locally asymptotically optimal, as for the standard case where a correct distributional condition applies along with minor regularity conditions, the test statistics of this study must be also be locally asymptotically optimal.

(b) Theorem 4 is shown by noting that $\widehat{\delta}_n$ and $\widehat{\lambda}_n$ posit similar asymptotic behavior, although their difference is not exactly zero but rather $\widehat{\delta}_n - \widehat{\lambda}_n = (2k^2)^{-1}\mathrm{tr}[(K_{o,n} + n^{-1/2}V_*)]^2 + o_\mathbb{P}(n^{-1})$, so that $\widehat{\delta}_n$ is asymptotically greater than $\widehat{\lambda}_n$.

(c) $V_* = 0$ under $\mathcal{H}_0$, and all of $\widehat{\mathfrak{B}}_n^{(i)}$, $\widehat{\mathfrak{D}}_n^{(i)}$, $\widehat{\mathfrak{S}}_n^{(i)}$ $(i = 1, 2)$, $\widehat{\mathfrak{E}}_n^{(i)}$ $(i = 1, 2, 3)$, and $\widehat{\mathfrak{M}}_n$ are asymptotically chi-squared under $\mathcal{H}_0$, just as $\mathfrak{LR}_n$ is chi-squared under $\mathcal{H}_0$.

Before moving to the next section, we add comments about extensions. The analysis given in this section can be easily adapted to other structural models. As Bartlett (1954, IIIa) shows, testing a hypothetical covariance matrix often involves testing the difference between $\widehat{\tau}_n$ and $\widehat{\lambda}_n$, as for the LR test statistic in Theorem 4. Therefore, if the structural assumption is embedded through the covariance matrix hypothesis,

this implies that the structural assumption can be tested in a similar way to the LR test statistic under the local alternative as in Theorem 4. For example, the standard multi-factor model often tests the number of factors by means of a structural model that assumes a normal distribution. If we let $\widehat{A}_n$ and $\widehat{B}_n$ be the ML estimators obtained under the null of a factor model constraint and an alternative, respectively, the standard LR test is written as $nk(\widehat{\tau}_n - \widehat{\lambda}_n)$ (*e.g.*, Geweke and Singleton, 1980), so that we can analyze its local limit distribution in a similar way to Theorem 4. It is also possible to test the factor model hypothesis without imposing the normal distribution assumption using the test statistics of the current study.

5   Monte Carlo Experiments

This section reports Monte Carlo experiments examining the performance of the tests analyzed in the previous section. We examine two applications. First, we assume a correctly specified model for the data generating process (DGP) and compare the performance of our tests with the power envelope of the LR test. Second, we consider application of residual bootstrapping to linear two stage least squares (TSLS) estimation (*e.g.* Efron and Tibshirani, 1993). Residual bootstrapping can also be applied to other GMM estimators.

5.1   Testing Using the Null Limit Distribution

Suppose we have the following DGP $Y_t := (y_{1t}, y_{2t})' \sim \text{IID } N(0, B_{n,*})$, where $B_{n,*} := I_2 + n^{-1/2}\bar{B}_*$. This simple DGP is assumed to enable comparison of the power of the tests in this paper with the power envelope. The Neyman-Pearson lemma implies that the LR test statistic is the most powerful test statistic for the following simple hypotheses: $H_0^\dagger : \bar{B}_* = 0$ versus $H_1^\dagger : \bar{B}_* = \Omega_*$, thereby yielding the power envelope, where

$$\Omega_* := \begin{bmatrix} a_* & d_* \\ d_* & c_* \end{bmatrix}.$$

We compare the powers of the test statistics in the current study with this power envelope.

Our test comparison is conducted as follows. First, we obtain the asymptotic power envelope by approximating the LR test statistic that tests the simple hypotheses:

$$\mathfrak{LR}_n^\dagger := 2(\mathcal{L}_n(B_{n,*}) - \mathcal{L}_n(I_2)) = \frac{1}{\sqrt{n}}\left[a_*\sum_{t=1}^n(y_{1t}^2 - 1) + 2d_*\sum_{t=1}^n y_{1t}y_{2t} + c_*\sum_{t=1}^n(y_{2t}^2 - 1)\right] - \frac{1}{2}\gamma_* + o_\mathbb{P}(1),$$

where $B_{n,*} = I_2 + n^{-1/2}\Omega_*$ and $\gamma_* := a_*^2 + c_*^2 + 2d_*^2$. Here, $\mathcal{L}_n(\cdot)$ is the same as above, based on $\widehat{B}_n :=$

$n^{-1}\sum Y_t Y_t'$. Note that $\mathfrak{LR}_n^\dagger$ is asymptotically constructed from the sufficient statistics $(\sum y_{1t}^2, \sum y_{1t}y_{2t}, \sum y_{2t}^2)$. Given this approximation, $\mathfrak{LR}_n^\dagger \overset{A}{\sim} N(-0.5\gamma_*, 2\gamma_*)$ under $H_0^\dagger$, and $\mathfrak{LR}_n^\dagger \overset{A}{\sim} N(0.5\gamma_*, 2\gamma_*)$ under $H_1^\dagger$ by Le Cam's first lemma, and the power envelope $\mathfrak{P}_e$ is obtained as $\mathbb{P}(Z > 1.96 - \sqrt{\gamma_*/2}) + \mathbb{P}(Z < -1.96 - \sqrt{\gamma_*/2})$ by two-sided testing at the significance level 5%, where $Z \sim N(0,1)$. Second, we apply the LR test statistic to test $H_0^\ddagger : \bar{B}_* = 0$ against the composite hypothesis $H_1^\ddagger : \bar{B}_* \neq 0$, leading to $\mathfrak{LR}_n^\ddagger := 2(\mathcal{L}_n(\widehat{B}_n) - \mathcal{L}_n(I_2))$, which follows a $\mathcal{X}_3^2$ distribution under $H_0^\ddagger$. Note that $\mathfrak{LR}_n^\ddagger$ differs from $\mathfrak{LR}_n^\dagger$ since nuisance parameters are estimated under the alternative. For this case, $\mathfrak{LR}_n^\ddagger$ is asymptotically equivalent to a locally optimal test statistic. Finally, we implement our tests using $\widehat{D}_n = \widehat{B}_n \widehat{A}_n^{-1}$ and $\widehat{A}_n = I_2$. All of $\widehat{\mathfrak{B}}_n^{(1)}, \widehat{\mathfrak{B}}_n^{(2)}, \widehat{\mathfrak{S}}_n^{(1)}, \widehat{\mathfrak{S}}_n^{(2)}, \widehat{\mathfrak{E}}_n^{(1)}, \widehat{\mathfrak{E}}_n^{(2)}$, and $\widehat{\mathfrak{M}}_n$ asymptotically follow a $\mathcal{X}_3^2$ distribution under $\mathcal{H}_0^\ddagger$ by Corollary 1, and they are asymptotically equivalent under $H_1^\ddagger$ by Theorem 2. Furthermore, Theorem 4 implies that $\mathfrak{LR}_n^\ddagger$ is equivalent to all of $\widehat{\mathfrak{B}}_n^{(1)}, \widehat{\mathfrak{B}}_n^{(2)}, \widehat{\mathfrak{S}}_n^{(1)}, \widehat{\mathfrak{S}}_n^{(2)}, \widehat{\mathfrak{E}}_n^{(1)}, \widehat{\mathfrak{E}}_n^{(2)}$, and $\widehat{\mathfrak{M}}_n$ under both $H_0^\ddagger$ and $H_1^\ddagger$.

For our simulation, we let $a_* = -h$, $d_* = h$, and $c_* = -h$ and generate DGPs close to or remote from the null DGP by letting $h \in \{0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0\}$. Using these DGPs, the powers of the test statistics are compared by repeating independent experiments 10,000 times with 5,000 observations. Table 2 collects the empirical levels and powers of the test statistics when the level of significance is 5%. We summarize the simulation results as follows.

1. All test statistics exhibit empirical levels close to the level of significance under $H_0^\ddagger$, affirming Corollary 1.

2. All of $\widehat{\mathfrak{B}}_n^{(1)}, \widehat{\mathfrak{B}}_n^{(2)}, \widehat{\mathfrak{S}}_n^{(1)}, \widehat{\mathfrak{S}}_n^{(2)}, \widehat{\mathfrak{E}}_n^{(1)}, \widehat{\mathfrak{E}}_n^{(2)}$, and $\widehat{\mathfrak{M}}_n$ show similar empirical rejection rates under both $H_0^\ddagger$ and $H_1^\ddagger$, affirming Theorem 2.

3. The power pattern of $\mathfrak{LR}_n^\ddagger$ exhibits a pattern similar to $\widehat{\mathfrak{M}}_n$, affirming Theorem 4. But the finite sample power of $\widehat{\mathfrak{M}}_n$ is superior to $\mathfrak{LR}_n^\ddagger$, which shows its practical advantage over $\mathfrak{LR}_n^\ddagger$.

4. Finally, the power pattern of $\widehat{\mathfrak{M}}_n$ is lower than than the asymptotic power envelope, although its power curve is closer to the power envelope than the other tests.

We next modify the distributional assumption and conduct another simulation. That is, we let $Y_t := B_*^{1/2} C_t$ and $B_* := I_2 + \bar{B}_*$, where $C_t := (w_{1t}^2 - 1, w_{2t}^2 - 1)'/\sqrt{2}$ and $W_t := (w_{1t}, w_{2t})' \sim$ IID $N(0, I_2)$. We test $H_0^\ddagger$ versus $H_1^\ddagger$ and also assume that this distributional assumption is unknown to the researcher, so that the LR test statistic cannot be used here and the parameter Σ_* in Corollary 1 must be estimated to obtain the null limit distribution of the test statistics in this study. We estimate it by the method of moments and

include the power patterns in Table 3 for $h \in \{0.0, 0.02, 0.04, 0.06, 0.08, 0.10\}$. The other conditions are the same as for Table 2. We summarize the simulation results as follows.

1. All test statistics exhibit empirical levels close to the level of significance under $H_0^\ddagger$, affirming Corollary 1.

2. The empirical powers of all tests approach unity as h increases, as implied by the omnibus power feature of the test statistics.

3. The power of $\widehat{\mathfrak{M}}_n$ always exceeds that of the other tests, which is implied by the fact that $\widehat{\mathfrak{M}}_n$ is the most powerful test statistic.

5.2   Residual Bootstrapping

As Horowitz (1994) noted, asymptotic theory may encounter difficulty in testing the equality of two positive-definite matrices because many tests suffer from size distortions. The information matrix equality test is a typical example. Horowitz (1994) showed that many such equality test methodologies can be implemented with little size distortion if the parametric bootstrap is applied. Accordingly, CW applied their test statistics for information matrix equality using the parametric bootstrap. This section extends the applicability of our tests by using the residual bootstrap, which does not require distributional information on the observations like the parametric bootstrap.

To illustrate suppose that the object is to estimate an unknown parameter using an asymptotically optimal weight matrix in GMM to ensure that the GMM estimator is asymptotically efficient. We propose a test to determine whether the weight matrix selected by the researcher is asymptotically optimal. If the selected weight matrix is optimal, the sandwich asymptotic covariance matrix of the GMM estimator simplifies, and we can test this feature directly using the testing methodology of the present study. Specifically, when $\widehat{\theta}_n$ is the GMM estimator obtained by minimizing $g_n(\cdot)' W_n g_n(\cdot)$ under the standard model and the DGP assumption (*e.g.*, Hamilton, 1994, chap. 14), it is well known that $\sqrt{n}(\widehat{\theta}_n - \theta_*) \overset{\text{A}}{\sim} N[0, (G'WG)^{-1}(G'WSWG)(G'WG)^{-1}]$, where θ_* is the probability limit of the GMM estimator, W is the probability limit of W_n, S is the asymptotic covariance matrix of $\sqrt{n} g_n(\theta_*)$, and G is the probability limit of $\nabla_\theta g_n(\theta_*)$. The researcher wishes to select W_n to estimate S^{-1} asymptotically, so that $\sqrt{n}(\widehat{\theta}_n - \theta_*) \overset{\text{A}}{\sim} N[0, (G'S^{-1}G)^{-1}]$.

In this general set up, we further suppose that the researcher assumes conditional homoskedasticity and we proceed with simulations on this basis: viz., for some $\sigma_*^2 > 0$ and positive-definite matrix Q, $S = \sigma_*^2 Q$. The conditional homoskedasticity assumption is violated if the error distribution is not conditionally

homoskedastic, the model assumption is incorrect, or the error is serially correlated. We therefore test the proportionality condition between the two matrices and use the testing outcome as a diagnostic for conditional heteroskedasticity, model misspecification, or autocorrelation. This hypothesis is a specific example of the sphericity condition.

As a particular example of this GMM framework, we consider TSLS estimation. Specifically, for the linear model $y_t = X_t'\beta + u_t$, where $X_t := (1, x_t)'$, we obtain the TSLS estimator $\widetilde{\beta}_n := (X'PX)^{-1}(X'PY)$ and residual $\widetilde{u}_t := y_t - X_t'\widetilde{\beta}_n$, where $Z_t := (1, z_t)'$ is the instrument, $P := Z(Z'Z)^{-1}Z'$ is the projection matrix onto the space of the instrument matrix $Z' = [z_1, ..., z_n]$, and $X' = [X_1, ..., X_n]$. The (conditionally homoskedastic) variance σ_*^2 is estimated by $\widetilde{\sigma}_n^2 := n^{-1}\sum_{t=1}^n \widetilde{u}_t^2$. The following residual bootstrap algorithm is implemented.

1. Step 1: Let $\widehat{A}_n := \widetilde{\sigma}_n^2(n^{-1}\sum_{t=1}^n Z_t Z_t')$ and $\widehat{B}_n$ be a heteroskedasticity autocorrelation consistent (HAC) covariance estimator of the asymptotic variance matrix of $n^{-1/2}\sum u_t Z_t$. Using $\widehat{A}_n$ and $\widehat{B}_n$, we compute the test statistics.

2. Step 2: We randomly draw $\{\widetilde{u}_t^b : t = 1, 2, \ldots, n\}$ from $\{\widetilde{u}_t : t = 1, 2, \ldots, n\}$ with replacement and let $\widehat{B}_n^b$ be the corresponding HAC estimator (constructed in the same fashion as $\widehat{B}_n$) for the asymptotic covariance of $n^{-1/2}\sum \widetilde{u}_t^b Z_t$. Using $\widehat{A}_n$ and $\widehat{B}_n^b$, we compute the test statistics.

3. Step 3: Replicate Step 2 many times and compute the percentage of bootstrapped test statistics greater than the tests. If this percentage is less than the significance level, we reject the null.

Note that if the independent draws of Step 2 $\{\widetilde{u}_t^b\}$ are conditionally homoskedastic given $\{Z_1, \ldots, Z_n\}$, the sphericity condition holds between the probability limits $\widehat{A}_n$ and $\widehat{B}_n^b$, and the null distribution of the test statistic is accordingly obtained. For our Monte Carlo simulations, the commonly used Bartlett-Newey-West HAC estimator is used in $\widehat{B}_n$ and $\widehat{B}_n^b$.

Our Monte Carlo experiments are conducted under null, local, and alternative DGPs. The following is the null DGP.

- Null: $y_t = \frac{1}{2}x_t + u_t$, $x_t := u_t + z_t$, and $(z_t, u_t)' \sim$ IID $N(0, I_2)$.

Note that the TSLS estimator is consistent for the true parameter vector $(0, \frac{1}{2})'$ because the model is correctly specified. Furthermore, the error is conditionally homoskedastic and serially uncorrelated. Therefore, the asymptotic limits of $\widehat{A}_n$ and $\widehat{B}_n$ are identical. Next, we consider three alternative DGPs.

- ALT1: $y_t = \frac{1}{2}x_t + (1 + \exp(x_t))u_t$, $x_t := u_t + z_t$, and $(z_t, u_t)' \sim$ IID $N(0, I_2)$;

22

- ALT2: $y_t = \frac{1}{2}x_t^4 + u_t$, $x_t := u_t + z_t$, and $(z_t, u_t)' \sim$ IID $N(0, I_2)$; and

- ALT3: $y_t = \frac{1}{2}x_t + u_t$, $x_t := z_t + \varepsilon_t$, $u_t := \frac{1}{10}u_{t-1} + \frac{1}{10}u_{t-2} + \varepsilon_t + \frac{1}{10}\varepsilon_{t-1}$, and $(z_t, \varepsilon_t)' \sim$ IID $N(0, I_2)$.

Note that ALT1, ALT2, and ALT3 possess conditional heteroskedasticity, model misspecification, and serially correlated errors, respectively, so that the estimates $\widehat{A}_n$ and $\widehat{B}_n$ have different asymptotic limits. Finally, we modify the given alternative DGPs into the following local alternative DGPs.

- LOC1: $y_t = \frac{1}{2}x_t + (1 + n^{-1/2}\exp(x_t)/2)u_t$, $x_t := u_t + z_t$, and $(z_t, u_t)' \sim$ IID $N(0, I_2)$;

- LOC2: $y_t = \frac{1}{10}n^{-1/2}x_t^4 + u_t$, $x_t := u_t + z_t$, and $(z_t, u_t)' \sim$ IID $N(0, I_2)$; and

- LOC3: $y_t = \frac{1}{2}x_t + u_t$, $x_t := z_t + \varepsilon_t$, $u_t := \rho_n(u_{t-1} + u_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$, $\rho_n := \frac{3}{4}n^{-1/2}$, and $(z_t, \varepsilon_t)' \sim$ IID $N(0, I_2)$.

The simulation results are given in Tables 4 to 10 and we summarize as follows. First, the null simulation results are reported in Table 4. As n increases, the empirical rejection rates converge to the nominal level. If n is small, say 50, the empirical rejection rates undervalue the nominal level for the seven test statistics. These discrepancies quickly disappear as n increases. The power and local power simulation results are reported in Tables 5 to 7 and Tables 8 to 10, respectively. We summarize the power results as follows.

1. When n increases, the empirical rejection rates converge to unity.

2. The most powerful test depends on the DGP and no test is uniformly most powerful.

3. The $\widehat{\mathfrak{M}}_n$ test possesses very respectable power overall.

The local power simulation results are summarized as follows.

1. As n increases, the empirical rejection rates converge to levels greater than 5%. This shows that the tests have nontrivial local powers.

2. $\widehat{\mathfrak{B}}_n^{(i)}$, $\widehat{\mathfrak{S}}_n^{(i)}$, $\widehat{\mathfrak{E}}_n^{(i)}$ ($i = 1, 2$), and $\widehat{\mathfrak{M}}_n$ exhibit approximately similar local powers as Theorem 4 predicts.

The simulations confirm theory, showing the tests of this study have nontrivial local discriminatory power against conditional heteroskedasticity, model misspecification, and autocorrelation.

One final comment. In addition to the simulations provided in this study, we also conducted many test procedures under numerous alternative environments and found that: (*i*) the test statistics obtained

by reversing the roles of A_* and B_* show similar or lower power patterns than those considered for the simulations reported above; (*ii*) the power patterns of $\widehat{\mathfrak{D}}_n^{(i)}$ ($i = 1, 2$) and $\widehat{\mathfrak{E}}_n^{(3)}$ are similar to one of the test statistics we consider under the fixed alternative, as expected from earlier discussion in view of Table 1.

6 Conclusion

The information matrix equality is a fundamental feature of correct specification in likelihood-based econometric work, and in GMM estimation the asymptotic variance matrix of the estimator takes a simple form if the models are correct and the errors are conditionally homoskedastic and serially uncorrelated. Our results provide a new methodology for testing such equalities in empirical applications. The approach is embedded in the general framework of testing the equality of two symmetric positive-definite matrices. The new approach improves earlier analytic attempts to control size in information matrix equality testing and provides tests for optimal weight matrix conditions in GMM estimation, delivering a class of test procedures that are easily implemented in practical work. The test mechanism extends earlier test statistics developed in the literature by exploiting a simple characterization of equality between two k dimensional symmetric positive-definite matrices A and B involving only the traces of the two matrices AB^{-1} and BA^{-1}, leading to a group of useful new omnibus test statistics and a recommended new procedure for testing the equality of covariance matrices.

Asymptotic theory for these tests under null, local, and alternatives are obtained under mild regularity conditions that support wide use of these procedures in empirical work including structural econometric models such as SVAR and multi-factor models. Our test factors can also be adapted even to high-dimensional matrix estimation settings. Simulation evidence shows that good size control is obtained and test power in testing optimal GMM estimation against various alternatives is generally strong.

Acknowledgements

References

Abadir, K. and Magnus, J., 2005, Matrix algebra. Cambridge university press, New York.

Amisano, G. and Giannini, C., 1997, Topics in structural VAR ecnometrics, 2nd edition. Springer-Verlag, Berlin/Heidelberg.

Anderson, T., 2003, An introduction to multivariate statistical analysis. John Wiley & Sons, Inc, Hoboken, NJ.

Baek, Y.I., Cho, J.S., and Phillips, P.C.B., 2015, Testing linearity using power transforms of regressors, Journal of Econometrics 187, 376–384.

Bai, Z., Jiang, D., Yao, J.-F., and Zheng, S., 2009, Correction to LRT on large-dimensional covariance matrix by RMT, Annals of Statistics 37, 3822–3840.

Bartlett, M., 1954, A note on the multiplying factors for various χ^2 approximations, Journal of the Royal Statistical Society, series B (methodological) 16, 296–298.

Bera, A. and Hall, A., 1991, An eigenvalue based test for heteroskedasticity, Department of economics, University of Illinois, mimeo.

Bernstein, D., 2005, Matrix mathematics. Princeton university press, Princeton, NJ.

Bozdogan, H., 2000, Akaike's information criterion and recent developments in information complexity, Journal of Mathematical Psychology 44, 62–91.

Chesher, A. and Spady, R., 1991, Asymptotic expansions of the information matrix test statistic, Econometrica 59, 787–815.

Cho, J. S. and White, H., 2011, Generalized runs tests for the IID hypothesis, Journal of Econometrics 162, 326–344.

Cho, J. S. and White, H., 2014, Testing the equality of two positive-definite matrices with application to information matrix testing, in: Yoosoon Chang, Thomas B. Fomby, and Joon Y. Park (Eds.), Advances in econometrics: Essays in honor of Peter C. B. Phillips, Vol. 33. Emerald group publishing limited, West Yorkshire, UK, pp. 491–556.

Dhaene, G. and Hoorelbeke, D., 2004, The information matrix test with bootstrap-based covariance matrix estimation, Economics Letters 82, 341–347.

Efron, B. and Tibshirani, R., 1993, Introduction to the bootstrap, Monographs on statistics & applied probability 57. Chapman and Hall/CRC, New York.

Geweke, J., and Singleton, 1925, Interpreting the likelihood ratio statistic in factor models when sample Size is small, Journal of the American Statistical Association 75, 133-137.

Golden, R., Henley, S., White, H., and Kashner, T., 2013, New directions in information matrix testing: eigenspectrum tests, in Norman Rasmus Swanson and Xiaohong Chen, (Eds.), Causality, prediction, and specification analysis: Recent advances and future directions essays in honour of Halbert L. White, Jr., Springer, New York, pp. 145–177.

Hall, A., 1987, The information matrix test for the linear model, Review of Economic Studies 54, 257–263.

Hamilton, J., 1994, Time series analysis. Princeton university proess, NJ.

Horowitz, J., 1994, Bootstrap-based critical values for the information matrix test, Journal of Econometrics 61, 395–411.

Hong, S.Y., Linton, O.B. and Zhang, H.J., 2017, An investigation into multivariate variance ratio statistics and their application to stock market predictability, Journal of Financial Econometrics 15, 173-222.

James, A. T., 1964, Distributions of matrix variates and latent roots derived from normal samples, Annals of Mathematical Statistics 35, 475–501.

John, S., 1971, Some optimal multivariate tests, Biometrika 58, 123–127.

Magnus, J., 1985, On differentiating eigenvalues and eigenvectors, Econometric Theory 1, 179–191.

Marshall, A.W., Olkin, I., and Arnold, B. C., 1979, Inequalities: Theory of majorization and its applications, Springer series in statistics, Springer, New York.

26

Mauchly, J., 1940, Significance test for sphericity of a normal n-variable distribution, Annals of Mathematical Statistics 11, 204–209.

Muirhead, R., 1982, Aspects of multivariate statistical theory. John Wiley & Sons, Inc, New York.

Nagao, H., 1967, Monotonicity of the modified likelihood ratio test for a covariance matrix, Journal of Science of the Hiroshima University, series A-I 31, 147–150.

Nagarsenker, B. and Pillai, K., 1973, The distribution of the sphericity test criterion, Journal of Multivariate Analysis 3, 226–235.

Onatski, A., Moreira, M., and Hallin, M., 2013, Asymptotic power of sphericity tests for high-dimensional data, Annals of Statistics 41, 1204–1231.

Orme, C., 1988, The calculation of the information matrix test for binary data models, The Manchester School 56, 370–376.

Sargan, J.D., 1988, Lectures on advanced econometric theory. Basil blackwell, Oxford.

Takeuchi, K., 1976, Distribution of information statistics and a criterion of model fitting, Suri-Kagaku (Mathematical Sciences) 153, 12–18 (in Japanese).

Wald, A., 1943, Tests of statistical hypotheses concerning several parameters when the number of observations is large, Transactions of the American Mathematical Society 54, 426–482.

White, H., 1982, Maximum likelihood estimation of misspecified models. Econometrica 50, 1–26.

	$\tau_*^2 > \max[\delta_*^2, \eta_*^2]$	$\delta_*^2 > \max[\tau_*^2, \eta_*^2]$	$\eta_*^2 > \max[\tau_*^2, \delta_*^2]$
$\sigma_* > \gamma_*$	$\widehat{\mathfrak{B}}_n^{(1)}$	$\widehat{\mathfrak{B}}_n^{(2)}$	$\widehat{\mathfrak{E}}_n^{(2)}$
$\sigma_* = \gamma_*$	$\widehat{\mathfrak{B}}_n^{(1)}, \widehat{\mathfrak{D}}_n^{(1)}, \widehat{\mathfrak{E}}_n^{(1)}$	$\widehat{\mathfrak{B}}_n^{(2)}, \widehat{\mathfrak{E}}_n^{(3)}, \widehat{\mathfrak{S}}_n^{(1)}$	$\widehat{\mathfrak{E}}_n^{(2)}, \widehat{\mathfrak{D}}_n^{(2)}, \widehat{\mathfrak{S}}_n^{(2)}$
$\sigma_* < \gamma_*$	$\widehat{\mathfrak{E}}_n^{(1)}$	$\widehat{\mathfrak{S}}_n^{(1)}$	$\widehat{\mathfrak{S}}_n^{(2)}$

Table 1: Test Statistics with the Greatest Leading Terms under the Fixed Alternative. The Test statistics in each cell indicates those with the greatest leading term under the fixed alternative hypothesis and the condition for each cell.

Statistics \ h	0.0	0.5	1.0	1.5	2.0	2.5	3.0
$\widehat{\mathfrak{B}}_n^{(1)}$	4.64	9.16	21.85	44.45	70.99	89.77	97.68
$\widehat{\mathfrak{B}}_n^{(2)}$	4.57	8.91	21.32	43.67	70.46	86.45	97.61
$\widehat{\mathfrak{S}}_n^{(1)}$	4.63	9.15	21.85	44.42	70.99	89.77	97.67
$\widehat{\mathfrak{S}}_n^{(2)}$	4.61	8.81	20.93	43.00	69.89	89.17	97.67
$\widehat{\mathfrak{E}}_n^{(1)}$	4.63	9.15	21.85	44.40	70.98	89.77	97.67
$\widehat{\mathfrak{E}}_n^{(2)}$	4.61	8.81	20.94	43.00	69.89	89.19	97.47
$\widehat{\mathfrak{M}}_n$	4.74	9.33	21.90	44.47	70.99	89.77	97.68
$\mathfrak{LR}_n^{\ddagger}$	4.55	8.14	19.88	41.22	68.26	88.17	97.22
Power Envelope ($\mathfrak{P}_e$)	4.99	10.89	29.29	56.41	80.74	94.24	98.87

Table 2: Empirical Levels of the Test Statistics (Level of Significance: 5%). Repetitions: 10,000. Sample Size: 5,000. Model for the test statistics: $Y_t \sim$ IID $N(0, B_{n,*})$, where $B_{n,*} = I_2 + n^{-1/2}\bar{B}_*$. $H_0^{\ddagger} : \bar{B}_* = 0$ versus $H_0^{\ddagger} : \bar{B}_* \neq 0$.

Statistics \ h	0.0	0.02	0.04	0.06	0.08	0.10
$\widehat{\mathfrak{B}}_n^{(1)}$	5.71	12.59	31.98	64.86	91.41	99.41
$\widehat{\mathfrak{B}}_n^{(2)}$	5.60	12.02	30.92	63.24	90.50	99.30
$\widehat{\mathfrak{S}}_n^{(1)}$	5.69	12.59	31.95	64.80	91.39	99.39
$\widehat{\mathfrak{S}}_n^{(2)}$	5.44	11.61	29.82	61.54	89.64	99.17
$\widehat{\mathfrak{E}}_n^{(1)}$	5.68	12.59	31.92	64.75	91.37	99.39
$\widehat{\mathfrak{E}}_n^{(2)}$	5.44	11.63	29.82	61.56	89.72	99.18
$\widehat{\mathfrak{M}}_n$	5.88	12.67	31.99	64.93	91.42	99.41

Table 3: Empirical Levels of the Test Statistics (Level of Significance: 5%). Repetitions: 10,000. Sample Size: 5,000. Model for the test statistics: $Y_t := B_*^{1/2}C_t$, where $C_t := (w_{1t}^2 - 1, w_{2t}^2 - 1)'/\sqrt{2}$, $W_t := (w_{1t}, w_{2t})' \sim$ IID $N(0, I_2)$, where $B_* = I_2 + \bar{B}_*$. $H_0^{\ddagger} : \bar{B}_* = 0$ versus $H_0^{\ddagger} : \bar{B}_* \neq 0$.

Statistics \ n	50	100	200	300	400	500
$\widehat{\mathfrak{B}}_n^{(1)}$	3.14	3.56	4.90	4.70	4.28	4.38
$\widehat{\mathfrak{B}}_n^{(2)}$	3.70	4.10	5.16	5.02	4.20	4.60
$\widehat{\mathfrak{S}}_n^{(1)}$	3.22	3.96	5.12	5.06	4.22	4.50
$\widehat{\mathfrak{S}}_n^{(2)}$	4.78	5.20	5.62	5.36	4.46	4.76
$\widehat{\mathfrak{E}}_n^{(1)}$	2.68	3.30	4.80	4.60	4.24	4.34
$\widehat{\mathfrak{E}}_n^{(2)}$	4.72	5.18	5.56	5.30	4.56	4.76
$\widehat{\mathfrak{M}}_n$	3.64	3.88	5.10	5.12	4.32	4.48

Table 4: Empirical Levels of the Test Statistics (Level of Significance: 5%). Repetitions: 5,000. Bootstrap Repetitions: 500. Model for TSLS: $X_t'\beta_*$ with $X_t = (1, x_t)'$ and IVs: $(1, z_t)'$. DGP: $y_t = \frac{1}{2}x_t + u_t$, $x_t := u_t + z_t$, and $(z_t, u_t)' \sim$ IID $N(0, I_2)$.

Statistics \ n	100	500	1,000	1,500	2,000	2,500
$\widehat{\mathfrak{B}}_n^{(1)}$	49.55	89.85	95.70	98.00	98.80	99.10
$\widehat{\mathfrak{B}}_n^{(2)}$	63.55	94.55	97.70	99.00	99.45	99.40
$\widehat{\mathfrak{S}}_n^{(1)}$	51.80	92.05	96.60	98.50	99.10	99.05
$\widehat{\mathfrak{S}}_n^{(2)}$	68.15	94.75	97.50	99.20	99.20	99.45
$\widehat{\mathfrak{E}}_n^{(1)}$	36.55	77.40	89.70	94.20	95.80	97.30
$\widehat{\mathfrak{E}}_n^{(2)}$	68.75	95.40	98.00	99.45	99.55	99.40
$\widehat{\mathfrak{M}}_n$	58.55	93.55	97.30	98.70	99.25	99.40

Table 5: Empirical Powers of the Test Statistics (Level of Significance: 5%). Repetitions: 2,000. Bootstrap Repetitions: 500. Model for TSLS: $X_t'\beta_*$ with $X_t = (1, x_t)'$ and IVs: $(1, z_t)'$. DGP: $y_t = \frac{1}{2}x_t + (1 + \exp(x_t))u_t$, $x_t := u_t + z_t$, and $(z_t, u_t)' \sim$ IID $N(0, I_2)$.

Statistics \ n	100	200	400	600	800	1,000
$\widehat{\mathfrak{B}}_n^{(1)}$	64.00	83.70	95.95	98.70	99.50	99.90
$\widehat{\mathfrak{B}}_n^{(2)}$	59.85	80.70	94.60	98.10	99.45	99.90
$\widehat{\mathfrak{S}}_n^{(1)}$	65.65	85.00	95.85	98.70	99.50	99.95
$\widehat{\mathfrak{S}}_n^{(2)}$	50.85	72.15	91.05	96.90	98.85	99.60
$\widehat{\mathfrak{E}}_n^{(1)}$	66.95	86.10	96.65	99.00	99.60	99.95
$\widehat{\mathfrak{E}}_n^{(2)}$	48.45	69.35	90.00	96.60	98.50	99.45
$\widehat{\mathfrak{M}}_n$	62.80	82.50	95.35	98.40	99.50	99.90

Table 6: Empirical Powers of the Test Statistics (Level of Significance: 5%). Repetitions: 2,000. Bootstrap Repetitions: 500. Model for TSLS: $X_t'\beta_*$ with $X_t = (1, x_t)'$ and IVs: $(1, z_t)'$. DGP: $y_t = \frac{1}{2}x_t^4 + u_t$, $x_t := u_t + z_t$ and $(z_t, u_t)' \sim$ IID $N(0, I_2)$.

Statistics \ n	50	100	300	500	700	900
$\widehat{\mathfrak{B}}_n^{(1)}$	13.30	29.00	72.45	92.25	96.45	99.20
$\widehat{\mathfrak{B}}_n^{(2)}$	13.15	27.40	69.15	90.85	96.05	99.00
$\widehat{\mathfrak{S}}_n^{(1)}$	13.25	27.55	69.40	91.05	96.10	99.00
$\widehat{\mathfrak{S}}_n^{(2)}$	12.65	24.00	63.95	89.15	95.05	98.60
$\widehat{\mathfrak{E}}_n^{(1)}$	13.20	29.45	72.95	92.50	96.45	99.20
$\widehat{\mathfrak{E}}_n^{(2)}$	13.00	24.20	63.35	89.10	95.00	98.50
$\widehat{\mathfrak{M}}_n$	13.05	26.70	69.45	91.10	96.15	98.95

Table 7: Empirical Powers of the Test Statistics (Level of Significance: 5%). Repetitions: 2,000. Bootstrap Repetitions: 500. Model for TSLS: $X_t'\beta_*$ with $X_t = (1, x_t)'$ and IVs: $(1, z_t)'$. DGP: $y_t = \frac{1}{2}x_t + u_t$, $x_t := z_t + \varepsilon_t$, $u_t = \frac{1}{10}u_{t-1} + \frac{1}{10}u_{t-2} + \varepsilon_t + \frac{1}{10}\varepsilon_{t-1}$, and $(z_t, \varepsilon_t)' \sim$ IID $N(0, I_2)$.

Statistics \ n	1,000	1,100	1,200	1,300	1,400	1,500
$\widehat{\mathfrak{B}}_n^{(1)}$	50.73	53.00	54.37	51.53	51.33	52.40
$\widehat{\mathfrak{B}}_n^{(2)}$	49.93	52.07	53.23	51.07	51.07	51.40
$\widehat{\mathfrak{S}}_n^{(1)}$	49.60	51.90	53.07	50.83	50.77	51.17
$\widehat{\mathfrak{S}}_n^{(2)}$	48.70	51.50	52.10	49.63	49.83	50.27
$\widehat{\mathfrak{E}}_n^{(1)}$	50.57	52.97	54.17	51.53	51.33	52.33
$\widehat{\mathfrak{E}}_n^{(2)}$	48.90	51.83	52.13	49.77	49.93	50.40
$\widehat{\mathfrak{M}}_n$	49.70	51.83	53.10	50.63	50.80	51.53

Table 8: Empirical Local Powers of the Test Statistics (Level of Significance: 5%). Repetitions: 3,000. Bootstrap Repetitions: 500. Model for TSLS: $X_t'\beta_*$ with $X_t = (1, x_t)'$ and IVs: $(1, z_t)'$. DGP: $y_t = \frac{1}{2}x_t + (1 + n^{-1/2}\exp(x_t)/2)u_t$, $x_t := u_t + z_t$, and $(z_t, u_t)' \sim$ IID $N(0, I_2)$.

Statistics \ n	500	600	700	800	900	1,000
$\widehat{\mathfrak{B}}_n^{(1)}$	37.27	39.57	37.97	36.60	35.77	36.57
$\widehat{\mathfrak{B}}_n^{(2)}$	37.40	39.63	37.93	37.33	36.20	37.10
$\widehat{\mathfrak{S}}_n^{(1)}$	37.07	39.30	37.47	37.03	36.00	36.83
$\widehat{\mathfrak{S}}_n^{(2)}$	36.70	39.80	37.50	37.03	36.30	37.10
$\widehat{\mathfrak{E}}_n^{(1)}$	36.73	39.07	37.73	36.33	35.50	36.53
$\widehat{\mathfrak{E}}_n^{(2)}$	37.10	40.10	37.80	37.27	36.50	37.53
$\widehat{\mathfrak{M}}_n$	37.10	39.33	37.50	37.10	35.73	36.80

Table 9: Empirical Local Powers of the Test Statistics (Level of Significance: 5%). Repetitions: 3,000. Bootstrap Repetitions: 500. Model for TSLS: $(1, z_t)'$. DGP: $y_t = \frac{1}{10}n^{-1/2}x_t^4 + u_t$, $x_t := u_t + z_t$, and $(z_t, u_t)' \sim$ IID $N(0, I_2)$.

Statistics \ n	200	300	400	500	600	700
$\widehat{\mathfrak{B}}_n^{(1)}$	16.40	16.23	16.57	16.03	16.67	15.23
$\widehat{\mathfrak{B}}_n^{(2)}$	15.67	15.17	15.53	14.83	15.40	13.87
$\widehat{\mathfrak{S}}_n^{(1)}$	15.87	15.40	15.53	14.83	15.47	14.07
$\widehat{\mathfrak{S}}_n^{(2)}$	14.00	13.37	14.23	13.90	14.10	13.00
$\widehat{\mathfrak{E}}_n^{(1)}$	16.57	16.27	16.70	16.10	16.60	15.23
$\widehat{\mathfrak{E}}_n^{(2)}$	14.10	13.33	14.20	13.73	14.10	13.10
$\widehat{\mathfrak{M}}_n$	15.37	15.10	15.43	14.77	15.00	13.97

Table 10: Empirical Local Powers of the Test Statistics (Level of Significance: 5%). Repetitions: 3,000. Bootstrap Repetitions: 500. Model for TSLS: $X_t'\beta_*$ with $X_t = (1, x_t)'$ and IVs: $(1, z_t)'$. DGP: $y_t = \frac{1}{2}x_t + u_t$, $x_t := z_t + \varepsilon_t$, $u_t = \frac{3}{4}n^{-1/2}u_{t-1} + \frac{3}{4}n^{-1/2}u_{t-2} + \varepsilon_t + \frac{3}{4}n^{-1/2}\varepsilon_{t-1}$, and $(z_t, \varepsilon_t)' \sim$ IID $N(0, I_2)$.