# Testing the Equality of Two Positive-Definite Matrices with Application to Information Matrix Testing*


JIN SEO CHO
School of Economics
Yonsei University
50 Yonsei-ro, Seodaemun-gu, Seoul 120-749, Korea
Email: jinseocho@yonsei.ac.kr

HALBERT WHITE
Department of Economics
University of California, San Diego
9500 Gilman Dr., La Jolla, CA, 92093-0508, U.S.A.
Email: hwhite@weber.ucsd.edu



First version: September 2011 This version: June 2014



**Abstract**

We provide a new characterization of the equality of two positive-definite matrices **A** and **B**, and we use this to propose several new computationally convenient statistical tests for the equality of two unknown positive-definite matrices. Our primary focus is on testing the information matrix equality (*e.g.*, White, 1982, 1994). We characterize the asymptotic behavior of our new trace-determinant information matrix test statistics under the null and the alternative and investigate their finite-sample performance for a variety of models: linear regression, exponential duration, probit, and Tobit. The parametric bootstrap suggested by Horowitz (1994) delivers critical values that provide admirable level behavior, even in samples as small as $n = 50$. Our new tests often have better power than the parametric-bootstrap version of the traditional IMT; when they do not, they nevertheless perform respectably.



**Key Words**: Matrix equality; Information matrix test; Eigenvalues; Trace; Determinant; Eigenspectrum test; Parametric Bootstrap.

**Subject Classification**: C01, C12, C52

**Acknowledgements:** The second author (Halbert White) had passed away before the final version was completed. The first author would like to acknowledge his contributions to this paper. He formed the outlines of the current paper. In addition, the authors are most grateful to the Co-editors, Yoosoon Chang, Tom Fomby, and Joon Y. Park, and one anonymous referee. We also acknowledge helpful discussions with Anil Bera, Stephane Bonhomme, Richard Golden, Steven Henley, Atsushi Inoue, Jinook Jeong, Kiho Jeong, Michael Kashner, Hak Bae Lee, Jin Lee, Ki Young Park, Sang Soo Park, Peter, C.B. Phillips, Robert Strawderman, Yoon Jae Whang, Byoungsam Yoo, Victoria Zinde-Walsh, and other participants at the 14th Advances in Econometrics Conference in Honor of Peter C.B. Phillips (Southern Methodist University, 2013), the Economics Joint Conference 2012 and the 17th Joint Workshop of Yonsei University, Hokkaido University (Hokkaido University, 2012), and KEA International Conference (Korea University, 2012). Also, Taeyoun Kim and Myungkoo Song provided excellent research assistance for our Monte Carlo experiments.


*A glossary of notation and the program codes written in GAUSS for our simulations are available in the following URL: http://web.yonsei.ac.kr/jinseocho/research.htm.

# 1 Introduction

Testing the equality of two positive-definite matrices, say **A** and **B**, has a distinguished history in multivariate statistics. Wilks (1935) showed that for data drawn from two independent multivariate normal samples, the likelihood ratio (LR) test for independence is a ratio of the determinants of the two maximum-likelihood covariance matrix estimators. This finding stimulated investigation of other determinant-based tests. For example, Pillai and Nagarsenker (1972) and Das Gupta and Giri (1973) examined the properties of a class of determinant ratio tests for equal covariance matrices. Other researchers have studied tests based on the trace of two covariance matrix estimators. Roy (1953), Pillai and Jayachandran (1968), and Nagao (1973, 1974) developed trace-based tests and compared their performance to that of determinant-based tests.

Mauchly (1940) considered a related but weaker hypothesis, sphericity, *i.e.*, that $\mathbf{A} = d\mathbf{B}$ for some unknown $d > 0$. For normal data, the LR statistic is a function of both the estimated traces and determinants.

Testing the equality of two positive-definite matrices also plays a central role in model specification analysis. Specifically, when a probability model used for maximum-likelihood estimation is correctly specified, the information matrix equality $\mathbf{A} = \mathbf{B}$ holds, where **A** is the opposite of the average log-likelihood Hessian and **B** is the average of the outer product of the log-likelihood scores, both evaluated at the true parameter (see, *e.g.*, Fisher, 1922, 1925). White (1982) proposed an information matrix test (IMT) for model misspecification based on a comparison of estimates of the elements of **A** and **B**. The omnibus or non-directional IMT compares all the non-redundant elements of **A** and **B**, yielding an asymptotic chi-squared test statistic. Directional specification tests (*e.g.*, for conditional heteroskedasticity, skewness, or kurtosis) are obtained by comparing selected elements of **A** and **B**, also yielding asymptotic chi-squared tests.

The intuitive appeal and generality of this proposal led to an extensive body of work studying the properties of the IMT, including papers of Chesher (1983), Lancaster (1984), Orme (1988, 1990), Taylor (1987), Hall (1987), Chesher and Spady (1991), Horowitz (1994), and Dhaene (2004), among others. One early finding was negative: Taylor (1987) and Orme (1990), among others, showed that previously proposed IMTs typically suffered from extreme level distortions, so much so as to render them impractical for empirical application. Chesher and Spady (1991) showed that these level distortions can be resolved by employing higher-order expansions for the IMT. As Horowitz (1994) pointed out, however, this approach is extremely cumbersome because it involves higher-order cumulants, making it impractical for all but the simplest models. Instead, Horowitz proposed use of the parametric bootstrap. This method is computationally straightforward, and, as Horowitz's Monte Carlo experiments show, parametric bootstrap-based IMTs exhibit outstanding level performance, even for modest sample sizes and in cases where the test statistic is not pivotal. Furthermore, it has power properties similar to those exhibited by level distortion-adjusted IMTs. The parametric bootstrap has been the preferred method for IMT implementation since its introduction. Nevertheless, it is unclear whether the parametric bootstrap still works well for other non-pivotal tests not used by Horowitz in his Monte Carlo experiments.

Until recently, little or no attention has been paid to the possibility of constructing IMTs based on the trace or determinant of $\mathbf{A}$ and $\mathbf{B}$. To the best of our knowledge, Golden, Henley, White, and Kashner (2013) were the first to systematically investigate this possibility by proposing IMTs based on functions of the log eigenspectra of $\mathbf{A}$ and $\mathbf{B}$, including tests based on the trace and determinant. A major contribution of the present paper is to extend this line of inquiry by proposing tests of $\mathbf{A} = \mathbf{B}$ based on trace and determinant properties that *characterize* the equality of two positive-definite matrices. This characterization, namely that $\mathbf{A} = \mathbf{B}$ if and only if $\det(\mathbf{B}\mathbf{A}^{-1})^{1/k} = k^{-1}\mathrm{tr}(\mathbf{B}\mathbf{A}^{-1}) = 1$, or $\det(\mathbf{A}\mathbf{B}^{-1})^{1/k} = k^{-1}\mathrm{tr}(\mathbf{A}\mathbf{B}^{-1}) = 1$, is apparently new; we have not been able to find its applications elsewhere. It is remarkably simple, and it is striking in that it enables an omnibus[1] IMT based on a comparison of only two quantities, regardless of the matrix dimension. This result also applies to the classical problem of testing the equality of two covariance matrices, but without requiring data to be generated by the multivariate normal distribution.

The resulting computational simplicity of these new trace-determinant tests is by itself appealing. As we further show, implementation using the parametric bootstrap yields tests with outstanding level performance, which implies that the parametric bootstraps show excellent performance even for our non-pivotal test statistics.

To date, the IMT literature has focused mostly on the level properties of the IMT. Here, we further contribute to the extant literature by studying the power properties of standard IMTs, other related tests, and our new trace-determinant IMTs by using the parametric bootstrap. Specifically, we examine the power of the various tests to detect misspecification in four models: linear regression, exponential duration, probit, and Tobit. We find that our new trace-determinant tests often outperform previous tests. When they do not, they nevertheless perform respectably. In addition, we examine the local power properties of our new tests and associate them with the standard LR statistics testing for covariance equality and sphericity under the multivariate normal distribution

[1] Directional versions can be constructed by selecting suitable submatrices of $\mathbf{A}$ and $\mathbf{B}$.

condition.

The plan of the paper is as follows. In Section 2, we state the lemma characterizing the equality of two positive-definite matrices, discuss its implications for information matrix testing, and propose three related IMTs. In Section 3, we study the asymptotic behavior of our test statistics under the null, alternative, and local alternative hypotheses. We also examine how our test statistics are interrelated with the LR test under the classical multivariate normal distribution condition and discuss implementation considerations; following Horowitz (1994), we recommend use of the parametric bootstrap. In Section 4, we investigate by Monte Carlo experiments the finite-sample level, power, and local power of our new trace-determinant IMTs, the Chesher (1983) and Lancaster (1984) IMTs, and some related tests. Section 5 contains a summary and concluding remarks. Mathematical proofs and additional assumptions are gathered in the Mathematical Appendix.

Before proceeding, we present the following notational details. A function will always be denoted using an empty argument: a function mapping $f$ from $\mathscr{X}$ to $\mathscr{Y}$ will be denoted by $f(\cdot)$ or $f : \mathscr{X} \mapsto \mathscr{Y}$. When $f$ alone is used, this denotes a variable in a Euclidean space. Also, for notational simplicity, we denote $f'(x)|_{x=x_*}$ by $f'(x_*)$ and let $\partial_x f(x)$ and $\partial^2_{x,y} f(x,y)$ denote $(\partial/\partial x)f(x)$ and $(\partial^2/\partial x \partial y)f(x,y)$, respectively. We use the following simple facts without reference: $\text{tr}[\mathbf{AB}] = \text{tr}[\mathbf{BA}]$, $\det[\mathbf{AB}] = \det[\mathbf{A}]\det[\mathbf{B}]$, and $\det[\mathbf{A}^{-1}] = 1/\det[\mathbf{A}]$.

## 2 A Basic Lemma and Its Testing Implications

Our tests for the equality of two positive-definite matrices are based on the following simple but striking lemma.

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

$$\det[\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}]^{1/k} = 1 \quad \textit{and} \quad \text{tr}[\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}]/k = 1,$$

*where* $\mathbf{A}^{-1/2}$ *is the real positive-definite* $k \times k$ *matrix square root of* $\mathbf{A}^{-1}$. □

Thus, $\mathbf{A} = \mathbf{B}$ whenever the geometric and arithmetic means of the eigenvalues of $\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}$ both equal 1. We state the result in terms of $\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}$ to make it obvious that this matrix has real positive eigenvalues. Equivalently, with $\mathbf{D} := \mathbf{B}\mathbf{A}^{-1}$, $\mathbf{A} = \mathbf{B}$ if and only if $\det(\mathbf{D})^{1/k} = k^{-1}\text{tr}(\mathbf{D}) = 1$. To the best of our knowledge, one of the exercise questions in Magnus and Neudecker (1999) is closest to ours; this states that for any positive semi-definite matrix $\mathbf{M}$, $\det[\mathbf{M}]^{1/k} = \text{tr}[\mathbf{M}]/k$ if and only if $\mathbf{M}$

is proportional to **I**. As another characterization, Strawderman (1994) also show that $\mathbf{A} = \mathbf{B}$ if and only if $k^{-1}\text{tr}(\mathbf{D}) = 1$ and $k^{-1}\text{tr}(\mathbf{D}^2) = 1$.

No IMT exists using Lemma 1, although similar motivations using eigenvalues exist. Bera (1986) and Bera and Hall (1991) regarded eigenvalues as key elements for testing two equal positive-definite matrices. They focused on the eigenvalues of $\mathbf{A}-\mathbf{B}$ and used their distances to test the equality of **A** and **B**. Similarly, many tests using the eigenvalues can be developed, although developing a general theory on their applications is not straightforward, mainly due to their non-differentiability.

To apply Lemma 1, we construct statistics testing any two of the following hypotheses:

$$\text{tr}[\mathbf{D}]/k = 1 \quad (1)$$

$$\det[\mathbf{D}]^{1/k} = 1 \quad (2)$$

$$\text{tr}[\mathbf{D}]/k = \det[\mathbf{D}]^{1/k}. \quad (3)$$

Any two of these imply the other. Thus, to test

$$\mathbb{H}_0 : \mathbf{A} = \mathbf{B} \ \text{ versus } \ \mathbb{H}_1 : \mathbf{A} \neq \mathbf{B},$$

we consider three equivalent hypotheses, each of which yields a corresponding test statistic:

$$\mathbb{H}_0^{(1)} : \text{tr}[\mathbf{D}]/k = 1 \text{ and } \det[\mathbf{D}]^{1/k} = 1 \quad \text{vs.} \quad \mathbb{H}_1^{(1)} : \text{tr}[\mathbf{D}]/k \neq 1 \text{ or } \det[\mathbf{D}]^{1/k} \neq 1;$$

$$\mathbb{H}_0^{(2)} : \text{tr}[\mathbf{D}]/k = 1 \text{ and } \text{tr}[\mathbf{D}]/k = \det[\mathbf{D}]^{1/k} \quad \text{vs.} \quad \mathbb{H}_1^{(2)} : \text{tr}[\mathbf{D}]/k \neq 1 \text{ or } \text{tr}[\mathbf{D}]/k \neq \det[\mathbf{D}]^{1/k};$$

$$\mathbb{H}_0^{(3)} : \det[\mathbf{D}]^{1/k} = 1 \text{ and } \text{tr}[\mathbf{D}]/k = \det[\mathbf{D}]^{1/k} \quad \text{vs.} \quad \mathbb{H}_1^{(3)} : \text{tr}[\mathbf{D}]/k \neq 1 \text{ or } \text{tr}[\mathbf{D}]/k \neq \det[\mathbf{D}]^{1/k}.$$

Clearly, $\mathbb{H}_0^{(1)}$ tests (1) and (2), $\mathbb{H}_0^{(2)}$ tests (1) and (3), and $\mathbb{H}_0^{(3)}$ tests (2) and (3).

To define test statistics corresponding to $\mathbb{H}_0^{(1)} - \mathbb{H}_0^{(3)}$, let $\mathbf{A}_n$ and $\mathbf{B}_n$ be consistent estimators for **A** and **B**, and write $\mathbf{D}_n := \mathbf{B}_n\mathbf{A}_n^{-1}$. Each test statistic combines two of the following:

$$T_n := \text{tr}[\mathbf{D}_n]/k - 1; \ \ D_n := \det[\mathbf{D}_n]^{1/k} - 1; \ \text{ and } \ S_n := \text{tr}[\mathbf{D}_n]/k - \det[\mathbf{D}_n]^{1/k}.$$

When $\mathbf{A} = \mathbf{B}$, the consistency of $(\mathbf{A}_n, \mathbf{B}_n)$ implies that each of these should be negligible in probability. For the estimators $(\mathbf{A}_n, \mathbf{B}_n)$ studied here, it turns out that $\mathbf{A} = \mathbf{B}$ implies $T_n = O_{\mathbb{P}}(n^{-1/2})$, $D_n = O_{\mathbb{P}}(n^{-1/2})$, and $S_n = O_{\mathbb{P}}(n^{-1})$, as we show in Lemma 3 below.

The given hypotheses and statistics are essential in testing $\mathbb{H}_0$ against $\mathbb{H}_1$ when the eigenvalues of $\mathbf{BA}^{-1}$ are exploited. We do not need to consider any other transformation of the eigenvalues. As pointed out by Bera (1986) and Bera and Hall (1991), the eigenvalues of $\mathbf{A}-\mathbf{B}$ can be transformed into many statistics in various ways. Instead, our statistics simply focus on the trace and determinant of $\mathbf{B}\mathbf{A}^{-1}$ without introducing any exceptional case for $\mathbf{A}=\mathbf{B}$. In addition, our statistics do not estimate eigenvectors. The level and power properties of their tests critically depend upon the way to estimate the eigenvectors, as in Bera and Hall (1991). Strawderman (1994) also considers another testing principle using his characterization. Comparing these test statistics with ours is left as a future research topic.

Consider the following statistics, corresponding to $\mathbb{H}_0^{(1)} - \mathbb{H}_0^{(3)}$:

$$\mathscr{B}_n^{(1)} := nk^2\left(\frac{1}{2}T_n^2 + \frac{1}{2}D_n^2\right);\;\; \mathscr{B}_n^{(2)} := 2nk\left(\frac{1}{2}T_n^2 + S_n\right);\;\; \text{and}\;\; \mathscr{B}_n^{(3)} := 2nk\left(\frac{1}{2}D_n^2 + S_n\right).$$

Under $\mathbb{H}_0$, each of these is bounded but not negligible in probability. Note that $T_n$ and $D_n$ are squared, while $S_n$ is not. This is due to their different orders of convergence. The coefficient 1/2 multiplying $T_n^2$ and $D_n^2$ ensures a straightforward asymptotic null distribution, and $S_n \geq 0$ because the arithmetic mean is always greater than or equal to the geometric mean.

To apply these methods to information matrix testing, we can let $\widehat{\boldsymbol{\theta}}_n$ be the (quasi-) maximum-likelihood ((Q)ML) estimator (White, 1982) consistent for $\boldsymbol{\theta}_*$, an interior element of $\Theta \subset \mathbb{R}^k$, such that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \overset{\text{A}}{\sim} N\left[\mathbf{0}, \mathbf{A}_*{}^{-1}\mathbf{B}_*\mathbf{A}_*{}^{-1}\right].$$

Here, $\mathbf{A}_*$ is the negative Hessian matrix of the (quasi-) likelihood function at $\boldsymbol{\theta}_*$, and $\mathbf{B}_*$ is the asymptotic covariance matrix of the (quasi-) log-likelihood scores. Correct specification implies the information matrix equality, $\mathbf{A}_* = \mathbf{B}_*$, under mild regularity conditions, justifying the classical maximum-likelihood covariance matrix estimators. Thus, if we let $\mathbf{A}(\cdot)$ and $\mathbf{B}(\cdot)$ be continuous functions of $\boldsymbol{\theta}$ (*i.e.*, $\mathbf{A}:\Theta \mapsto \mathbb{R}^{k\times k}$ and $\mathbf{B}:\Theta \mapsto \mathbb{R}^{k\times k}$), we can test $\mathbf{A}_* := \mathbf{A}(\boldsymbol{\theta}_*) = \mathbf{B}_* := \mathbf{B}(\boldsymbol{\theta}_*)$ using estimators $\widehat{\mathbf{A}}_n := \mathbf{A}_n(\widehat{\boldsymbol{\theta}}_n)$ and $\widehat{\mathbf{B}}_n := \mathbf{B}_n(\widehat{\boldsymbol{\theta}}_n)$, where $\mathbf{A}_n(\cdot)$ and $\mathbf{B}_n(\cdot)$ are sample analogs of $\mathbf{A}(\cdot)$ and $\mathbf{B}(\cdot)$. We also let $\widehat{\mathbf{D}}_n := \widehat{\mathbf{B}}_n\widehat{\mathbf{A}}_n^{-1}$ and define

$$\widehat{T}_n := \text{tr}[\widehat{\mathbf{D}}_n]/k - 1;\;\; \widehat{D}_n := \det[\widehat{\mathbf{D}}_n]^{1/k} - 1;\;\; \text{and}\;\; \widehat{S}_n := \text{tr}[\widehat{\mathbf{D}}_n]/k - \det[\widehat{\mathbf{D}}_n]^{1/k},$$

from which we construct

$$\widehat{\mathscr{B}}_n^{(1)} := nk^2\left(\frac{1}{2}\widehat{T}_n^2 + \frac{1}{2}\widehat{D}_n^2\right);\;\; \widehat{\mathscr{B}}_n^{(2)} := 2nk\left(\frac{1}{2}\widehat{T}_n^2 + \widehat{S}_n\right);\;\; \text{and}\;\; \widehat{\mathscr{B}}_n^{(3)} := 2nk\left(\frac{1}{2}\widehat{D}_n^2 + \widehat{S}_n\right).$$

As it turns out, estimating $\boldsymbol{\theta}_*$ using $\widehat{\boldsymbol{\theta}}_n$ impacts the asymptotic null behavior of our test statistics, introducing certain additive terms.

Furthermore, we notice that the dimension condition can be relaxed to handle general cases. That is, as we detail below, our constructions are valid even when $\Theta \subset \mathbb{R}^{\ell}$ with $\ell \in \mathbb{N}$. We, therefore, suppose this from now on and impose $\ell = k$ only when testing the information matrix equality.

We also consider the simpler but important special case of testing the equality of the population covariance matrices of $\mathbf{a}(Y_t,\mathbf{X}_t)$ and $\mathbf{b}(Y_t,\mathbf{X}_t)$, where $\mathbf{a}(Y_t,\mathbf{X}_t)$ and $\mathbf{b}(Y_t,\mathbf{X}_t)$ are possibly nonlinear transformations of $(Y_t,\mathbf{X}_t)$. For this, we compare two sample covariance matrices $\widetilde{\mathbf{A}}_n$ and $\widetilde{\mathbf{B}}_n$, defined as

$$\widetilde{\mathbf{A}}_n := \widehat{\Sigma}_{\mathbf{a},n} - \widehat{\boldsymbol{\mu}}_{\mathbf{a},n}\widehat{\boldsymbol{\mu}}'_{\mathbf{a},n}, \qquad \text{and} \qquad \widetilde{\mathbf{B}}_n := \widehat{\Sigma}_{\mathbf{b},n} - \widehat{\boldsymbol{\mu}}_{\mathbf{b},n}\widehat{\boldsymbol{\mu}}'_{\mathbf{b},n},$$

where, for the given vector-valued functions $\mathbf{a}(\cdot,\cdot)$ and $\mathbf{b}(\cdot,\cdot)$,

$$\widehat{\boldsymbol{\mu}}_{\mathbf{a},n} := n^{-1}\sum_{t=1}^{n}\mathbf{a}(Y_t,\mathbf{X}_t), \qquad \widehat{\Sigma}_{\mathbf{a},n} := n^{-1}\sum_{t=1}^{n}\mathbf{a}(Y_t,\mathbf{X}_t)\mathbf{a}(Y_t,\mathbf{X}_t)',$$

$$\widehat{\boldsymbol{\mu}}_{\mathbf{b},n} := n^{-1}\sum_{t=1}^{n}\mathbf{b}(Y_t,\mathbf{X}_t), \qquad \text{and} \qquad \widehat{\Sigma}_{\mathbf{b},n} := n^{-1}\sum_{t=1}^{n}\mathbf{b}(Y_t,\mathbf{X}_t)\mathbf{b}(Y_t,\mathbf{X}_t)'.$$

Here, $(\widehat{\boldsymbol{\mu}}_{\mathbf{a},n}, \widehat{\Sigma}_{\mathbf{a},n}, \widehat{\boldsymbol{\mu}}_{\mathbf{b},n}, \widehat{\Sigma}_{\mathbf{b},n})$ is consistent for $(\boldsymbol{\mu}_{\mathbf{a},*}, \Sigma_{\mathbf{a},*}, \boldsymbol{\mu}_{\mathbf{b},*}, \Sigma_{\mathbf{b},*}) := (E[\mathbf{a}(Y_t,\mathbf{X}_t)], E[\mathbf{a}(Y_t,\mathbf{X}_t)\mathbf{a}(Y_t,\mathbf{X}_t)'],$ $E[\mathbf{b}(Y_t,\mathbf{X}_t)], E[\mathbf{b}(Y_t,\mathbf{X}_t)\mathbf{b}(Y_t,\mathbf{X}_t)'])$, and the other regularity conditions are provided in the Appendix for our test statistics defined in Theorem 2 below.

# 3 Test Statistic Asymptotic Behavior

## 3.1 Asymptotic Null Behavior

To examine the asymptotic null behavior of $\widehat{T}_n$, $\widehat{D}_n$, and $\widehat{S}_n \equiv \widehat{T}_n - \widehat{D}_n$, we impose the following regularity conditions:

**Assumption 1.** $(i)$ $(\Omega, \mathscr{F}, \mathbb{P})$ *is a complete probability space;*

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

$(iii)$ *there is a sequence of measurable mappings* $\{\widehat{\boldsymbol{\theta}}_n : \Omega \mapsto \Theta\}$ *consistent for a unique* $\boldsymbol{\theta}_* \in int(\Theta)$.

□

**Assumption 2.** $(i)$ *For* $k \in \mathbb{N}$, *the symmetric matrix mapping* $\mathbf{A} : \Theta \mapsto \mathbb{R}^{k\times k}$ *is in* $\mathscr{C}^{(2)}(\Theta)$ *and is such that* $\mathbf{A}_* := \mathbf{A}(\boldsymbol{\theta}_*)$ *is positive definite; and* (*ii*) *the symmetric matrix mapping* $\mathbf{B} : \Theta \mapsto \mathbb{R}^{k \times k}$ *is in* $\mathscr{C}^{(2)}(\Theta)$ *and is such that* $\mathbf{B}_* := \mathbf{B}(\boldsymbol{\theta}_*)$ *is positive definite.* □

**Assumption 3.** (*i*) *There are symmetric matrix estimators* $\mathbf{A}_n(\cdot)$ *and* $\mathbf{B}_n(\cdot)$ *consistent for* $\mathbf{A}(\cdot)$ *and* $\mathbf{B}(\cdot)$*, respectively, uniformly on* $\Theta$*; and*

(*ii*) $\sqrt{n}[(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)', \text{vech}[\mathbf{A}_n - \mathbf{A}_*]', \text{vech}[\mathbf{B}_n - \mathbf{B}_*]']' \overset{A}{\sim} N(\mathbf{0}, \Omega_*)$*, where* $\mathbf{A}_n := \mathbf{A}_n(\boldsymbol{\theta}_*)$*,* $\mathbf{B}_n := \mathbf{B}_n(\boldsymbol{\theta}_*)$*, and* $\Omega_*$ *is a* $(k^2 + k + \ell) \times (k^2 + k + \ell)$ *positive semi-definite matrix.* □

Here, $\text{vech}(\cdot)$ is used in Assumption 3(*ii*) to accommodate the symmetry of $\mathbf{A}_n$, $\mathbf{A}_*$, $\mathbf{B}_n$, and $\mathbf{B}_*$. An appropriate central limit theorem (CLT) typically ensures Assumption 3(*ii*). Also, $\mathbf{A}_n$ and $\mathbf{B}_n$ used for defining $T_n$, $D_n$, and $S_n$ correspond to $\mathbf{A}_n(\boldsymbol{\theta}_*)$ and $\mathbf{B}_n(\boldsymbol{\theta}_*)$, respectively. That is, these are estimators with known parameters. Similarly, we have $\mathbf{D}_n := \mathbf{B}_n(\boldsymbol{\theta}_*)\mathbf{A}_n(\boldsymbol{\theta}_*)^{-1}$. Finally, $\ell = k$ when testing the information matrix equality, and the first $k \times k$ block of $\Omega_*$ is equal to $\mathbf{A}_*^{-1}\mathbf{B}_*\mathbf{A}_*^{-1}$.

To simplify notation, we let $\mathbf{M}_n := \mathbf{A}_*^{-1}(\mathbf{B}_n - \mathbf{A}_n)$ and $\mathbf{S}_{j,*} := \mathbf{A}_*^{-1}(\partial_j \mathbf{B}_* - \partial_j \mathbf{A}_*)$ for $j = 1, \ldots, \ell$, where $\partial_j \mathbf{B}_* := (\partial/\partial\theta_j)\mathbf{B}_*$ and $\partial_j \mathbf{A}_* := (\partial/\partial\theta_j)\mathbf{A}_*$. We also write

$$\mathbf{K}_n := \mathbf{M}_n + \sum_{j=1}^{\ell} (\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{S}_{j,*}.$$

The $k \times k$ statistic $\mathbf{K}_n$ plays the central role in determining the asymptotic distributions of all our test statistics. Given Assumption 3(*ii*), $\mathbf{M}_n = O_{\mathbb{P}}(n^{-1/2})$, $(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) = O_{\mathbb{P}}(n^{-1/2})$, and $\mathbf{K}_n = O_{\mathbb{P}}(n^{-1/2})$. Furthermore, Assumption 3(*ii*) implies that $\mathbf{K}_n$ is asymptotically a $k \times k$ matrix of normal random variables. We omit deriving the asymptotic covariance matrix of $\mathbf{K}_n$. This is mainly because we will implement our testing by using the parametric bootstrap, which renders estimation of the covariance matrix unnecessary.

We obtain asymptotic approximations by applying Taylor expansions to $\widehat{T}_n$ and $\widehat{D}_n$. The following lemma relies on first-order expansions.

**Lemma 2.** *Given Assumptions 1, 2, 3, and* $\mathbb{H}_0$*,*

(*i*) $\widehat{T}_n = \text{tr}[\mathbf{K}_n]/k + o_{\mathbb{P}}(n^{-1/2})$*;*

(*ii*) $\widehat{D}_n = \text{tr}[\mathbf{K}_n]/k + o_{\mathbb{P}}(n^{-1/2})$*; and*

(*iii*) $\widehat{S}_n \equiv \widehat{T}_n - \widehat{D}_n = o_{\mathbb{P}}(n^{-1/2})$. □

Given Assumption 3, both $\widehat{T}_n$ and $\widehat{D}_n$ are $O_{\mathbb{P}}(n^{-1/2})$ under $\mathbb{H}_0$, and their asymptotic null distributions are normal. Asymptotic null distributions for $\widehat{T}_n^2$ and $\widehat{D}_n^2$ follow straightforwardly. In contrast, a first-order approximation is not sufficient to obtain an asymptotic distribution for $\widehat{S}_n$, regardless of whether there are estimated parameters, as $\widehat{S}_n = o_{\mathbb{P}}(n^{-1/2})$.

To fully specify the null behavior of our statistics, it turns out that second-order expansions for $\mathrm{tr}[\widehat{\mathbf{D}}_n]$ and $\det[\widehat{\mathbf{D}}_n]$ are required. For these, we write $\mathbf{W}_n := \mathbf{B}_*^{-1}(\mathbf{B}_n - \mathbf{B}_*)$, $\mathbf{U}_n := \mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_*)$, and, for $i, j = 1, \ldots, \ell$, $\mathbf{G}_{j,n} := \mathbf{B}_*^{-1}\partial_j(\mathbf{B}_n - \mathbf{B}_*)$ and $\mathbf{H}_{j,n} := \mathbf{A}_*^{-1}\partial_j(\mathbf{A}_n - \mathbf{A}_*)$. We also write $\partial_{ji}^2 := \partial^2/\partial\theta_j\partial\theta_i$, and we impose the following conditions.

**Assumption 4.** (*i*) *For $j = 1, \ldots, \ell$, $\partial_j \mathbf{A}_n(\cdot)$ and $\partial_j \mathbf{B}_n(\cdot)$ are consistent for $\partial_j \mathbf{A}(\cdot)$ and $\partial_j \mathbf{B}(\cdot)$, uniformly on $\Theta$; and*

(*ii*) *for $i, j = 1, \ldots, \ell$, $\partial_{ji}^2 \mathbf{A}_n(\cdot)$ and $\partial_{ji}^2 \mathbf{B}_n(\cdot)$ are consistent for $\partial_{ji}^2 \mathbf{A}(\cdot)$ and $\partial_{ji}^2 \mathbf{B}(\cdot)$, uniformly on $\Theta$.* □

**Assumption 5.** *For $j = 1, \ldots, \ell$, $\mathbf{H}_{j,n} = O_{\mathbb{P}}(n^{-1/2})$ and $\mathbf{G}_{j,n} = O_{\mathbb{P}}(n^{-1/2})$.* □

Assumptions 4 and 5 easily hold for many interesting models such as those considered in Section 4.

The following lemma provides the required second-order asymptotic approximations. For this, we let $\mathbf{J}_{j,n} := \mathbf{G}_{j,n} - \mathbf{H}_{j,n}$, we let $\nabla_{\boldsymbol{\theta}}^2$ denote the Hessian operator with respect to $\boldsymbol{\theta}$, and we define $\mathbf{D}_* := \mathbf{B}_*\mathbf{A}_*^{-1}$.

**Lemma 3.** *Given Assumptions 1, 2, 3, 4, 5, and $\mathbb{H}_0$,*

*(i) $\widehat{T}_n = \widehat{T}_n^\star + o_{\mathbb{P}}(n^{-1})$, where $\widehat{T}_n^\star := \widehat{T}_{n,1}^\star + \widehat{T}_{n,2}^\star$, $\widehat{T}_{n,1}^\star := \frac{1}{k}\mathrm{tr}[\mathbf{K}_n]$, and*

$$\widehat{T}_{n,2}^\star := -\frac{1}{k}\mathrm{tr}[\mathbf{K}_n\mathbf{U}_n] + \frac{1}{k}[\mathrm{tr}[\mathbf{J}_{j,n} - \mathbf{M}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + \frac{1}{2k}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla_{\boldsymbol{\theta}}^2\mathrm{tr}[\mathbf{D}_*](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)$$

*with $\widehat{T}_{n,1}^\star = O_{\mathbb{P}}(n^{-1/2})$ and $\widehat{T}_{n,2}^\star = O_{\mathbb{P}}(n^{-1})$;*

*(ii) $\widehat{D}_n = \widehat{D}_n^\star + o_{\mathbb{P}}(n^{-1})$, where $\widehat{D}_n^\star := \widehat{D}_{n,1}^\star + \widehat{D}_{n,2}^\star$, $\widehat{D}_{n,1}^\star := \frac{1}{k}\mathrm{tr}[\mathbf{K}_n]$, and*

$$\begin{aligned}\widehat{D}_{n,2}^\star := &\frac{1}{2k}\left(\frac{1}{k} - 1\right)\mathrm{tr}[\mathbf{K}_n]^2 + \frac{1}{2k}(\mathrm{tr}[\mathbf{M}_n]^2 + \mathrm{tr}[\mathbf{U}_n^2] - \mathrm{tr}[\mathbf{W}_n^2]) + \frac{1}{k}[\mathrm{tr}[\mathbf{M}_n]\mathrm{tr}[\mathbf{S}_{j,*}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)\\ &+ \frac{1}{k}[\mathrm{tr}[\mathbf{J}_{j,n} + \mathbf{U}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_* - \mathbf{W}_n\mathbf{A}_*^{-1}\partial_j\mathbf{B}_*]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + \frac{1}{2k}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla_{\boldsymbol{\theta}}^2\det[\mathbf{D}_*](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)\end{aligned}$$

*with $\widehat{D}_{n,1}^\star = O_{\mathbb{P}}(n^{-1/2})$ and $\widehat{D}_{n,2}^\star = O_{\mathbb{P}}(n^{-1})$;*

*(iii) $\widehat{S}_n = \widehat{S}_n^\star + o_{\mathbb{P}}(n^{-1})$, where*

$$\begin{aligned}\widehat{S}_n^\star := &-\frac{1}{2k}\left(\frac{1}{k} - 1\right)\mathrm{tr}[\mathbf{K}_n]^2 - \frac{1}{2k}(\mathrm{tr}[\mathbf{M}_n]^2 - \mathrm{tr}[\mathbf{M}_n^2]) + \frac{1}{k}\mathrm{tr}[\mathbf{M}_n\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{S}_{j,*}]\\ &-\frac{1}{k}\mathrm{tr}[\mathbf{M}_n]\mathrm{tr}[\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{S}_{j,*}] + \frac{1}{2k}\mathrm{tr}[(\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{S}_{j,*})^2] - \frac{1}{2k}\mathrm{tr}[\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{S}_{j,*}]^2\end{aligned}$$

*with $\widehat{S}_n^\star = O_{\mathbb{P}}(n^{-1})$.* □

Here, we let $[V_j]$ say, denote a vector $\mathbf{V}$ with $j$-th row element $V_j$.

These results suffice to obtain asymptotic distributions for our test statistics. First, the asymptotic null distribution of $\widehat{\mathscr{B}}_n^{(1)}$ is easily obtained from Lemma 2($i$, $ii$), because

$$\widehat{\mathscr{B}}_n^{(1)} := nk^2\left(\frac{1}{2}\widehat{T}_n^2 + \frac{1}{2}\widehat{D}_n^2\right) = n(\mathrm{tr}[\mathbf{K}_n])^2 + o_{\mathbb{P}}(1).$$

Given that $\sqrt{n}\mathrm{tr}[\mathbf{K}_n]$ is asymptotically normal by Assumption 3, the asymptotic null distribution of $\widehat{\mathscr{B}}_n^{(1)}$ is the square of a normal random variable. Second, we can combine Lemma 2($i$) and Lemma 3($iii$) to obtain the asymptotic null distribution of $\widehat{\mathscr{B}}_n^{(2)}$. This follows from

$$\begin{aligned}\widehat{\mathscr{B}}_n^{(2)} = n\Bigg\{&\mathrm{tr}[\mathbf{K}_n]^2 - (\mathrm{tr}[\mathbf{M}_n]^2 - \mathrm{tr}[\mathbf{M}_n^2]) + 2\mathrm{tr}[\mathbf{M}_n\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n}-\theta_{j,*})\mathbf{S}_{j,*}] + \mathrm{tr}[(\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n}-\theta_{j,*})\mathbf{S}_{j,*})^2]\\
&- 2\mathrm{tr}[\mathbf{M}_n]\mathrm{tr}[\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n}-\theta_{j,*})\mathbf{S}_{j,*}] - \mathrm{tr}[\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n}-\theta_{j,*})\mathbf{S}_{j,*}]^2\Bigg\} + o_{\mathbb{P}}(1)\\
= n\Bigg\{&\mathrm{tr}[(\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n}-\theta_{j,*})\mathbf{S}_{j,*})^2] + 2\mathrm{tr}[(\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n}-\theta_{j,*})\mathbf{S}_{j,*})\mathbf{M}_n] + \mathrm{tr}[\mathbf{M}_n^2]\Bigg\} + o_{\mathbb{P}}(1) = n\,\mathrm{tr}[\mathbf{K}_n^2] + o_{\mathbb{P}}(1).\end{aligned}$$

The asymptotic null approximation of $\widehat{\mathscr{B}}_n^{(3)}$ is obtained in the same way. The definition of $\widehat{\mathscr{B}}_n^{(3)}$ and the approximation for $\widehat{D}_n$ give $\widehat{\mathscr{B}}_n^{(3)} = n\mathrm{tr}[\mathbf{K}_n^2] + o_{\mathbb{P}}(1)$, so that $\widehat{\mathscr{B}}_n^{(2)} - \widehat{\mathscr{B}}_n^{(3)} = o_{\mathbb{P}}(1)$ under $\mathbb{H}_0$.

We summarize these results as follows.

**Theorem 1.** *Given Assumptions 1, 2, 3, 4, 5, and* $\mathbb{H}_0$*:*

($i$) $\widehat{\mathscr{B}}_n^{(1)} = n\mathrm{tr}[\mathbf{K}_n]^2 + o_{\mathbb{P}}(1)$;

($ii$) $\widehat{\mathscr{B}}_n^{(2)} = n\mathrm{tr}[\mathbf{K}_n^2] + o_{\mathbb{P}}(1)$*; and*

($iii$) $\widehat{\mathscr{B}}_n^{(3)} = n\mathrm{tr}[\mathbf{K}_n^2] + o_{\mathbb{P}}(1)$. □

For the case in which $\widetilde{\mathbf{A}}_n$ and $\widetilde{\mathbf{B}}_n$ are sample covariance matrices, replacing $\mathbf{A}_*$ and $\mathbf{B}_*$ to provide $\widetilde{\mathbf{M}}_n$, we have a simplified result that does not involve $\widehat{\boldsymbol{\theta}}_n$: we have

**Theorem 2.** *Given Assumptions A 1, A 2, A 3, and* $\mathbb{H}_0$*,*

($i$) $\widetilde{\mathscr{B}}_n^{(1)} = n\mathrm{tr}[\widetilde{\mathbf{M}}_n]^2 + o_{\mathbb{P}}(1)$;

($ii$) $\widetilde{\mathscr{B}}_n^{(2)} = n\mathrm{tr}[\widetilde{\mathbf{M}}_n^2] + o_{\mathbb{P}}(1)$*; and*

($iii$) $\widetilde{\mathscr{B}}_n^{(3)} = n\mathrm{tr}[\widetilde{\mathbf{M}}_n^2] + o_{\mathbb{P}}(1)$*, where*

$$\widetilde{\mathscr{B}}_n^{(1)} := nk^2\left(\frac{1}{2}\widetilde{T}_n^2 + \frac{1}{2}\widetilde{D}_n^2\right);\ \ \widetilde{\mathscr{B}}_n^{(2)} := 2nk\left(\frac{1}{2}\widetilde{T}_n^2 + \widetilde{S}_n\right);\ \ \widetilde{\mathscr{B}}_n^{(3)} := 2nk\left(\frac{1}{2}\widetilde{D}_n^2 + \widetilde{S}_n\right);$$

*and* $\widetilde{\mathbf{M}}_n := \mathbf{A}_*^{-1}(\widetilde{\mathbf{B}}_n - \widetilde{\mathbf{A}}_n)$ *with* $\widetilde{T}_n := \mathrm{tr}[\widetilde{\mathbf{D}}_n]/k - 1$*;* $\widetilde{D}_n := \det[\widetilde{\mathbf{D}}_n]^{1/k} - 1$*;* $\widetilde{S}_n := \mathrm{tr}[\widetilde{\mathbf{D}}_n]/k - \det[\widetilde{\mathbf{D}}_n]^{1/k}$*; and*

$\widetilde{\mathbf{D}}_n := \widetilde{\mathbf{B}}_n\widetilde{\mathbf{A}}_n^{-1}$. □

To obtain this result, we write

$$\sqrt{n}(\widetilde{\mathbf{A}}_n - \mathbf{A}_*) = \sqrt{n}(\widehat{\Sigma}_{\mathbf{a},n} - \Sigma_{\mathbf{a},*}) - \sqrt{n}(\widehat{\boldsymbol{\mu}}_{\mathbf{a},n} - \boldsymbol{\mu}_{\mathbf{a},*})\widehat{\boldsymbol{\mu}}'_{\mathbf{a},n} - \sqrt{n}\boldsymbol{\mu}_{\mathbf{a},*}(\widehat{\boldsymbol{\mu}}_{\mathbf{a},n} - \boldsymbol{\mu}_{\mathbf{a},*})' \text{ and}$$

$$\sqrt{n}(\widetilde{\mathbf{B}}_n - \mathbf{B}_*) = \sqrt{n}(\widehat{\Sigma}_{\mathbf{b},n} - \Sigma_{\mathbf{b},*}) - \sqrt{n}(\widehat{\boldsymbol{\mu}}_{\mathbf{b},n} - \boldsymbol{\mu}_{\mathbf{b},*})\widehat{\boldsymbol{\mu}}'_{\mathbf{b},n} - \sqrt{n}\boldsymbol{\mu}_{\mathbf{b},*}(\widehat{\boldsymbol{\mu}}_{\mathbf{b},n} - \boldsymbol{\mu}_{\mathbf{b},*})',$$

and require that the multivariate CLT applies to $\sqrt{n}[\text{vech}[(\widehat{\Sigma}_{\mathbf{a},n} - \Sigma_{\mathbf{a},*})]', \text{vech}[(\widehat{\Sigma}_{\mathbf{b},n} - \Sigma_{\mathbf{b},*})]', (\widehat{\boldsymbol{\mu}}_{\mathbf{a},n} - \boldsymbol{\mu}_{\mathbf{a},*})', (\widehat{\boldsymbol{\mu}}_{\mathbf{b},n} - \boldsymbol{\mu}_{\mathbf{b},*})']'$. The difference between Theorems 1 and 2 is that, due to the structure of $\widetilde{\mathbf{A}}_n$ and $\widetilde{\mathbf{B}}_n$, $\widehat{\boldsymbol{\theta}}_n$ no longer plays an explicit role. Consequently, $\widetilde{\mathbf{M}}_n$ replaces $\mathbf{K}_n$. We state the referenced regularity conditions in the Appendix. Because the proof of Theorem 2 closely parallels that of Theorem 1, we omit this from the Appendix.

### 3.2 Asymptotic Behavior under the Alternative

We next examine the asymptotic behavior of $\widehat{T}_n$, $\widehat{D}_n$, and $\widehat{S}_n$ under the alternative. To simplify notation, for $j = 1, \ldots, \ell$, we let $\mathbf{P}_n := \mathbf{W}_n - \mathbf{U}_n$ and $\mathbf{R}_{j,*} := \mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{A}_*^{-1}\partial_j\mathbf{A}_*$; these definitions correspond to $\mathbf{M}_n$ and $\mathbf{S}_{j,*}$. Under $\mathbb{H}_0$, $\mathbf{P}_n = \mathbf{M}_n$ and $\mathbf{R}_{j,*} = \mathbf{S}_{j,*}$. We also let

$$\mathbf{L}_n := \mathbf{P}_n + \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{R}_{j,*},$$

so that $\mathbf{L}_n = \mathbf{K}_n$ under $\mathbb{H}_0$. In particular, $\mathbf{P}_n = O_{\mathbb{P}}(n^{-1/2})$ and $\mathbf{L}_n = O_{\mathbb{P}}(n^{-1/2})$ by Assumptions 3 and 5, even under the alternative. For the next result, we let $T_* := k^{-1}\text{tr}[\mathbf{D}_*] - 1$ and $D_* := \det[\mathbf{D}_*]^{1/k} - 1$. Recall that $\mathbf{D}_* := \mathbf{B}_*\mathbf{A}_*^{-1}$. We also note that $T_* \geq D_*$ from the property that the arithmetic mean is always greater than or equal to the geometric mean.

**Lemma 4.** *Given Assumptions 1, 2, 3, 4, and 5, the following hold:*

$$(i)\ \widehat{T}_n = T_* + \frac{1}{k}\text{tr}[\mathbf{L}_n\mathbf{A}_*^{-1}\mathbf{B}_*] - \frac{1}{k}\text{tr}[\mathbf{L}_n\mathbf{U}_n\mathbf{A}_*^{-1}\mathbf{B}_*] + \frac{1}{k}[\text{tr}[(\mathbf{J}_{j,n} - \mathbf{P}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)\mathbf{A}_*^{-1}\mathbf{B}_*]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + \frac{1}{2k}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla^2_{\boldsymbol{\theta}}\text{tr}[\mathbf{D}_*](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1}); \text{ and} \quad (4)$$

$$(ii)\det[\mathbf{D}_*]^{-\frac{1}{k}}\widehat{D}_n = \det[\mathbf{D}_*]^{-\frac{1}{k}}D_* + \frac{1}{k}\mathrm{tr}[\mathbf{L}_n] + \frac{1}{2k}\left(\frac{1}{k}-1\right)\mathrm{tr}[\mathbf{L}_n]^2 + \frac{1}{k}\mathrm{tr}[\mathbf{P}_n][\mathrm{tr}[\mathbf{R}_{j,*}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \quad (5)$$
$$+ \frac{1}{2k}(\mathrm{tr}[\mathbf{P}_n]^2 + \mathrm{tr}[\mathbf{U}_n^2] - \mathrm{tr}[\mathbf{W}_n^2]) + \frac{1}{k}[\mathrm{tr}[\mathbf{J}_{j,n} + \mathbf{U}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_* - \mathbf{W}_n\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)$$
$$+ \frac{1}{2k}\det[\mathbf{D}_*]^{-1}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla^2_{\boldsymbol{\theta}}\det[\mathbf{D}_*](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1}). \qquad \square$$

Lemma 3 follows easily from Lemma 4 given $\mathbb{H}_0$. Under $\mathbb{H}_0$, $T_* = D_* = 0$, $\mathbf{L}_n = \mathbf{K}_n$, $\mathbf{P}_n = \mathbf{M}_n$, $\mathbf{A}_*^{-1}\mathbf{B}_* = \mathbf{I}$, $\det[\mathbf{D}_*] = 1$, and $\mathbf{R}_{j,*} = \mathbf{S}_{j,*}$ for each $j = 1, 2, \ldots, \ell$. Using these, we obtain Lemma 3 from Lemma 4. We also note that only the second terms on the right of Eqs. (4) and (5) are $O_{\mathbb{P}}(n^{-1/2})$; the rest are $O_{\mathbb{P}}(n^{-1})$. Thus, asymptotic behavior under the alternative is mainly determined by the first two terms in each case. It follows that the behavior of $\widehat{S}_n := \widehat{T}_n - \widehat{D}_n$ is similarly determined. We have

**Corollary 1.** *Given Assumptions 1, 2, 3, 4, and 5,*

$(i)$ *if for all* $d > 0$, $\mathbf{B}_* \neq d\mathbf{A}_*$, $\widehat{S}_n = (T_* - D_*) + \frac{1}{k}\mathrm{tr}[(\mathbf{A}_*^{-1}\mathbf{B}_* - \det[\mathbf{D}_*]^{\frac{1}{k}}\mathbf{I})\mathbf{L}_n] + o_{\mathbb{P}}(n^{-1/2})$*; and*

$(ii)$ *if for some* $d_* > 0$, $\mathbf{B}_* = d_*\mathbf{A}_*$, $\widehat{S}_n = -\frac{d_*}{2k^2}\mathrm{tr}[\mathbf{L}_n]^2 + \frac{d_*}{2k}\mathrm{tr}[\mathbf{L}_n^2] + o_{\mathbb{P}}(n^{-1})$. $\square$

Although Corollary 1 follows easily from Lemma 4, we provide its proof in the Appendix to make the role of the second-order derivatives clear.

Indeed, $\mathrm{tr}[(\mathbf{A}_*^{-1}\mathbf{B}_* - \det[\mathbf{D}_*]^{\frac{1}{k}}\mathbf{I})\mathbf{L}_n] = 0$ if and only if $\mathbf{B}_*$ is proportional to $\mathbf{A}_*$. Also, if $\mathbf{B}_* = d_*\mathbf{A}_*$, then $T_* = D_* = d_* - 1$. We use this fact in obtaining Corollary 1(*ii*), which implies that $\widehat{S}_n = O_{\mathbb{P}}(n^{-1})$. This suggests that the analysis of the statistics under the alternative must separately treat the cases where $\mathbf{B}_*$ *is* proportional to $\mathbf{A}_*$ and where $\mathbf{B}_*$ is *not* proportional to $\mathbf{A}_*$. In terms of correct model specification, the literature typically assumes that for some $d_*$, $\mathbf{B}_* = d_*\mathbf{A}_*$ by assuming that conditional disturbances obey conditional homoskedasticity.

The power properties of our test statistics now follow easily. We have

$$\widehat{\mathscr{B}}_n^{(1)} = \frac{nk^2}{2}(T_*^2 + D_*^2) + nk(\mathrm{tr}[(T_*\mathbf{A}_*^{-1}\mathbf{B}_* + D_*\det[\mathbf{D}_*]^{\frac{1}{k}}\mathbf{I})\mathbf{L}_n]) + O_{\mathbb{P}}(1) \quad (6)$$

from the definition of $\widehat{\mathscr{B}}_n^{(1)}$ and Lemma 4. The first term is $O(n)$, and the second is $O_{\mathbb{P}}(\sqrt{n})$. Thus, for any $c_n = o(n)$, $\lim_{n\to\infty}\mathbb{P}(\widehat{\mathscr{B}}_n^{(1)} \geq c_n) = 1$.

As seen above, the behavior of $\widehat{\mathscr{B}}_n^{(2)}$ and $\widehat{\mathscr{B}}_n^{(3)}$ depends on whether $\mathbf{B}_*$ is proportional to $\mathbf{A}_*$. Specifically, we cannot expect power from $\widehat{S}_n$ if $\mathbf{B}_*$ is proportional to $\mathbf{A}_*$. When $\mathbf{B}_*$ is not proportional to $\mathbf{A}_*$, we have

$$\widehat{\mathscr{B}}_n^{(2)} = nk(T_*^2 + 2T_* - 2D_*) + 2n\mathrm{tr}[[(T_* + 1)\mathbf{A}_*^{-1}\mathbf{B}_* - \det[\mathbf{D}_*]^{\frac{1}{k}}\mathbf{I}]\mathbf{L}_n] + O_{\mathbb{P}}(1); \text{ and} \quad (7)$$

$$\widehat{\mathscr{B}}_n^{(3)} = nk(D_*^2 - 2D_* + 2T_*) + 2n\{(D_*+1)(D_*-1)\mathrm{tr}[\mathbf{L}_n] + \mathrm{tr}[\mathbf{A}_*^{-1}\mathbf{B}_*\mathbf{L}_n]\} + O_{\mathbb{P}}(1) \tag{8}$$

by the definitions of $\widehat{\mathscr{B}}_n^{(2)}$ and $\widehat{\mathscr{B}}_n^{(3)}$, Lemma 4, and Corollary 1($i$). The first term of $\widehat{\mathscr{B}}_n^{(2)}$ is $O(n)$ if $T_*^2 + 2T_* - 2D_* \neq 0$, and that of $\widehat{\mathscr{B}}_n^{(3)}$ is $O(n)$ if $D_*^2 - 2D_* + 2T_* \neq 0$. The second terms are $O_{\mathbb{P}}(\sqrt{n})$. Also, if $T_* \neq D_*$, $T_* > D_*$ from the inequality of the arithmetic and geometric means. Using this, we can obtain the power properties of $\widehat{\mathscr{B}}_n^{(2)}$ and $\widehat{\mathscr{B}}_n^{(3)}$ as follows: if $T_*^2 + 2T_* - 2D_* > 0$, for the same $c_n = o(n)$ as above, $\lim_{n\to\infty}\mathbb{P}(\widehat{\mathscr{B}}_n^{(2)} \geq c_n) = 1$; and if $D_*^2 + 2T_* - 2D_* > 0$, $\lim_{n\to\infty}\mathbb{P}(\widehat{\mathscr{B}}_n^{(3)} \geq c_n) = 1$. Here, we do not have to consider the absolute values of $\widehat{\mathscr{B}}_n^{(2)}$ and $\widehat{\mathscr{B}}_n^{(3)}$, as the sign of $T_*^2 + 2T_* - 2D_*$ or of $T_*^2 + 2T_* - 2D_*$ cannot be negative. A two-sided test is therefore inappropriate in general.

Furthermore, we can compare the global powers of the tests using the power properties. First, we compare $\widehat{\mathscr{B}}_n^{(2)}$ with $\widehat{\mathscr{B}}_n^{(3)}$. The leading term of $\widehat{\mathscr{B}}_n^{(2)}$ diverges more rapidly than $\widehat{\mathscr{B}}_n^{(3)}$ if and only if $|T_*| > |D_*|$. Second, the leading term of $\widehat{\mathscr{B}}_n^{(1)}$ diverges more rapidly than $\widehat{\mathscr{B}}_n^{(3)}$ if and only if $\frac{k}{2}(T_*^2 + D_*^2) > D_*^2 + 2(T_* - D_*)$. Likewise, the leading term of $\widehat{\mathscr{B}}_n^{(1)}$ diverges more rapidly than $\widehat{\mathscr{B}}_n^{(2)}$ if and only if $\frac{k}{2}(T_*^2 + D_*^2) > T_*^2 + 2(T_* - D_*)$. From these aspects, it follows that the global powers of the tests are interrelated in $3!(=6)$ different ways.

- Condition A: $|T_*| \geq |D_*|$, $\frac{k}{2}(T_*^2 + D_*^2) \geq D_*^2 + 2(T_* - D_*)$, and $\frac{k}{2}(T_*^2 + D_*^2) \geq T_*^2 + 2(T_* - D_*)$;
- Condition B: $|T_*| \geq |D_*|$, $\frac{k}{2}(T_*^2 + D_*^2) \geq D_*^2 + 2(T_* - D_*)$, and $\frac{k}{2}(T_*^2 + D_*^2) \leq T_*^2 + 2(T_* - D_*)$;
- Condition C: $|T_*| \geq |D_*|$, $\frac{k}{2}(T_*^2 + D_*^2) \leq D_*^2 + 2(T_* - D_*)$, and $\frac{k}{2}(T_*^2 + D_*^2) \leq T_*^2 + 2(T_* - D_*)$;
- Condition D: $|T_*| \leq |D_*|$, $\frac{k}{2}(T_*^2 + D_*^2) \leq D_*^2 + 2(T_* - D_*)$, and $\frac{k}{2}(T_*^2 + D_*^2) \leq T_*^2 + 2(T_* - D_*)$;
- Condition E: $|T_*| \leq |D_*|$, $\frac{k}{2}(T_*^2 + D_*^2) \leq D_*^2 + 2(T_* - D_*)$, and $\frac{k}{2}(T_*^2 + D_*^2) \geq T_*^2 + 2(T_* - D_*)$;
- Condition F: $|T_*| \leq |D_*|$, $\frac{k}{2}(T_*^2 + D_*^2) \geq D_*^2 + 2(T_* - D_*)$, and $\frac{k}{2}(T_*^2 + D_*^2) \geq T_*^2 + 2(T_* - D_*)$.

Under different conditions, the tests behave differently. For example, if $(T_*, D_*)$ satisfies Condition F, the leading term of $\widehat{\mathscr{B}}_n^{(1)}$ is greater than those of $\widehat{\mathscr{B}}_n^{(2)}$ and $\widehat{\mathscr{B}}_n^{(3)}$, and the leading term of $\widehat{\mathscr{B}}_n^{(3)}$ is greater than that of $\widehat{\mathscr{B}}_n^{(2)}$. Thus, for $c_n = o(n)$, $\lim_{n\to\infty}\mathbb{P}(\widehat{\mathscr{B}}_n^{(1)} \geq \widehat{\mathscr{B}}_n^{(3)} + c_n \geq \widehat{\mathscr{B}}_n^{(2)} + c_n) = 1$, or, for notational simplicity, we may denote this as $\widehat{\mathscr{B}}_n^{(1)} \gtrsim \widehat{\mathscr{B}}_n^{(3)} \gtrsim \widehat{\mathscr{B}}_n^{(2)}$. Table 1 summarizes the global power relationships arranged by Conditions A to F.

These global power behaviors provide guidance for selecting better tests. From the given condition, $(\widehat{T}_n - 1, \widehat{D}_n - 1) \to (T_*, D_*)$ in probability, so that if $(\widehat{T}_n - 1, \widehat{D}_n - 1)$ satisfies Condition A, say, we should expect better power properties from $\widehat{\mathscr{B}}_n^{(1)}$ than from the other tests. As another example, if $T_* = D_*$ under $\mathbb{H}_1$, Conditions A or F holds, and the others do not hold. This implies that $\widehat{\mathscr{B}}_n^{(1)}$ or $\widehat{\mathscr{B}}_n^{(2)}$ are globally most powerful.

The dimension of $\widehat{\mathbf{A}}_n$ or $\widehat{\mathbf{B}}_n$ also delivers various power properties. Figure 1 shows how the space of $(T_*, D_*)$ is partitioned into the regions satisfying Conditions A to F for $k = 2$, 3, 4, 10, 20, and 50.

For example, if $(T_*, D_*)$ belongs to the region indexed by D, this implies that Condition D governs the data. Here, only the region below the 45-degree line is considered because $T_*$ has to be greater than or equal to $D_*$. As we can see from Figure 1, the regions governed by Conditions B, C, D, and E get smaller and eventually vanish to the origin as $k$ tends to infinity. Most space is indexed by A or F. This fact implies that if $(\widehat{T}_n, \widehat{D}_n)$ is moderately different from the origin and the number of the dimension is fairly large, $\widehat{\mathscr{B}}_n^{(1)}$ or $\widehat{\mathscr{B}}_n^{(2)}$ delivers more powerful results. We can also see that the regions indexed by E and F are relatively smaller than other regions when $k$ is small.

Now suppose that for some $d_*(\neq 1)$, $\mathbf{B}_* = d_*\mathbf{A}_*$. The definitions of $\widehat{\mathscr{B}}_n^{(1)}$, $\widehat{\mathscr{B}}_n^{(2)}$, and $\widehat{\mathscr{B}}_n^{(3)}$, Lemma 4, and Corollary 1(*ii*) imply that

$$\widehat{\mathscr{B}}_n^{(1)} = nk^2(d_*-1)^2 + 2nk(d_*-1)d_*\mathrm{tr}[\mathbf{L}_n] + O_{\mathbb{P}}(1); \tag{9}$$

$$\widehat{\mathscr{B}}_n^{(2)} = nk(d_*-1)^2 + 2n(d_*-1)d_*\mathrm{tr}[\mathbf{L}_n] + O_{\mathbb{P}}(1); \text{ and} \tag{10}$$

$$\widehat{\mathscr{B}}_n^{(3)} = nk(d_*-1)^2 + 2n(d_*-1)d_*\mathrm{tr}[\mathbf{L}_n] + O_{\mathbb{P}}(1). \tag{11}$$

Alternatively, we can also obtain these results from Eqs. (6), (7), and (8) by letting $\mathbf{B}_* = d_*\mathbf{A}_*$ for $d_* \neq 1$. The first terms on the right are $O(n)$, and the second are $O_{\mathbb{P}}(\sqrt{n})$. Thus, for any $c_n = o(n)$, $\lim_{n\to\infty}\mathbb{P}(\widehat{\mathscr{B}}_n^{(2)} \geq c_n) = 1$ and $\lim_{n\to\infty}\mathbb{P}(\widehat{\mathscr{B}}_n^{(3)} \geq c_n) = 1$.

We can also compare the global powers of the tests in a similar manner to the general case. First, the leading terms of $\widehat{\mathscr{B}}_n^{(2)}$ and $\widehat{\mathscr{B}}_n^{(3)}$ are identical. This implies that their global powers are similar. Second, the leading term of $\widehat{\mathscr{B}}_n^{(1)}$ is always greater than or equal to those of $\widehat{\mathscr{B}}_n^{(2)}$ and $\widehat{\mathscr{B}}_n^{(3)}$, so that we can always expect the global power of $\widehat{\mathscr{B}}_n^{(1)}$ to be bigger than those of $\widehat{\mathscr{B}}_n^{(2)}$ and $\widehat{\mathscr{B}}_n^{(3)}$. Thus, we obtain that $\lim_{n\to\infty}\mathbb{P}(\widehat{\mathscr{B}}_n^{(1)} \geq \max[\widehat{\mathscr{B}}_n^{(2)}, \widehat{\mathscr{B}}_n^{(3)}] + c_n) = 1$. When there is additional information that for some $d_*$, $\mathbf{B}_* = d_*\mathbf{A}_*$ (*e.g.*, conditional homoskedasticity in testing correct model specification), the testing results of $\widehat{\mathscr{B}}_n^{(1)}$ should be more valuable than the other tests.

We summarize these results as follows.

**Theorem 3.** *Given Assumptions 1, 2, 3, 4, 5, and* $\mathbb{H}_1$, *for any* $c_n = o(n)$,

*(i)* $\lim_{n\to\infty}\mathbb{P}(\widehat{\mathscr{B}}_n^{(i)} \geq c_n) = 1$, *where* $i = 1, 2$, *and 3; and*

*(ii) when for all* $d > 0$, $\mathbf{B}_* \neq d\mathbf{A}_*$,

*(ii.a)* $\lim_{n\to\infty}\mathbb{P}(\widehat{\mathscr{B}}_n^{(2)} \geq \widehat{\mathscr{B}}_n^{(3)} + c_n) = 1$ *if and only if* $|T_*| > |D_*|$;

*(ii.b)* $\lim_{n\to\infty}\mathbb{P}(\widehat{\mathscr{B}}_n^{(1)} \geq \widehat{\mathscr{B}}_n^{(2)} + c_n) = 1$ *if and only if* $(\frac{k}{2}-1)T_*^2 + \frac{k}{2}D_*^2 - 2(T_* - D_*) > 0$;

*(ii.c)* $\lim_{n\to\infty}\mathbb{P}(\widehat{\mathscr{B}}_n^{(1)} \geq \widehat{\mathscr{B}}_n^{(3)} + c_n) = 1$ *if and only if* $(\frac{k}{2}-1)D_*^2 + \frac{k}{2}T_*^2 - 2(T_* - D_*) > 0$;

*(iii) when for some* $d_* > 0$, $\mathbf{B}_* = d_*\mathbf{A}_*$, $\lim_{n\to\infty}\mathbb{P}(\widehat{\mathscr{B}}_n^{(1)} \geq \max[\widehat{\mathscr{B}}_n^{(2)}, \widehat{\mathscr{B}}_n^{(3)}] + c_n) = 1$. □

The results are similar for the case when the estimators are sample covariance matrices: for the case in which $\widetilde{\mathbf{A}}_n$ and $\widetilde{\mathbf{B}}_n$ are sample covariance matrices, we have

**Theorem 4.** *Given Assumptions A 1, A 2, A 3, and* $\mathbb{H}_1$*, for any* $c_n = o(n)$*,*

*(i)* $\lim_{n\to\infty} \mathbb{P}(\widetilde{\mathscr{B}}_n^{(i)} \geq c_n) = 1$*, where* $i = 1,2,3$*;*

*(ii) when for all* $d > 0$*,* $\mathbf{B}_* \neq d\mathbf{A}_*$*,*

*(ii.a)* $\lim_{n\to\infty} \mathbb{P}(\widetilde{\mathscr{B}}_n^{(2)} \geq \widetilde{\mathscr{B}}_n^{(3)} + c_n) = 1$ *if and only if* $|T_*| > |D_*|$*;*

*(ii.b)* $\lim_{n\to\infty} \mathbb{P}(\widetilde{\mathscr{B}}_n^{(1)} \geq \widetilde{\mathscr{B}}_n^{(2)} + c_n) = 1$ *if and only if* $(\frac{k}{2} - 1)T_*^2 + \frac{k}{2}D_*^2 - 2(T_* - D_*) > 0$*;*

*(ii.c)* $\lim_{n\to\infty} \mathbb{P}(\widetilde{\mathscr{B}}_n^{(1)} \geq \widetilde{\mathscr{B}}_n^{(3)} + c_n) = 1$ *if and only if* $(\frac{k}{2} - 1)D_*^2 + \frac{k}{2}T_*^2 - 2(T_* - D_*) > 0$*;*

*(iii) when for some* $d_* > 0$*,* $\mathbf{B}_* = d_*\mathbf{A}_*$*,* $\lim_{n\to\infty} \mathbb{P}(\widetilde{\mathscr{B}}_n^{(1)} \geq \max[\widetilde{\mathscr{B}}_n^{(2)}, \widetilde{\mathscr{B}}_n^{(3)}] + c_n) = 1$. □

Because we explained earlier the key idea of Theorem 3 and because the proof of Theorem 4 closely parallels that of Theorem 3, we omit their proofs from the Appendix.

### 3.3 Asymptotic Behavior under the Local Alternative

The tests we consider here are consistent under the alternative as we analyzed in the previous subsection. Their local powers, which we consider in this subsection, can be similarly analyzed.

There can be many local alternatives to the null hypothesis. An analytically convenient local alternative is constructed by letting $\mathbf{B}_{*,n}$ approach $\mathbf{A}_*$. To keep our presentation concise, we consider the following Pitman type local alternative:

$$\mathbb{H}_a^{(1)} : \mathbf{B}_{*,n} = \mathbf{B}_* + n^{-1/2}\bar{\mathbf{B}}_* \text{ and } \mathbf{B}_* = \mathbf{A}_*,$$

where $\bar{\mathbf{B}}_* := \bar{\mathbf{B}}(\boldsymbol{\theta}_*)$, and $\bar{\mathbf{B}}(\cdot)$ is a continuously differentiable function of $\boldsymbol{\theta}$. That is, $\bar{\mathbf{B}} : \Theta \mapsto \mathbb{R}^{k\times k}$. If $\bar{\mathbf{B}}_* = \mathbf{0}$, the local alternative $\mathbb{H}_a^{(1)}$ reduces to the null hypothesis. Therefore, as $n$ tends to infinity, $\mathbf{B}_{*,n}$ approaches $\mathbf{B}_*$, and the two matrices become identical.

The motivation for considering $\mathbb{H}_a^{(1)}$ is that it has a simple form. Our statistics defined by $\widehat{T}_n$, $\widehat{D}_n$, and $\widehat{S}_n$ have the core element $\mathbf{B}_{*,n}\mathbf{A}_*^{-1}$ under $\mathbb{H}_a^{(1)}$, and this feature simplifies the analysis of $\mathbf{B}_{*,n}\mathbf{A}_*^{-1}$ without losing the insights of the tests. If we had considered the following local alternative:

$$\mathbb{H}_a^{(2)} : \mathbf{A}_{*,n} = \mathbf{A}_* + n^{-1/2}\bar{\mathbf{A}}_* \text{ and } \mathbf{B}_* = \mathbf{A}_*,$$

where $\bar{\mathbf{A}}_* := \bar{\mathbf{A}}(\boldsymbol{\theta}_*)$ and $\bar{\mathbf{A}} : \Theta \mapsto \mathbb{R}^{k\times k}$, the analysis becomes more complicated. This also implies that there are many local alternatives other than $\mathbb{H}_a^{(1)}$, and they can be used to examine the local powers of the tests in other directions.

Before examining the asymptotic behaviors of $\widehat{T}_n$, $\widehat{D}_n$, and $\widehat{S}_n$ under $\mathbb{H}_a^{(1)}$, we first fix our ideas by defining notations relevant to our discussions. We let $\mathbf{W}_{o,n} := \mathbf{B}_*^{-1}(\mathbf{B}_n - \mathbf{B}_{*,n})$, which is identical to $\mathbf{A}_*^{-1}(\mathbf{B}_n - \mathbf{B}_{*,n})$ under $\mathbb{H}_a^{(1)}$;

$$\mathbf{M}_{o,n} := \mathbf{W}_{o,n} - \mathbf{U}_n;$$

$$\mathbf{K}_{o,n} := \mathbf{M}_{o,n} + \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{S}_{j,*};$$

$$\mathbf{J}_{j,o,n} := \mathbf{G}_{j,o,n} - \mathbf{H}_{j,n} := \mathbf{B}_*^{-1}\partial_j(\mathbf{B}_n - \mathbf{B}_{*,n}) - \mathbf{H}_{j,n};$$

$$\widehat{T}_{o,n} := \frac{1}{k}\mathrm{tr}[\mathbf{K}_{o,n}(\mathbf{I} - \mathbf{U}_n)] + \frac{1}{k}[\mathrm{tr}[\mathbf{J}_{j,o,n} - \mathbf{M}_{o,n}\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + \frac{1}{2k}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla_{\boldsymbol{\theta}}^2\mathrm{tr}[\mathbf{D}_*](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*);$$

$$\begin{aligned}\widehat{D}_{o,n} := &\frac{1}{k}\mathrm{tr}[\mathbf{K}_{o,n}] + \frac{1}{2k}\left(\frac{1}{k} - 1\right)\mathrm{tr}[\mathbf{K}_{o,n}]^2 + \frac{1}{2k}(\mathrm{tr}[\mathbf{M}_{o,n}]^2 + \mathrm{tr}[\mathbf{U}_n^2] - \mathrm{tr}[\mathbf{W}_{o,n}^2])\\ &+ \frac{1}{k}[\mathrm{tr}[\mathbf{J}_{j,o,n} + \mathbf{U}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_* - \mathbf{W}_{o,n}\mathbf{A}_*^{-1}\partial_j\mathbf{B}_*]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)\\ &+ \frac{1}{k}[\mathrm{tr}[\mathbf{M}_{o,n}]\mathrm{tr}[\mathbf{S}_{j,*}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + \frac{1}{2k}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla_{\boldsymbol{\theta}}^2\det[\mathbf{D}_*](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*);\end{aligned}$$

$$\begin{aligned}\widehat{S}_{o,n} := &-\frac{1}{2k}\left(\frac{1}{k} - 1\right)\mathrm{tr}[\mathbf{K}_{o,n}]^2 - \frac{1}{2k}(\mathrm{tr}[\mathbf{M}_{o,n}]^2 - \mathrm{tr}[\mathbf{M}_{o,n}^2])\\ &+ \frac{1}{k}\mathrm{tr}[\mathbf{M}_{o,n}\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{S}_{j,*}] - \frac{1}{k}\mathrm{tr}[\mathbf{M}_{o,n}]\mathrm{tr}[\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{S}_{j,*}]\\ &+ \frac{1}{2k}\mathrm{tr}[(\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{S}_{j,*})^2] - \frac{1}{2k}\mathrm{tr}[\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{S}_{j,*}]^2.\end{aligned}$$

Note that they are identical to the previous statistics if $\bar{\mathbf{B}}_* = \mathbf{0}$. All these statistics are defined to capture the local asymptotic behaviors of the previous statistics. Mainly, their difference stems from that the key parameter of interest is not $\mathbf{B}_*$ but $\mathbf{B}_{*,n}$, so that it is now $\sqrt{n}(\mathbf{B}_n - \mathbf{B}_{*,n})$ that obeys the CLT asymptotically.

We formally impose the following assumptions.

**Assumption 6.** $(i)$ *There are symmetric matrix estimators* $\mathbf{A}_n(\cdot)$ *and* $\mathbf{B}_n(\cdot)$ *consistent for* $\mathbf{A}(\cdot)$ *and* $\mathbf{B}(\cdot)$*, respectively, uniformly on* $\Theta$*; and*

$(ii)$ $\sqrt{n}[(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)', \mathrm{vech}[\mathbf{A}_n - \mathbf{A}_*]', \mathrm{vech}[\mathbf{B}_n - \mathbf{B}_{*,n}]']' \overset{A}{\sim} N(\mathbf{0}, \Omega_*)$*, where* $\mathbf{A}_n := \mathbf{A}_n(\boldsymbol{\theta}_*)$*,* $\mathbf{B}_n := \mathbf{B}_n(\boldsymbol{\theta}_*)$*, and* $\Omega_*$ *is a* $(k^2 + k + \ell) \times (k^2 + k + \ell)$ *positive semi-definite matrix.* □

**Assumption 7.** *For* $j = 1, \ldots, \ell$*,* $\mathbf{H}_{j,n} = O_{\mathbb{P}}(n^{-1/2})$ *and* $\mathbf{G}_{j,o,n} = O_{\mathbb{P}}(n^{-1/2})$*.* □

**Assumption 8.** *The symmetric mapping* $\bar{\mathbf{B}} : \Theta \mapsto \mathbb{R}^{k\times k}$ *is in* $\mathscr{C}^{(1)}(\Theta)$ *and is such that* $\bar{\mathbf{B}}_* := \bar{\mathbf{B}}(\boldsymbol{\theta}_*)$ *is positive definite.* □

Assumptions 6 and 7 are provided to replace Assumptions 3 and 5, respectively. We replaced the key parameter $\mathbf{B}_*$ with $\mathbf{B}_{*,n}$, so that we can expect $\mathbf{W}_{o,n}$, $\mathbf{M}_{o,n}$, $\mathbf{K}_{o,n}$, $\mathbf{J}_{j,o,n}$, $\widehat{T}_{o,n}$, $\widehat{D}_{o,n}$, and $\widehat{S}_{o,n}$ under $\mathbb{H}_a^{(1)}$ to have asymptotic behaviors identical to those of $\mathbf{W}_n$, $\mathbf{M}_n$, $\mathbf{K}_n$, $\mathbf{J}_{j,n}$, $\widehat{T}_n$, $\widehat{D}_n$, and $\widehat{S}_n$ under $\mathbb{H}_0$, respectively. Therefore, all of them are $O_{\mathbb{P}}(n^{-1/2})$ under $\mathbb{H}_a^{(1)}$. Furthermore, Assumption 8 is added to endow the local parameter $\bar{\mathbf{B}}_*$ with an appropriate structure of positive definiteness.

The following lemma shows the asymptotic expansions of the statistics under the local alternative.

**Lemma 5.** *Given Assumptions 1, 2, 4, 6, 7, 8, and* $\mathbb{H}_a^{(1)}$,

$$(i)\quad \widehat{T}_n = \widehat{T}_{o,n} + \frac{1}{\sqrt{n}k}\{\mathrm{tr}[\mathbf{N}_*] - \mathrm{tr}[\mathbf{N}_*\mathbf{U}_n] + [\mathrm{tr}[\mathbf{A}_*^{-1}(\partial_j\bar{\mathbf{B}}_*) - \mathbf{N}_*\mathbf{A}_*^{-1}(\partial_j\mathbf{A}_*)]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)\} + o_{\mathbb{P}}(n^{-1}),$$

*where* $\mathbf{N}_* := \mathbf{B}_*^{-1}\bar{\mathbf{B}}_*$;

$$(ii)\quad \widehat{D}_n = \widehat{D}_{o,n} + \frac{1}{\sqrt{n}k}\{\mathrm{tr}[\mathbf{N}_*] - \mathrm{tr}[\mathbf{N}_*\mathbf{W}_{o,n}] + [\mathrm{tr}[\mathbf{C}_{j,*}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)\} + \frac{1}{\sqrt{n}k^2}\mathrm{tr}[\mathbf{N}_*]\mathrm{tr}[\mathbf{K}_{o,n}] - \frac{1}{2nk}\mathrm{tr}[\mathbf{N}_*^2] + \frac{1}{2nk^2}\mathrm{tr}[\mathbf{N}_*]^2 + o_{\mathbb{P}}(n^{-1}),$$

*where* $\mathbf{C}_{j,*} := \mathbf{B}_*^{-1}\partial_j\bar{\mathbf{B}}_* - \mathbf{N}_*\mathbf{B}_*^{-1}(\partial_j\mathbf{B}_*)$; *and*

$$(iii)\quad \widehat{S}_n = \widehat{S}_{o,n} + \frac{1}{\sqrt{n}k}\{\mathrm{tr}[\mathbf{N}_*\mathbf{M}_{o,n}] + [\mathrm{tr}[\mathbf{N}_*\mathbf{S}_{j,*}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)\} - \frac{1}{\sqrt{n}k^2}\mathrm{tr}[\mathbf{N}_*]\mathrm{tr}[\mathbf{K}_{o,n}] + \frac{1}{2nk}\mathrm{tr}[\mathbf{N}_*^2] - \frac{1}{2nk^2}\mathrm{tr}[\mathbf{N}_*]^2 + o_{\mathbb{P}}(n^{-1}).$$ □

Proving Lemma 5 is not difficult. Using Lemma 4 and the expansion that $(\mathbf{I} - n^{-1/2}\mathbf{B}_*^{-1}(-\bar{\mathbf{B}}_*))^{-1} = \mathbf{I} - n^{-1/2}\mathbf{B}_*^{-1}\bar{\mathbf{B}}_* + n^{-1}\mathbf{B}_*^{-1}\bar{\mathbf{B}}_*\mathbf{B}_*^{-1}\bar{\mathbf{B}}_* + \ldots$ provides the proof of Lemma 5. Here, this expansion is applicable because for sufficiently large $n$, $\|n^{-1/2}\mathbf{B}_*^{-1}(-\bar{\mathbf{B}}_*)\| < 1$.

The local asymptotic behaviors of the statistics in Lemma 5 are not identical. Lemmas 5(*i* and *ii*) imply that $\sqrt{n}k\widehat{T}_n = \mathrm{tr}[\mathbf{N}_*] + O_{\mathbb{P}}(1)$ and $\sqrt{n}k\widehat{D}_n = \mathrm{tr}[\mathbf{N}_*] + O_{\mathbb{P}}(1)$, whereas $2nk^2\widehat{S}_n = k\,\mathrm{tr}[\mathbf{N}_*^2] - \mathrm{tr}[\mathbf{N}_*]^2 + O_{\mathbb{P}}(1)$ by Lemma 5(*iii*). Using these features, we derive the following theorem.

**Theorem 5.** *Given Assumptions 1, 2, 3, 4, 5, 6, and* $\mathbb{H}_a^{(1)}$,

(*i*) $\widehat{\mathscr{B}}_n^{(1)} = \mathrm{tr}[\mathbf{N}_* + \sqrt{n}\mathbf{K}_{o,n}]^2 + o_{\mathbb{P}}(1)$;

(*ii*) $\widehat{\mathscr{B}}_n^{(2)} = \mathrm{tr}[(\mathbf{N}_* + \sqrt{n}\mathbf{K}_{o,n})^2] + o_{\mathbb{P}}(1)$*; and*

(*iii*) $\widehat{\mathscr{B}}_n^{(3)} = \mathrm{tr}[(\mathbf{N}_* + \sqrt{n}\mathbf{K}_{o,n})^2] + o_{\mathbb{P}}(1)$. □

We note that Theorem 1 can also be derived from Theorem 5 if we let $\bar{\mathbf{B}}_* = \mathbf{0}$. In addition, the local parameter $\mathbf{N}_*$ is added to $\sqrt{n}\mathbf{K}_{o,n}$ as a location parameter. Theorem 5 further implies that the local power of $\widehat{\mathscr{B}}_n^{(1)}$ is not automatically acquired. The product of two positive-definite matrices is not necessarily positive definite, so that $\mathbf{N}_* := \mathbf{B}_*^{-1}\bar{\mathbf{B}}_*$ can have $\mathrm{tr}[\mathbf{N}_*] = 0$, implying that $\widehat{\mathscr{B}}_n^{(1)}$ may not have local power for such a case. Similarly, if $\mathrm{tr}[\mathbf{N}_*^2] = 0$, $\widehat{\mathscr{B}}_n^{(2)}$ and $\widehat{\mathscr{B}}_n^{(3)}$ do not have local powers, either. For them to have their local powers, it is necessary to have $\mathrm{tr}[\mathbf{N}_*] \neq 0$ and $\mathrm{tr}[\mathbf{N}_*^2] \neq 0$.

When $\widetilde{\mathbf{A}}_n$ and $\widetilde{\mathbf{B}}_n$ are sample covariance matrices, we have

**Corollary 2.** *Given Assumptions A 1, A 2, A 4, A 5, and* $\mathbb{H}_a^{(1)}$*,*

(*i*) $\widetilde{\mathscr{B}}_n^{(1)} = \mathrm{tr}[(\mathbf{N}_* + \sqrt{n}\widetilde{\mathbf{M}}_{o,n})]^2 + o_{\mathbb{P}}(1)$*;*

(*ii*) $\widetilde{\mathscr{B}}_n^{(2)} = \mathrm{tr}[(\mathbf{N}_* + \sqrt{n}\widetilde{\mathbf{M}}_{o,n})^2] + o_{\mathbb{P}}(1)$*; and*

(*iii*) $\widetilde{\mathscr{B}}_n^{(3)} = \mathrm{tr}[(\mathbf{N}_* + \sqrt{n}\widetilde{\mathbf{M}}_{o,n})^2] + o_{\mathbb{P}}(1)$ *and* $\widetilde{\mathscr{B}}_n^{(2)} - \widetilde{\mathscr{B}}_n^{(3)} = o_{\mathbb{P}}(1)$*, where* $\widetilde{\mathbf{M}}_{o,n} := \mathbf{A}_*^{-1}(\widetilde{\mathbf{B}}_n - \mathbf{B}_{*,n}) - \mathbf{A}_*^{-1}(\widetilde{\mathbf{A}}_n - \mathbf{A}_*)$. □

As before, $\widetilde{\mathbf{M}}_{o,n}$ replaces $\mathbf{K}_{o,n}$ because $\widehat{\boldsymbol{\theta}}_n$ no longer plays an explicit role. The referenced regularity conditions are in the Appendix. As the proof of Corollary 2 is parallel to that of Theorem 5, we omit it from the Appendix.

### 3.4 The Trace-Determinant Tests and the Likelihood Ratio Tests

In this subsection, we consider how the trace-determinant tests are associated with the LR test statistic under the local alternative. The LR test statistic gives reasonable power in all directions under the local alternative, and it is uniformly most powerful under some restrictive assumptions (*e.g.*, proposition 15.2 of van der Vaart, 2000). This motivates us to compare the trace-determinant test statistics with the LR test statistic.

For the purpose of comparison, we first assume a particular distribution. Specifically, an independently and identically distributed (IID) multivariate normal random variable is assumed. We assume this because the LR test can be easily computed for the equality of two matrices. In particular, the normal distribution condition enables the LR test to be associated with the trace and determinant tests. As it turns out, our tests can be reformulated into a test which is asymptotically equivalent to the LR test.

We consider two different null hypotheses to examine the LR test. First, Nagao (1967) and Nagarsenker and Pillai (1973) tested whether the covariance matrix is equal to a specified matrix under the same data condition. We exploit their results to examine the LR test. Second, Mauchly (1940) assumed a similar condition to ours and examined the LR statistic to test whether the covariance matrix of the random variable is a diagonal matrix with the same diagonal elements. We consider the LR test under the same condition and examine how the LR test is associated with our test statistics.

Before investigating the LR test, we first define statistics relevant to our investigations and provide their asymptotic behaviors under the local alternative. We first let $\widehat{Q}_n := \ln[\mathrm{tr}[\widehat{\mathbf{D}}_n]/k]$ and $\widehat{L}_n := \ln[\det(\widehat{\mathbf{D}}_n)]/k$. They are the logarithms of the trace and the determinant, respectively, so that $\widehat{T}_n$ and $\widehat{D}_n$ now correspond to $\widehat{Q}_n$ and $\widehat{L}_n$, respectively. We next let $\widehat{W}_n := \widehat{Q}_n - \widehat{L}_n$ and $\widehat{M}_n := \widehat{T}_n - \widehat{L}_n$. The motivations of these statistics are identical to $\widehat{S}_n$. As the same roles are played by $\widehat{Q}_n$ and $\widehat{T}_n$, we are able to define two different test statistics with identical roles. The following lemma provides their asymptotic behaviors under the local alternative.

**Lemma 6.** *Given Assumptions 1, 2, 4, 6, 7, 8, and* $\mathbb{H}_a^{(1)}$,

$$(i)\ \widehat{L}_n = \frac{1}{\sqrt{n}k}\mathrm{tr}[\mathbf{N}_*] - \frac{1}{\sqrt{n}k}\mathrm{tr}[\mathbf{N}_*\mathbf{W}_{o,n}] + \frac{1}{\sqrt{n}k}[\mathrm{tr}[\mathbf{C}_{j,*}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) - \frac{1}{2nk}\mathrm{tr}[\mathbf{N}_*^2] - \frac{1}{2k^2}\mathrm{tr}[\mathbf{K}_{o,n}]^2 + \widehat{D}_{o,n} + o_{\mathbb{P}}(n^{-1});$$

$$(ii)\ \widehat{Q}_n = \frac{1}{\sqrt{n}k}\mathrm{tr}[\mathbf{N}_*] - \frac{1}{\sqrt{n}k}\mathrm{tr}[\mathbf{N}_*\mathbf{U}_n] + \frac{1}{\sqrt{n}k}[\mathrm{tr}[\mathbf{A}_*^{-1}\partial_j\bar{\mathbf{B}}_* - \mathbf{N}_*\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) - \frac{1}{2nk}\mathrm{tr}[\mathbf{N}_*^2] - \frac{1}{2k^2}\mathrm{tr}[\mathbf{K}_{o,n}]^2 - \frac{1}{\sqrt{n}k^2}\mathrm{tr}[\mathbf{N}_*]\mathrm{tr}[\mathbf{K}_{o,n}] + \widehat{T}_{o,n} + o_{\mathbb{P}}(n^{-1});$$

*(iii)* $\widehat{M}_n = \frac{1}{2nk}\mathrm{tr}[(\mathbf{N}_* + \sqrt{n}\mathbf{K}_{o,n})^2] + +o_{\mathbb{P}}(n^{-1})$*; and*

*(iv)* $\widehat{W}_n = \frac{1}{2nk}\mathrm{tr}[(\mathbf{N}_* + \sqrt{n}\mathbf{K}_{o,n})^2] - \frac{1}{2nk^2}\mathrm{tr}[\mathbf{N}_* + \sqrt{n}\mathbf{K}_{o,n}]^2 + o_{\mathbb{P}}(n^{-1})$. □

Proving Lemma 6 is straightforward. We note that $\widehat{L}_n = \ln[\widehat{D}_n + 1]$ and $\widehat{Q}_n = \ln[\widehat{T}_n + 1]$, and Lemma 5 already provided the asymptotic behaviors of $\widehat{D}_n$ and $\widehat{T}_n$. We exploit them to prove Lemma 6(*i* and *ii*). The asymptotic behaviors of $\widehat{M}_n$ and $\widehat{W}_n$ are obtained by combining these results with Lemma 5(*i* and *ii*). The statistics $\widehat{M}_n$ and $\widehat{W}_n$ form the expansion of the LR statistic as we detail below.

In this section, we also provide assumptions distinct from those in the previous section.

**Assumption 9.** (*i*) $\{\mathbf{X}_t \in \mathbb{R}^k\}$ *is a set of IID multivariate normal random variables such that* $\mathbf{X}_t \sim N(\boldsymbol{\theta}_*, \mathbf{B}_*)$ *such that* $\boldsymbol{\theta}_* \in \mathbb{R}^k$. □

Here, it follows that $k = \ell$ by the normal distribution condition. Although this assumption is too strong for general data and we can easily generalize the assumption, we stick to this condition so that a direct comparison of the tests can be made in a straightforward manner. Given Assumption 9, if we let $L_n(\boldsymbol{\theta},\mathbf{B})$ be the likelihood, the ML estimator is

$$\widehat{\boldsymbol{\theta}}_n := \frac{1}{n}\sum_{t=1}^{n}\mathbf{X}_t \quad \text{and} \quad \widehat{\mathbf{B}}_n := \frac{1}{n}\sum_{t=1}^{n}(\mathbf{X}_t-\widehat{\boldsymbol{\theta}}_n)(\mathbf{X}_t-\widehat{\boldsymbol{\theta}}_n)',$$

and the maximum-likelihood is obtained as $L_n(\widehat{\boldsymbol{\theta}}_n,\widehat{\mathbf{B}}_n) = (2\pi)^{-nk/2}(\det[\widehat{\mathbf{B}}_n])^{-n/2}\exp(-nk/2)$. In addition to this, we further impose the following assumption on an estimator $\widehat{\mathbf{A}}_n$.

**Assumption 10.** $(i)$ *There is a symmetric matrix estimator $\mathbf{A}_n(\cdot)$ consistent for $\mathbf{A}(\cdot)$ uniformly on $\Theta$;*

$(ii)$ *the symmetric matrix mapping $\mathbf{A}:\Theta\mapsto\mathbb{R}^{k\times k}$ is in $\mathscr{C}^{(2)}(\Theta)$ and is such that $\mathbf{A}_* := \mathbf{A}(\boldsymbol{\theta}_*)$ is positive definite;*

$(iii)$ *for $j=1,\ldots,\ell$, $\partial_j\mathbf{A}_n(\cdot)$ is consistent for $\partial_j\mathbf{A}(\cdot)$ uniformly on $\Theta$;*

$(iv)$ *for $i,j=1,\ldots,\ell$, $\partial^2_{ji}\mathbf{A}_n(\cdot)$ is consistent for $\partial^2_{ji}\mathbf{A}(\cdot)$;*

$(v)$ *for $j=1,\ldots,\ell$, $\mathbf{H}_{j,n} = O_{\mathbb{P}}(n^{-1/2})$;*

$(vi)$ *$\sqrt{n}[(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_*)', \text{vech}[\mathbf{A}_n-\mathbf{A}_*]', \text{vech}[\mathbf{B}_n-\mathbf{B}_{*,n}]']' \overset{A}{\sim} N(\mathbf{0},\Omega_*)$, where $\mathbf{A}_n := \mathbf{A}_n(\boldsymbol{\theta}_*)$, $\mathbf{B}_n := \mathbf{B}_n(\boldsymbol{\theta}_*)$, and $\Omega_*$ is a $(k^2+k+\ell)\times(k^2+k+\ell)$ positive semi-definite matrix; and*

$(vii)$ *The symmetric mapping $\bar{\mathbf{B}}:\Theta\mapsto\mathbb{R}^{k\times k}$ is in $\mathscr{C}^{(1)}(\Theta)$ and is such that $\bar{\mathbf{B}}_* := \bar{\mathbf{B}}(\boldsymbol{\theta}_*)$ is positive definite.* □

The conditions in Assumption 10 are virtually identical assumptions on $\widehat{\mathbf{A}}_n$ in the previous sections. We slightly modified the structure to fit Assumption 9.

When testing the unknown covariance matrix $\mathbf{B}_*$, we consider it to be identical to $\mathbf{A}_*$, which can be consistently estimated by $\widehat{\mathbf{A}}_n$. That is, we are interested in testing $\mathbf{B}_* = \mathbf{A}_*$ as before, so that we are supposing the same environment as in Theorem 5.

Given this, the LR test can be easily obtained. If we let $L_n(\boldsymbol{\theta},\mathbf{B})$ be the likelihood, the maximum-likelihood constrained by $\mathbf{B}=\widehat{\mathbf{A}}_n$ is obtained as $L_n(\widehat{\boldsymbol{\theta}}_n,\widehat{\mathbf{A}}_n) = (2\pi)^{-nk/2}(\det[\widehat{\mathbf{A}}_n])^{-n/2}\exp\left(-\frac{nk}{2}\text{tr}[\widehat{\mathbf{A}}_n^{-1}\widehat{\mathbf{B}}_n]\right)$, and the LR test defined as $\mathscr{LR}^{(1)}_n := 2\{\ln[L_n(\widehat{\boldsymbol{\theta}}_n,\widehat{\mathbf{B}}_n)]-\ln[L_n(\widehat{\boldsymbol{\theta}}_n,\widehat{\mathbf{A}}_n)]\}$ has the following asymptotic behavior under the local alternative.

**Theorem 6.** *Given Assumptions 9, 10, and $\mathbb{H}^{(1)}_a$,*

*(i) $\mathscr{LR}^{(1)}_n = \frac{1}{2}\text{tr}[(\mathbf{N}_*+\sqrt{n}\mathbf{K}_{o,n})^2] + o_{\mathbb{P}}(n^{-1})$;*

*(ii) $\mathscr{LR}^{(1)}_n = \frac{1}{2}\widehat{\mathscr{B}}^{(2)}_n + o_{\mathbb{P}}(n^{-1})$; and*

*(iii) $\mathscr{LR}^{(1)}_n = \frac{1}{2}\widehat{\mathscr{B}}^{(3)}_n + o_{\mathbb{P}}(n^{-1})$.* □

Therefore, Theorem 6$(i)$ implies that the LR test is asymptotically equivalent to $nk\widehat{M}_n$ under the local alternative.

Mauchly (1940) tested the sphericity assumption under the same condition as above. That is, he tested whether for some $d_*$, $\mathbf{B}_* = d_*\mathbf{A}_*$. Here, $\mathbf{B}_*$, $\mathbf{A}_*$, and $d_*$ are unknown, and $\mathbf{A}_*$ can be consistently estimated by $\widehat{\mathbf{A}}_n$ as before, so that the relevant local alternative hypothesis is given as follows: for some $d_*$, $\mathbb{H}_a^{(s)} : \mathbf{B}_{*,n} = \mathbf{B}_* + n^{-1/2}\bar{\mathbf{B}}_*$ and $\mathbf{B}_* = d_*\mathbf{A}_*$.

The sphericity assumption can be associated with our test statistics, too. The constrained maximum-likelihood is obtained by maximizing $L_n(\boldsymbol{\theta}, d\widehat{\mathbf{A}}_n)$ with respect to $\boldsymbol{\theta}$ and $d$, and it turns out equal to $L_n(\widetilde{\boldsymbol{\theta}}_n, \widetilde{d}_n\widehat{\mathbf{A}}_n) = (2\pi)^{-nk/2}(\mathrm{tr}[\widehat{\mathbf{A}}_n^{-1}\widehat{\mathbf{B}}_n]/k)^{-nk/2}(\det[\widehat{\mathbf{A}}_n])^{-n/2}\exp(-nk/2)$, where $(\widetilde{\boldsymbol{\theta}}_n, \widetilde{d}_n)$ is the constrained ML estimator. Specifically, $(\widetilde{\boldsymbol{\theta}}_n, \widetilde{d}_n) = (\widehat{\boldsymbol{\theta}}_n, \mathrm{tr}[\widehat{\mathbf{A}}_n^{-1}\widehat{\mathbf{B}}_n]/k)$, and the LR test is obtained as $\mathscr{LR}_n^{(2)} := 2\{\ln[L_n(\widehat{\boldsymbol{\theta}}_n, \widehat{\mathbf{B}}_n)] - \ln[L_n(\widetilde{\boldsymbol{\theta}}_n, \widetilde{d}_n\widehat{\mathbf{A}}_n)]\}$. The LR test has the following asymptotic behavior under the local alternative.

**Theorem 7.** *Given Assumptions 9, 10, and* $\mathbb{H}_a^{(s)}$,

*(i)* $\mathscr{LR}_n^{(2)} = \frac{1}{2}(\mathrm{tr}[(\mathbf{N}_* + \sqrt{n}\mathbf{K}_{o,n})^2] - \frac{1}{k}\mathrm{tr}[\mathbf{N}_* + \sqrt{n}\mathbf{K}_{o,n}]^2) + o_{\mathbb{P}}(n^{-1})$;

*(ii)* $\mathscr{LR}_n^{(2)} = \frac{1}{2}(\widehat{\mathscr{B}}_n^{(2)} - \frac{1}{k}\widehat{\mathscr{B}}_n^{(1)}) + o_{\mathbb{P}}(n^{-1})$; *and*

*(iii)* $\mathscr{LR}_n^{(2)} = \frac{1}{2}(\widehat{\mathscr{B}}_n^{(3)} - \frac{1}{k}\widehat{\mathscr{B}}_n^{(1)}) + o_{\mathbb{P}}(n^{-1})$. □

Theorem 7($i$) implies that the LR test is asymptotically equivalent to $nk\widehat{W}_n$ under the local alternative.

We can see that the trace-determinant tests form the major components of the LR tests obtained under different circumstances from Theorems 6 and 7. They provide several further implications. First, the meanings of the trace-determinant tests become clear by using the LR test statistics. If we let $\mathscr{LR}_n^{(3)} := 2\{\ln[L_n(\widetilde{\boldsymbol{\theta}}_n, \widetilde{d}_n\widehat{\mathbf{A}}_n)] - \ln[L_n(\widetilde{\boldsymbol{\theta}}_n, \widehat{\mathbf{A}}_n)]\}$, $\mathscr{LR}_n^{(1)} \equiv \mathscr{LR}_n^{(2)} + \mathscr{LR}_n^{(3)}$, and this implies that $\mathscr{LR}_n^{(1)}$ is the sum of $\mathscr{LR}_n^{(2)}$ and $\mathscr{LR}_n^{(3)}$ such that $\mathscr{LR}_n^{(3)}$ tests whether the sphericity parameter ($d_*$) is one. Second, the asymptotic behavior of $\mathscr{LR}_n^{(3)}$ is determined from this. That is, $\mathscr{LR}_n^{(3)} = \frac{1}{2k}\mathrm{tr}[\mathbf{N}_* + \sqrt{n}\mathbf{K}_{o,n}]^2 + o_{\mathbb{P}}(n^{-1})$ under the local alternative, so that $\mathscr{LR}_n^{(3)} = \frac{1}{2k}\widehat{\mathscr{B}}_n^{(1)} + o_{\mathbb{P}}(n^{-1})$. Finally, therefore, $\frac{1}{2}(\widehat{\mathscr{B}}_n^{(2)} - \frac{1}{k}\widehat{\mathscr{B}}_n^{(1)})$ or $\frac{1}{2}(\widehat{\mathscr{B}}_n^{(3)} - \frac{1}{k}\widehat{\mathscr{B}}_n^{(1)})$ tests the sphericity condition, and $\frac{1}{2k}\widehat{\mathscr{B}}_n^{(1)}$ tests the sphericity parameter condition under the given conditions, so that testing the equal covariance matrix condition can be accomplished by summing the two test statistics: $\frac{1}{2}(\widehat{\mathscr{B}}_n^{(2)} - \frac{1}{k}\widehat{\mathscr{B}}_n^{(1)})$ and $\frac{1}{2k}\widehat{\mathscr{B}}_n^{(1)}$; or $\frac{1}{2}(\widehat{\mathscr{B}}_n^{(3)} - \frac{1}{k}\widehat{\mathscr{B}}_n^{(1)})$ and $\frac{1}{2k}\widehat{\mathscr{B}}_n^{(1)}$.

### 3.5 Implementation Considerations

The previous literature provides useful guidance for implementing our tests. As Taylor (1987) and Orme (1990), among others, have pointed out, the level distortions of the classical IMTs can be huge. A major reason for this is the ill-conditioning of the estimated asymptotic covariance matrix used to form the IMT (see, *e.g.*, Golden, Henley, White, and Kashner, 2013). We may therefore expect that our test statistics will suffer from similar finite-sample level distortions. Chesher and Spady (1991) examined higher-order expansions of the IMT as a way to reduce the finite-sample level distortion. Their Monte Carlo experiments showed that these methods can work well. As Horowitz (1994) pointed out, however, this approach is extremely cumbersome because it requires use of higher-order cumulants, which differ from model to model. This suggests that higher-order approximations and methods that estimate an asymptotic covariance matrix should be avoided.

The parametric bootstrap proposed by Horowitz (1994) avoids both of these pitfalls, and, as Horowitz showed, gives tests with negligible finite-sample level distortions, even for small sample sizes and when the tests are non-pivotal. The parametric bootstrap is also appealing because it can deliver well-behaved asymptotic critical values for our statistics, despite their non-standard asymptotic distributions and non-pivotal properties. Furthermore, the Monte Carlo experiments by Horowitz showed that powers obtained by the parametric bootstrap are very close to those obtained by correcting level distortions; on this fact, he suggested using the parametric bootstrap to compare powers of IMT statistics. In particular, the parametric bootstrap is appealing when comparing various IMTs, given that the maximum-likelihood-based approach is hard to apply to construct IMTs. Thus, we implement the parametric bootstrap to apply our tests.

The use of the parametric bootstrap is theoretically justified by the fact that the asymptotic approximations of $\widehat{T}_n$, $\widehat{D}_n$, and $\widehat{S}_n$ are already derived. More specifically, Lemma 4 and Corollary 1 show that the distributions of $\widehat{T}_n$, $\widehat{D}_n$, and $\widehat{S}_n$ are mainly determined by their own location parameters and $\mathbf{L}_n := \mathbf{B}_*^{-1}(\mathbf{B}_n - \mathbf{B}_*) - \mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_*) + \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{R}_{j,*}$, which are zeros and $\mathbf{K}_n$, respectively, under $\mathbb{H}_0$. Furthermore, Assumption 3(*ii*) implies that $\mathbf{L}_n$ asymptotically follows a normal distribution, implying that the asymptotic distributions of $\widehat{T}_n$, $\widehat{D}_n$, and $\widehat{S}_n$ are well defined under the null and are first-order correct. These observations, along with the arguments given by Horowitz (1994), justify use of the parametric bootstrap.

Although the parametric bootstrap is well known, we explain its procedure in order for our paper to be self-contained and to fix our test statistics. We proceed as follows:

1. Given the model $\mathscr{M} := \{f(\cdot|\cdot,\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$ for $Y_t|\mathbf{X}_t$, obtain the (Q)MLE $\widehat{\boldsymbol{\theta}}_n$ by maximizing $L_n(\cdot)$ on $\Theta$, where $L_n(\boldsymbol{\theta}) := n^{-1}\sum_{t=1}^{n}\ell_t(\boldsymbol{\theta})$ and $\ell_t(\boldsymbol{\theta}) := \log[f(Y_t|\mathbf{X}_t,\boldsymbol{\theta})]$, with IID $\{(Y_t,\mathbf{X}_t')' \in \mathbb{R}^{1+d}\}$;

2. Estimate $\mathbf{A}_*$ and $\mathbf{B}_*$ as $\widehat{\mathbf{A}}_n := -\nabla^2_{\boldsymbol{\theta}}L_n(\widehat{\boldsymbol{\theta}}_n)$ and $\widehat{\mathbf{B}}_n := n^{-1}\sum_{t=1}^{n}\nabla_{\boldsymbol{\theta}}\ell_t(\widehat{\boldsymbol{\theta}}_n)\nabla'_{\boldsymbol{\theta}}\ell_t(\widehat{\boldsymbol{\theta}}_n)$. Using these, compute $\widehat{\mathscr{B}}_n^{(1)}$, $\widehat{\mathscr{B}}_n^{(2)}$, and $\widehat{\mathscr{B}}_n^{(3)}$;

3. For each $\mathbf{X}_t$, generate $B$ independent samples of $Y_t^b$ from $f(\cdot|\mathbf{X}_t, \widehat{\boldsymbol{\theta}}_n)$, yielding data sets $(Y_t^b, \mathbf{X}_t)$, $t = 1, 2, \ldots, n$, $b = 1, 2, \ldots, B$;

4. Form the bootstrap (Q)MLE $\widehat{\boldsymbol{\theta}}_n^b$ by maximizing $L_n^b(\cdot) := n^{-1}\sum_{t=1}^n \ell_t^b(\cdot)$, where $\ell_t^b(\boldsymbol{\theta}) := \log[f(Y_t^b \mid \mathbf{X}_t, \boldsymbol{\theta})]$. Compute $\mathbf{A}_n^b := -\nabla_{\boldsymbol{\theta}}^2 L_n^b(\widehat{\boldsymbol{\theta}}_n^b)$ and $\mathbf{B}_n^b := n^{-1}\sum_{t=1}^n \nabla_{\boldsymbol{\theta}} \ell_t^b(\widehat{\boldsymbol{\theta}}_n^b)\nabla_{\boldsymbol{\theta}}' \ell_t^b(\widehat{\boldsymbol{\theta}}_n^b)$, and use these to form test statistics $\mathscr{B}_n^{(1),b}$, $\mathscr{B}_n^{(2),b}$, and $\mathscr{B}_n^{(3),b}$;

5. Estimate $p$-values by $\widehat{p}_n^{(1)} := B^{-1}\sum_{b=1}^B I(\mathscr{B}_n^{(1),b} \geq \widehat{\mathscr{B}}_n^{(1)})$, $\widehat{p}_n^{(2)} := B^{-1}\sum_{b=1}^B I(\mathscr{B}_n^{(2),b} \geq \widehat{\mathscr{B}}_n^{(2)})$, and $\widehat{p}_n^{(3)} := B^{-1}\sum_{b=1}^B I(\mathscr{B}_n^{(3),b} \geq \widehat{\mathscr{B}}_n^{(3)})$, respectively.

6. If the estimated $p$-values are less than the pre-specified level of significance, we reject the null; otherwise, we do not.

# 4 Monte Carlo Experiments

In this section, we examine the performance of our statistics when applied to a variety of parametric models, estimated by quasi-maximum-likelihood applied to IID data. We consider four different models: linear regression, exponential duration, probit, and Tobit.[2]

Our Monte Carlo experiments replicate the steps of the parametric bootstrap multiple times. When a test statistic, $\widehat{\mathscr{B}}_n^{(1)}$, say, is evaluated using the parametric bootstrap, we also write this as $\widehat{\mathscr{B}}_n^{(1,p)}$.

## 4.1 Linear Regression

Suppose that data are generated as

$$Y_t = \mathbf{X}_t'\boldsymbol{\beta}_* + U_t,$$

where $U_t \mid \mathbf{X}_t \sim N(0, \sigma_*^2)$. We estimate the unknown $\boldsymbol{\beta}_*$ using the normal MLE with the typical model element $Y_t \mid \mathbf{X}_t \sim N(\mathbf{X}_t'\boldsymbol{\beta}, \sigma^2)$. We consider the case when $\mathbf{X}_t = (1, X_t)'$ and $X_t \sim N(0,1)$.

We compare our tests to other familiar test statistics. Specifically, we consider Chesher's (1983) and Lancaster's (1984) IMT,[3] denoted $IM_n$. We also consider Jarque and Bera's (1987) test for the normality of $U_t$, denoted $JB_n$. Both $IM_n$ and $JB_n$ obey standard chi-square distributions asymp-

[2] GAUSS codes for the linear regression, exponential duration, Weibull duration, probit, and Tobit models are downloadable at the following URL: http://web.yonsei.ac.kr/jinseocho/research.htm.

[3] To construct this IMT, we need to include the first-order and second-order derivatives as testing components. Nevertheless, including the second-order derivative with respect to the intercept coefficient yields a singular covariance matrix. We thus omit this redundant term.

totically under the null. We also apply the parametric bootstrap to these tests, denoted $IM_n^p$ and $JB_n^p$.

We present our simulation results in Table 2. Here, we take $\boldsymbol{\beta}_* = (0.5, 1)'$ with $\sigma_*^2 = 1$. We let $B = 1{,}000$ and perform $20{,}000$ Monte Carlo replications. The simulation results under the null can be summarized as follows. First, our tests based on the parametric bootstrap perform quite well, with nominal levels closely comparable to actual levels even for $n = 50$. As expected, $IM_n$ performs quite poorly, exhibiting the familiar level distortions. In contrast, $JB_n$ performs much better, although it does suffer modest level distortion for smaller samples. The parametric bootstrap nicely fixes these problems. Only $IM_n$ is adversely affected by increasing the number of parameters.

To examine power, we conduct further Monte Carlo experiments using six different alternative DGPs:

- ALT 1: $U_t|\mathbf{X}_t \sim N(0, \exp(\mathbf{X}_t'\boldsymbol{\beta}_*))$;
- ALT 2: $U_t|\mathbf{X}_t \sim N(0, \exp(2\mathbf{X}_t'\boldsymbol{\beta}_*))$;
- ALT 3: $U_t|\mathbf{X}_t \sim MxN(-1, 1; 1, 1; 0.5)$;
- ALT 4: $U_t|\mathbf{X}_t \sim MxN(-1.5, 1; 1.5, 1; 0.5)$;
- ALT 5: $U_t|\mathbf{X}_t \sim t_{30}$; and
- ALT 6: $U_t|\mathbf{X}_t \sim t_{20}$.

Here, $Z \sim MxN(a, b; c, d; p)$ denotes a mixed normal distribution where $Z \sim N(a, b)$ with probability $p$ and $Z \sim N(c, d)$ with probability $1 - p$.

Although the normal model is correctly specified for the conditional mean, the conditional variance or conditional distribution is misspecified for these alternatives. The first two alternatives exhibit conditional heteroskedasticity, with the first closer to the null than the second. The next two alternatives have PDFs with two peaks and dispersed distributions. The third is closer to the null than the fourth. The final two alternatives have fat tails, with the first closer to the null than the second.

We present our simulation results in Table 3. The results are for the nominal 5% level, obtained using the parametric bootstrap, and are somewhat nuanced. First, with conditional heteroskedasticity, our tests perform better than the others. $\widehat{\mathscr{B}}_n^{(2,p)}$ and $\widehat{\mathscr{B}}_n^{(3,p)}$ outperform $\widehat{\mathscr{B}}_n^{(1,p)}$. Second, with normal mixture disturbances, the best test is $IM_n^p$, followed by $\widehat{\mathscr{B}}_n^{(1,p)}$, which noticeably outperforms $\widehat{\mathscr{B}}_n^{(2,p)}$ and $\widehat{\mathscr{B}}_n^{(3,p)}$ in smaller samples. $JB_n^p$ underperforms in smaller samples, although it catches up in larger samples. The $t-$distribution alternatives are harder for all tests to detect, as these alternatives are not as far from the null. The $JB_n^p$ test is best, followed fairly closely by our trace-determinant tests. Overall, it appears that our tests are consistent against these alternatives, sometimes performing best and otherwise performing respectably.

We also conduct other experiments to examine the global powers of the trace-determinant tests. More specifically, Table 1 shows how the most powerful test is globally determined by the conditions for $(T_*, D_*)$. As the sample size increases, the leading terms in Eqs. (6), (7), and (8) become the main factors to determine the powers. We investigate this by Monte Carlo experiments. For this purpose, we use the following procedure: First, for each DGP and sample size, we estimate $(T_*, D_*)$ by Monte Carlo simulations. The number of replications is 10,000. This experiment is conducted to avoid the difficulties in computing $(T_*, D_*)$ analytically. Second, we predict the most powerful test by using this estimate and Table 1. Third, we compare the empirical powers in Table 3 with our predictions and indicate the results as follows: if the prediction is correct (resp. wrong), we denote this by "○" (resp. "●"); and if all three tests reject the null hypothesis, so that we cannot say which one is the most powerful test, we denote this by "△." If the global powers are effectively determined by Table 1, "○" should eventually appear as the sample size increases.

The results of this experiment are contained in Table 4. For ALTs 1, 3, 4, and 5, "○" is observed more frequently, or it appears eventually as the sample size increases before all three tests reject the null. On the other hand, for ALTs 2 and 6, the most powerful test is not correctly predicted by the global power patterns.

These wrong predictions occur mainly because $(T_*, D_*)$ is too close to the other regions. For example, the estimated $(T_*, D_*)$ of ALT 6 is approximately $(0.0608, 0.0544)$ when $n = 2{,}000$, and it belongs to the region indexed by C, which predicts $\widehat{\mathscr{B}}_n^{(2,p)}$ to be the most powerful test statistic. Nevertheless, this value is very close to the origin, whose neighbors are the regions indexed by A, B, and D. We can see this from the second panel of Figure 1. Under Conditions A, B, and D, $\widehat{\mathscr{B}}_n^{(2,p)}$, $\widehat{\mathscr{B}}_n^{(1,p)}$, and $\widehat{\mathscr{B}}_n^{(3,p)}$ are the most powerful test statistics, respectively, so that the null hypothesis can also be frequently rejected by $\widehat{\mathscr{B}}_n^{(1,p)}$ and $\widehat{\mathscr{B}}_n^{(3,p)}$. This leads to the wrong prediction, and it also explains why one test does not have a dominant power over the other tests, as we can see from Table 3.

We also conducted parallel simulations with $\mathbf{X}_t$ drawn from a uniform distribution and/or $\mathbf{X}_t = (1, X_t, Z_t)'$ with $(X_t, Z_t)' \sim N(\mathbf{0}, \mathbf{I}_2)$. The results were almost identical, so we do not report them here.

### 4.2 Exponential Duration

Next, we consider duration structures following a conditionally exponential distribution. The DGP is

$$Y_t|(\delta_t, \mathbf{X}_t) \sim \text{Exp}(\delta_t \exp(\mathbf{X}_t' \boldsymbol{\beta}_*)),$$

where 'Exp' stands for the exponential distribution, and $\mathbf{X}_t = X_t$ with $X_t \sim N(0,1)$, as before. We assume that $\delta_t$ is not observable and that the researcher observes only $Y_t$ and $\mathbf{X}_t$. The conditional distribution of $Y_t|\mathbf{X}_t$ is the exponential if $\delta_t$ is a constant with probability one. Otherwise, the simple exponential model is misspecified. We consider the following null and alternative DGPs for $\delta_t$:

- NULL (Constant case): $\delta = 1$ with probability 1;
- ALT 1 (Discrete): $\delta_t \sim$ DM(0.7370, 1.9296; 0.5);
- ALT 2 (Gamma): $\delta_t \sim$ Gamma(5, 5);
- ALT. 3 (Log-normal): $\delta_t \sim$ Log-normal($-\ln(1.2)/2, \ln(1.2)$);
- ALT 4 (Uniform I): $\delta_t \sim$ Uniform[0.30053, 2.3661]; and
- ALT 5 (Uniform II): $\delta_t \sim$ Uniform[1, 5/3],

where DM$(a, b; p)$ denotes a discrete mixture with $P(\delta_t = a) = p$ and $P(\delta_t = b) = 1 - p$. These DGPs are identical to those considered by Cho and White (2010) and Cho, Cheong, and White (2011). In each case, we let $\boldsymbol{\beta}_*$ be the unit vector.

The alternative DGPs represent various forms of unobservable heterogeneity, a subject of considerable interest in the literature. As Heckman (1984) pointed out, unobservable heterogeneity can have serious adverse consequences for estimation and inference. Among others, Lancaster (1979) developed a test for unobserved heterogeneity; Lancaster's test is designed to detect misspecifications in the direction of a gamma distribution. The other three DGPs also represent alternative lines of inquiry in the literature. See Cho and White (2010) for further discussion.

We suppose that the researcher specifies a model for $Y_t|\mathbf{X}_t$ that treats $\delta_t$ as a constant:

$$\{f(\cdot|\cdot : \alpha, \boldsymbol{\beta}) : f(y|\mathbf{x} : \alpha, \boldsymbol{\beta}) = \alpha \exp(\mathbf{x}'\boldsymbol{\beta}) \exp(-\alpha \exp(\mathbf{x}'\boldsymbol{\beta}) y), (\alpha, \boldsymbol{\beta}) \in \Theta\},$$

where $\Theta$ is a parameter space for $(\alpha, \boldsymbol{\beta})$ such that each parameter has a lower bound and upper bound equal to 0 and 10, respectively. Thus, the above model is correctly specified only for the first DGP and misspecified for the others. The parameter space is relatively large, ensuring that the probability limits of the (Q)MLEs are interior to $\Theta$. We test model misspecification using our tests, the Chesher

(1983) and Lancaster (1984) IMT, and Lancaster's (1979) Lagrange multiplier test for unobserved heterogeneity, which we denote $LM_n$. As before, we also consider parametric bootstrap versions of the IMTs and Lancaster's (1979) LM test, denoted $LM_n^p$.

Our simulation results appear in Tables 5 and 6. Table 5 focuses on the null behavior of the tests, whereas Table 6 shows the behaviors under the alternative. Here, $B = 1{,}000$, and we perform $20{,}000$ Monte Carlo replications. We summarize the results as follows. The levels for all parametric bootstrap procedures are good, with actual levels close to nominal levels. Except for $IM_n$, we observe similar null behavior in Table 5.

Second, we see from Table 6 that our tests outperform $IM_n^p$ and $LM_n^p$ for every alternative DGP. Among our tests, the most powerful test for smaller samples is $\widehat{\mathscr{B}}_n^{(2,p)}$, followed by $\widehat{\mathscr{B}}_n^{(3,p)}$ and $\widehat{\mathscr{B}}_n^{(1,p)}$, in that sequence. In particular, the finite sample power of $LM_n^p$ is very weak. Even when we apply the parametric bootstrap to $LM_n$, it can yield poor performance under the alternative. This indicates that choosing the appropriate IMT is important to have better finite sample properties. Our IMTs appear to avoid the poor properties. This ordering changes in larger samples.

We finally examine the predictions of the most powerful test by using the global powers. The process is the same as in the linear regression case. The prediction results are contained in Table 7, which shows that many alternative DGPs exhibit correct predictions as the sample size increases. On the other hand, a wrong prediction is obtained for ALT 5.

The reason for this wrong prediction is the same as before. That is, $(T_*, D_*)$ is too close to other regions. The estimated $(T_*, D_*)$ using 2,000 observations is approximately $(0.0401, 0.0382)$, and this value belongs to the region indexed by C but is also very close to the regions indexed by A, B, and D. We can see this from the first panel of Figure 1. Thus, the null hypothesis is frequently rejected by the other test statistics.

We also conducted experiments using $\mathbf{X}_t \sim N(\mathbf{0}, \mathbf{I}_2)$ and/or the Weibull specifications studied by Cho and White (2010) and Cho, Cheong, and White (2011) for the same DGPs, and we obtained similar level and power patterns. We omit reporting these results for the sake of brevity. Thus, our parametric bootstrap tests appear promising for detecting unobserved heterogeneity in duration models.

### 4.3 Probit

For our third experiment, we consider a probit specification, as did Horowitz (1994).

Again, we consider $\mathbf{X}_t = (1, X_t)'$, where $X_t \sim N(0,1)$. The probit model has the typical element

$$f(Y_t|\mathbf{X}_t;\boldsymbol{\beta}) = \Phi(\mathbf{X}_t'\boldsymbol{\beta}),$$

where $\Phi$ is the standard normal CDF. We take $\boldsymbol{B} = [-5, 5]$.

We examine the following four DGPs in our experiments:

- NULL: $Y_t|\mathbf{X}_t \sim \text{Probit}(\mathbf{X}_t'\boldsymbol{\beta}_*)$;
- ALT 1: $Y_t|\mathbf{X}_t \sim \text{Probit}(\mathbf{X}_t'\boldsymbol{\beta}_*/\exp(0.5\mathbf{X}_t'\boldsymbol{\beta}_*))$;
- ALT 2: $Y_t|\mathbf{X}_t \sim \text{Probit}((\mathbf{X}_t'\boldsymbol{\beta}_*)^2)$; and
- ALT 3: $Y_t|\mathbf{X}_t \sim \text{Logit}(\mathbf{X}_t'\boldsymbol{\beta}_*)$,

where 'Probit($\mathbf{X}_t'\boldsymbol{\beta}_*$)' denotes that the conditional distribution of $Y_t|\mathbf{X}_t$ satisfies $P(Y_t = 1|\mathbf{X}_t) = \Phi(\mathbf{X}_t'\boldsymbol{\beta}_*)$. The model is correctly specified for the first DGP but misspecified for the rest. 'Logit($\mathbf{X}_t'\boldsymbol{\beta}_*$)' denotes that the conditional distribution of $Y_t|\mathbf{X}_t$ satisfies $P(Y_t = 1|\mathbf{X}_t) = \{1 + \exp(\mathbf{X}_t'\boldsymbol{\beta}_*)\}^{-1}$. We take $\boldsymbol{\beta}_* = (0.5, 1)'$.

We conduct the same simulations as before, and we again compare our tests to parametric bootstrap versions of the Chesher (1983) and Lancaster (1984) IMT. As the asymptotic covariance matrix for the full IMT is singular, we drop redundant terms.

We present our simulation results in Tables 8 and 9. Table 8 presents the null behaviors of the test statistics, and Table 9 examines the alternatives. Here, the number of bootstrap replications is $B = 500$. We conduct 10,000 Monte Carlo replications under the null and 3,000 under the alternative.

We summarize the results as follows. As before, the levels of all the parametric bootstrap tests accord well with nominal values, and again the raw $IM_n$ statistic performs poorly. With regard to power, $IM_n^p$ performs best for the third DGP in smaller samples, although it is dominated in the other two cases. Overall, the trace-determinant tests perform respectably.

We also examine the global powers of the test statistics as before. For all alternative DGPs, Table 10 shows that as the sample size increases, "○" is observed more frequently or eventually before all three tests reject the null, implying that Table 1 correctly predicts the most powerful test statistic.

We also conducted experiments using $\mathbf{X}_t := (1, X_t, Z_t)'$ and $(X_t, Z_t)' \sim N(\mathbf{0}, \mathbf{I}_2)$ and obtained similar level and power patterns. We omit reporting these results for the sake of brevity.

### 4.4 Tobit

As a final experiment, we consider the Tobit model. Horowitz (1994) also examined the performance of the IMT for the Tobit model, describing it as a successful application of the parametric bootstrap.

The experimental design is similar to that for the probit model. As before, $\mathbf{X}_t = (1, X_t)'$, where $X_t \sim N(0,1)$. The Tobit model has the typical element

$$Y_t = \max[0, \mathbf{X}_t'\boldsymbol{\beta} + U_t],$$

where $U_t \sim N(0, \sigma^2)$. Again, we take $\boldsymbol{B} = [-5, 5]$, and we let $\sigma$ lie between 0.1 and 10. The unknown parameters $\boldsymbol{\beta}$ and $\sigma$ are estimated by (Q)ML estimation. We consider the following four DGPs:

- NULL: $Y_t = \max[0, \mathbf{X}_t'\boldsymbol{\beta}_* + U_t]$ and $U_t|\mathbf{X}_t \sim N(0,1)$;
- ALT 1: $Y_t = \max[0, \mathbf{X}_t'\boldsymbol{\beta}_* + U_t]$ and $U_t|\mathbf{X}_t \sim N(0, \exp(0.5\mathbf{X}_t'\boldsymbol{\beta}_*))$;
- ALT 2: $Y_t = \max[0, (\mathbf{X}_t'\boldsymbol{\beta}_*)^2 + U_t]$ and $U_t|\mathbf{X}_t \sim N(0,1)$; and
- ALT 3: $Y_t = \max[0, \mathbf{X}_t'\boldsymbol{\beta}_* + U_t]$ and $U_t|\mathbf{X}_t \sim t_{30}$.

We let $\boldsymbol{\beta}_* = (0,1)'$. Thus, the Tobit model is correctly specified for the first DGP and misspecified for the others.

We again compare our tests with parametric bootstrap versions of the Chesher (1983) and Lancaster (1984) IMT. As before, we drop redundant terms.

The results appear in Tables 11 and 12. Table 11 contains results for levels, and Table 12 reports results for the alternatives. As for the probit case, the number of bootstrap replications is $B = 500$, with 10,000 Monte Carlo replications under the null and 3,000 under the alternatives.

Again, we see that all the parametric bootstrap tests have levels close to nominal, while $IM_n$ does not perform well. In each alternative, our new tests outperform $IM_n^p$.

We also predict the most powerful test statistic using the global powers as before. For all the alternative DGPs we consider here, Table 13 shows that as the sample size increases, Table 1 correctly predicts the most powerful test statistic. This is affirmed from the fact that as the sample size increases, "○" is more frequently observed, or it appears eventually before "△" is observed.

We also conducted experiments using $\mathbf{X}_t := (1, X_t, Z_t)'$ and $(X_t, Z_t)' \sim N(\mathbf{0}, \mathbf{I}_2)$ and obtained similar level and power patterns. We omit reporting these results for the sake of brevity.

# 5 Conclusion

We provide a new characterization of the equality of two positive-definite matrices **A** and **B**, namely that $\det(\mathbf{BA}^{-1})^{1/k} = k^{-1}\text{tr}(\mathbf{BA}^{-1}) = 1$, and we use this to propose several new, computationally convenient statistical tests for the equality of two unknown positive-definite matrices. Our primary focus is on testing the information matrix equality (*e.g.* White, 1982, 1994), where the matrices **A** and **B** depend on underlying parameters that require estimation. We characterize the asymptotic behavior of our test statistics under the null, alternative, and local alternative, and we investigate their finite-sample performance for a variety of models: linear regression, exponential duration, probit, and Tobit. Although our statistics have non-standard asymptotic distributions, the parametric bootstrap easily accommodates these and delivers critical values that provide admirable level behavior, even in samples as small as $n = 50$. Reinforcing Horowitz's (1994) findings, we find that the parametric bootstrap also eradicates the well-documented poor level performance of the traditional IMT in all of our experiments. Our new tests often have better power than the traditional IMT; when they do not, they nevertheless perform respectably. In particular, the trace-determinant tests are always better than the conventional IMTs for duration data.

The simplicity, reliable level performance, and respectable power properties of our new parametric bootstrap trace-determinant IMTs remove the obstacles that might have previously dissuaded researchers from applying an IMT. Our findings here, and the versatility of information matrix testing methods generally, lead us to strongly recommend routine application of parametric bootstrap-based IMTs.

# A Appendix

## A.1 Proofs

**Proof of Lemma 1:** If $\mathbf{A} = \mathbf{B}$, then clearly $[\det(\mathbf{A}^{-1/2}\mathbf{BA}^{-1/2})]^{1/k} = \text{tr}(\mathbf{A}^{-1/2}\mathbf{BA}^{-1/2})/k = 1$.

For the converse, we have that $[\det(\mathbf{A}^{-1/2}\mathbf{BA}^{-1/2})]^{1/k} = \text{tr}(\mathbf{A}^{-1/2}\mathbf{BA}^{-1/2})/k$. Equivalently, $(\prod_{j=1}^{k} \lambda_j)^{1/k} = k^{-1}\sum_{j=1}^{k}\lambda_j$, where $\lambda_j$, $j = 1,\ldots,k$, are the real non-negative eigenvalues of $\mathbf{A}^{-1/2}\mathbf{BA}^{-1/2}$. It is well-known that equality of the geometric and arithmetic means follows if and only if the elements of the means are identical, *i.e.*, $\lambda_1 = \ldots = \lambda_k = \lambda$, say. As $[\det(\mathbf{A}^{-1/2}\mathbf{BA}^{-1/2})]^{1/k} = 1$, we have $\det(\mathbf{A}^{-1/2}\mathbf{BA}^{-1/2})^{1/k} = \lambda = 1$. That is, $\mathbf{A}^{-1/2}\mathbf{BA}^{-1/2}$ has eigenvalues identically equal to 1, so that $\mathbf{A}^{-1/2}\mathbf{BA}^{-1/2} = \mathbf{CIC}' = \mathbf{I}$, where **C** is the orthonormal matrix of the eigenvectors of $\mathbf{A}^{-1/2}\mathbf{BA}^{-1/2}$. Now $\mathbf{A}^{-1/2}\mathbf{BA}^{-1/2} = \mathbf{I}$ implies $\mathbf{A}^{1/2}\mathbf{A}^{-1/2}\mathbf{BA}^{-1/2}\mathbf{A}^{1/2} = \mathbf{A}^{1/2}\mathbf{A}^{1/2}$, which simplifies to $\mathbf{B} = \mathbf{A}$. ■

Before proving our results, we first state some useful lemmas, allowing us to avoid repeating certain arguments. We provide proofs only for those lemmas that are not elementary.

**Lemma A1.** *Given Assumptions 1, 2, and 3,*

*(i)* $\mathbf{A}_n^{-1} - \mathbf{A}_*^{-1} = -\mathbf{U}_n\mathbf{A}_*^{-1} + \mathbf{U}_n^2\mathbf{A}_*^{-1} + o_{\mathbb{P}}(n^{-1})$;

*(ii)* $\mathbf{B}_n^{-1} - \mathbf{B}_*^{-1} = -\mathbf{W}_n\mathbf{B}_*^{-1} + \mathbf{W}_n^2\mathbf{B}_*^{-1} + o_{\mathbb{P}}(n^{-1})$;

*(iii)* $\mathbf{D}_n - \mathbf{D}_* = \mathbf{B}_*\mathbf{P}_n(\mathbf{I} - \mathbf{U}_n)\mathbf{A}_*^{-1} + o_{\mathbb{P}}(n^{-1})$*, where* $\mathbf{P}_n := \mathbf{W}_n - \mathbf{U}_n$*; and*

*(iv) if* $\mathbb{H}_0$ *further holds,* $\mathbf{D}_n - \mathbf{D}_* = \mathbf{B}_*\mathbf{M}_n(\mathbf{I} - \mathbf{U}_n)\mathbf{A}_*^{-1} + o_{\mathbb{P}}(n^{-1})$. □

**Proof of Lemma A1:** ($i$) We apply exercise 13.21 of Abadir and Magnus (2005): $d\mathbf{X}^{-1} = -\mathbf{X}^{-1}(d\mathbf{X})\mathbf{X}^{-1}$, so that $d^2\mathbf{X}^{-1} = -d\mathbf{X}^{-1}(d\mathbf{X})\mathbf{X}^{-1} - \mathbf{X}^{-1}(d^2\mathbf{X})\mathbf{X}^{-1} - \mathbf{X}^{-1}(d\mathbf{X})d\mathbf{X}^{-1}$. This implies that the second-order approximation of $\mathbf{X}^{-1} - \mathbf{X}_o^{-1}$ is obtained as $d\mathbf{X}^{-1} + \frac{1}{2}d^2\mathbf{X}^{-1} = -\mathbf{X}_o^{-1}(d\mathbf{X})\mathbf{X}_o^{-1} + \{\mathbf{X}_o^{-1}(d\mathbf{X})\mathbf{X}_o^{-1}(d\mathbf{X})\mathbf{X}_o^{-1}\} - \frac{1}{2}\mathbf{X}_o^{-1}(d^2\mathbf{X})\mathbf{X}_o^{-1}$, where $d\mathbf{X}$ captures $(\mathbf{X} - \mathbf{X}_o)$.

We now obtain the second-order approximation of $\mathbf{A}_n^{-1} - \mathbf{A}_*^{-1}$ by letting $\mathbf{X}_o = \mathbf{A}_*$ and $d\mathbf{X} = \mathbf{A}_n - \mathbf{A}_*$, where the remainder term is $o_{\mathbb{P}}(n^{-1})$ from the fact that $\mathbf{A}_n - \mathbf{A}_* = O_{\mathbb{P}}(n^{-1/2})$. We also use the definition of $\mathbf{U}_n := \mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_*)$ and obtain the desired result.

($ii$) The proof is identical to ($i$). We can replace $\mathbf{A}_n$, $\mathbf{A}_*$, and $\mathbf{U}_n$ with $\mathbf{B}_n$, $\mathbf{B}_*$, and $\mathbf{W}_n$, respectively.

($iii$) We note the following identity: $\mathbf{D}_n - \mathbf{D}_* \equiv (\mathbf{B}_n - \mathbf{B}_*)\mathbf{A}_*^{-1} + (\mathbf{B}_n - \mathbf{B}_*)(\mathbf{A}_n^{-1} - \mathbf{A}_*^{-1}) + \mathbf{B}_*(\mathbf{A}_n^{-1} - \mathbf{A}_*^{-1})$. By reasoning analogous to Lemma A1($i$) for $(\mathbf{A}_n^{-1} - \mathbf{A}_*^{-1})$ we obtain $(\mathbf{B}_n - \mathbf{B}_*)\mathbf{U}_n^2\mathbf{A}_*^{-1} = O_{\mathbb{P}}(n^{-3/2})$. This yields the desired approximation.

($iv$) We have $\mathbf{P}_n = \mathbf{M}_n$ in ($iii$) under $\mathbb{H}_0$. This completes the proof. ■

**Lemma A2.** *Given Assumptions 1, 2, and 3,*

*(i)* $\det[\mathbf{A}_n] - \det[\mathbf{A}_*] = \det[\mathbf{A}_*](\mathrm{tr}[\mathbf{U}_n] + \frac{1}{2}\mathrm{tr}[\mathbf{U}_n]^2 - \frac{1}{2}\mathrm{tr}[\mathbf{U}_n^2]) + o_{\mathbb{P}}(n^{-1})$*;*

*(ii)* $\det[\mathbf{B}_n] - \det[\mathbf{B}_*] = \det[\mathbf{B}_*](\mathrm{tr}[\mathbf{W}_n] + \frac{1}{2}\mathrm{tr}[\mathbf{W}_n]^2 - \frac{1}{2}\mathrm{tr}[\mathbf{W}_n^2]) + o_{\mathbb{P}}(n^{-1})$*;*

*(iii)* $\det[\mathbf{A}_n]^{-1} - \det[\mathbf{A}_*]^{-1} = -\det[\mathbf{A}_*]^{-1}(\mathrm{tr}[\mathbf{U}_n] + \frac{1}{2}\mathrm{tr}[\mathbf{U}_n]^2 - \frac{1}{2}\mathrm{tr}[\mathbf{U}_n^2]) + o_{\mathbb{P}}(n^{-1})$*;*

*(iv)* $\det[\mathbf{D}_n] - \det[\mathbf{D}_*] = \det[\mathbf{D}_*]\{\mathrm{tr}[\mathbf{P}_n] + \frac{1}{2}\mathrm{tr}[\mathbf{P}_n]^2 - \frac{1}{2}(\mathrm{tr}[\mathbf{W}_n^2] - \mathrm{tr}[\mathbf{U}_n^2])\} + o_{\mathbb{P}}(n^{-1})$*; and*

*(v) if* $\mathbb{H}_0$ *further holds,* $\det[\mathbf{D}_n] - 1 = \mathrm{tr}[\mathbf{M}_n] + \frac{1}{2}(\mathrm{tr}[\mathbf{M}_n])^2 - \frac{1}{2}(\mathrm{tr}[\mathbf{W}_n^2] - \mathrm{tr}[\mathbf{U}_n^2]) + o_{\mathbb{P}}(n^{-1})$. □

**Proof of Lemma A2:** ($i$) By exercise 13.32 of Abadir and Magnus (2005): $d|\mathbf{X}| = |\mathbf{X}|\mathrm{tr}(\mathbf{X}^{-1}d\mathbf{X})$. This also implies that $d^2|\mathbf{X}| = d|\mathbf{X}|\mathrm{tr}(\mathbf{X}^{-1}d\mathbf{X}) + |\mathbf{X}|d\mathrm{tr}(\mathbf{X}^{-1}d\mathbf{X}) = |\mathbf{X}|\mathrm{tr}(\mathbf{X}^{-1}d\mathbf{X})^2 + |\mathbf{X}|\mathrm{tr}(d\mathbf{X}^{-1}\ d\mathbf{X} + \mathbf{X}^{-1}d^2\mathbf{X})$. We also note that $d\mathbf{X}^{-1} = -\mathbf{X}^{-1}(d\mathbf{X})\mathbf{X}^{-1}$ by exercise 13.21 of Abadir and Magnus (2005). Thus, $d^2|\mathbf{X}| = |\mathbf{X}|\{\mathrm{tr}(\mathbf{X}^{-1}d\mathbf{X})^2 - \mathrm{tr}[\mathbf{X}^{-1}(d\mathbf{X})\mathbf{X}^{-1}d\mathbf{X}] + \mathrm{tr}[\mathbf{X}^{-1}d^2\mathbf{X}]\}$, which implies that the second-order approximation of $|\mathbf{X}| - |\mathbf{X}_o|$ is obtained as $d|\mathbf{X}| + \frac{1}{2}d^2|\mathbf{X}| = |\mathbf{X}_o|\mathrm{tr}(\mathbf{X}_o^{-1}d\mathbf{X}) + \frac{1}{2}|\mathbf{X}_o|\{\mathrm{tr}(\mathbf{X}_o^{-1}d\mathbf{X})^2 - \mathrm{tr}[\mathbf{X}_o^{-1}(d\mathbf{X})\mathbf{X}_o^{-1}d\mathbf{X}] + \mathrm{tr}[\mathbf{X}_o^{-1}d^2\mathbf{X}]\}$, where $d\mathbf{X}$ captures $(\mathbf{X} - \mathbf{X}_o)$.

We now apply this result by letting $\mathbf{X}_o$, $|\mathbf{X}_o|$, and $d\mathbf{X}$ be $\mathbf{A}_*$, $\det[\mathbf{A}_*]$, and $\mathbf{A}_n - \mathbf{A}_*$, respectively. Thus,

$$\det[\mathbf{A}_n] - \det[\mathbf{A}_*] = \det[\mathbf{A}_*]\mathrm{tr}[\mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_*)] \\ + \frac{1}{2}\det[\mathbf{A}_*]\{\mathrm{tr}[\mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_*)]^2 - \mathrm{tr}[\mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_*)\mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_*)]\} + o_{\mathbb{P}}(n^{-1}),$$

where the final terms are obtained from Assumption 3, that $(\mathbf{A}_n - \mathbf{A}_*) = O_{\mathbb{P}}(n^{-1/2})$. Finally, we note that $\mathbf{U}_n := \mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_*)$, so that $\det[\mathbf{A}_n] - \det[\mathbf{A}_*] = \det[\mathbf{A}_*](\mathrm{tr}[\mathbf{U}_n] + \frac{1}{2}\mathrm{tr}[\mathbf{U}_n]^2 - \frac{1}{2}\mathrm{tr}[\mathbf{U}_n^2]) + o_{\mathbb{P}}(n^{-1})$.

(*ii*) We apply the same argument, replacing $\mathbf{A}_n$, $\mathbf{A}_*$, and $\mathbf{U}_n$ with $\mathbf{B}_n$, $\mathbf{B}_*$, and $\mathbf{W}_n$, respectively.

(*iii*) Applying Taylor's expansion gives $\det[\mathbf{A}_n]^{-1} - \det[\mathbf{A}_*]^{-1} = -\det[\mathbf{A}_*]^{-2}(\det[\mathbf{A}_n] - \det[\mathbf{A}_*]) + \det[\mathbf{A}_*]^{-3}(\det[\mathbf{A}_n] - \det[\mathbf{A}_*])^2 + o_{\mathbb{P}}(n^{-1})$ using Assumption 3, that $(\mathbf{A}_n - \mathbf{A}_*) = O_{\mathbb{P}}(n^{-1/2})$. Replacing $(\det[\mathbf{A}_n] - \det[\mathbf{A}_*])$ with the expression in Lemma A2(*i*) and rearranging gives the desired result.

(*iv*) We note that $\det[\mathbf{D}_n] - \det[\mathbf{D}_*] = \det[\mathbf{B}_n]/\det[\mathbf{A}_n] - \det[\mathbf{B}_*]/\det[\mathbf{A}_*]$ by the definitions of $\mathbf{D}_n$ and $\mathbf{D}_*$, so that the following identity holds:

$$\det[\mathbf{D}_n] - \det[\mathbf{D}_*] \equiv \frac{1}{\det[\mathbf{A}_*]}(\det[\mathbf{B}_n] - \det[\mathbf{B}_*]) \\ + \left(\frac{1}{\det[\mathbf{A}_n]} - \frac{1}{\det[\mathbf{A}_*]}\right)(\det[\mathbf{B}_n] - \det[\mathbf{B}_*]) + \left(\frac{1}{\det[\mathbf{A}_n]} - \frac{1}{\det[\mathbf{A}_*]}\right)\det[\mathbf{B}_*].$$

Using LemmaA2(*ii* and *iii*) and rearranging the terms according to the rate of convergence gives the result.

(*v*) We note that $\mathbf{P}_n = \mathbf{M}_n$ under $\mathbb{H}_0$. The desired result follows from Lemma A2(*iv*). ■

**Lemma A3.** *Given Assumptions 1, 2, and 3,*

*(i)* $\nabla_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_n] = [\mathrm{tr}[\mathbf{R}_{j,n}\mathbf{A}_n^{-1}\mathbf{B}_n]]$, *where* $\mathbf{R}_{j,n} := \mathbf{B}_n^{-1}\partial_j\mathbf{B}_n - \mathbf{A}_n^{-1}\partial_j\mathbf{A}_n$*;*

*(ii)* $\nabla_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*] = [\mathrm{tr}[\mathbf{R}_{j,*}\mathbf{A}_*^{-1}\mathbf{B}_*]]$, *where* $\mathbf{R}_{j,*} := \mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{A}_*^{-1}\partial_j\mathbf{A}_*$*; and*

*(iii) if* $\mathbb{H}_0$ *further holds,* $\nabla_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*] = [\mathrm{tr}[\mathbf{S}_{j,*}]]$. □

**Proof of Lemma A3:** (*i*) By the fact that $\partial_j\mathbf{A}_n^{-1} = -\mathbf{A}_n^{-1}(\partial_j\mathbf{A}_n)\,\mathbf{A}_n^{-1}$ and the definition of $\mathbf{D}_n := \mathbf{B}_n\mathbf{A}_n^{-1}$, the desired result holds.

(*ii*) The proof is the same as that of (*i*). That is, $\mathbf{D}_* := \mathbf{B}_*\mathbf{A}_*^{-1}$ and $\partial_j\mathbf{A}_*^{-1} = -\mathbf{A}_*^{-1}(\partial_j\mathbf{A}_*)\mathbf{A}_*^{-1}$.

(*iii*) If $\mathbb{H}_0$ holds, $\mathbf{B}_*\mathbf{A}_*^{-1} = \mathbf{I}$, so that $\nabla_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*] = [\mathrm{tr}[(\partial_j\mathbf{B}_* - \partial_j\mathbf{A}_*)\mathbf{A}_*^{-1}]] = [\mathrm{tr}[\mathbf{S}_{j,*}]]$. ■

**Lemma A4.** *Given Assumptions 1, 2, 3, and 4,*

*(i)* $\nabla_{\boldsymbol{\theta}}\det[\mathbf{D}_n] = \det[\mathbf{D}_n][\mathrm{tr}[\mathbf{R}_{j,n}]]$*;*

*(ii)* $\nabla_{\boldsymbol{\theta}}\det[\mathbf{D}_*] = \det[\mathbf{D}_*][\mathrm{tr}[\mathbf{R}_{j,*}]]$*;*

*(iii) if* $\mathbb{H}_0$ *further holds,* $\nabla_{\boldsymbol{\theta}} \det[\mathbf{D}_*] = [\mathrm{tr}[\mathbf{S}_{j,*}]]$*;*

*(iv)* $\nabla_{\boldsymbol{\theta}}(\det[\mathbf{D}_n]-\det[\mathbf{D}_*]) = \det[\mathbf{D}_*]\{\mathrm{tr}[\mathbf{P}_n][\mathrm{tr}[\mathbf{R}_{j,*}]]+[\mathrm{tr}[\mathbf{A}_*^{-1}(\partial_j\mathbf{A}_*)\mathbf{U}_n-\mathbf{B}_*^{-1}(\partial_j\mathbf{B}_*)\,\mathbf{W}_n]]+[\mathrm{tr}[\mathbf{J}_{j,n}]]\}+ o_{\mathbb{P}}(n^{-1/2})$*; and*

*(v) if* $\mathbb{H}_0$ *further holds,* $\nabla_{\boldsymbol{\theta}}(\det[\mathbf{D}_n]-\det[\mathbf{D}_*]) = \mathrm{tr}[\mathbf{M}_n][\mathrm{tr}[\mathbf{S}_{j,*}]]+[\mathrm{tr}[\mathbf{A}_*^{-1}\{(\partial_j\mathbf{A}_*)\mathbf{U}_n-(\partial_j\mathbf{B}_*)\mathbf{W}_n\}]]+ [\mathrm{tr}[\mathbf{J}_{j,n}]]+o_{\mathbb{P}}(n^{-1/2})$. □

**Proof of Lemma A4:** ($i$) By exercise 13.32 of Abadir and Magnus (2005): $d|\mathbf{X}| = |\mathbf{X}|\mathrm{tr}(\mathbf{X}^{-1}d\mathbf{X})$, with $d|\mathbf{X}|$ be $\partial_j \det[\mathbf{D}_n]$, and $\partial_j \det[\mathbf{D}_n] = \det[\mathbf{D}_n]\mathrm{tr}[\mathbf{D}_n^{-1}\partial_j\mathbf{D}_n]$. Also, $\partial_j\mathbf{D}_n = (\partial_j\mathbf{B}_n)\mathbf{A}_n^{-1}-\mathbf{B}_n\mathbf{A}_n^{-1}(\partial_j\mathbf{A}_n)\mathbf{A}_n^{-1}$. If we combine these, $\partial_j \det[\mathbf{D}_n] = \det[\mathbf{D}_n]\mathrm{tr}[\mathbf{R}_{j,n}]$, and the definition of the gradient yields the result.

($ii$) We repeat the proof of ($i$), replacing the subscript '$n$' with the subscript '$*$'.

($iii$) Under $\mathbb{H}_0$, $\det[\mathbf{D}_*]=1$ and $\mathbf{A}_*=\mathbf{B}_*$, so that $\partial_j\det[\mathbf{D}_*] = \mathrm{tr}[\mathbf{A}_*^{-1}\partial_j\mathbf{B}_*-\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*] = \mathrm{tr}[\mathbf{S}_{j,*}]$.

($iv$) We note the following identity:

$$\begin{aligned}\nabla_{\boldsymbol{\theta}}(\det[\mathbf{D}_n]-\det[\mathbf{D}_*]) &\equiv (\det[\mathbf{D}_n]-\det[\mathbf{D}_*])[\mathrm{tr}[\mathbf{R}_{j,n}]]\\ &+\det[\mathbf{D}_*][\,\mathrm{tr}[\mathbf{B}_n^{-1}\partial_j\mathbf{B}_n-\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*]]-\det[\mathbf{D}_*][\,\mathrm{tr}[\mathbf{A}_n^{-1}\partial_j\mathbf{A}_n-\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*]].\end{aligned} \tag{12}$$

We examine the asymptotic approximation of each component in the RHS. First, Lemma A2($iv$) implies that $\det[\mathbf{D}_n]-\det[\mathbf{D}_*] = \det[\mathbf{D}_*](\mathrm{tr}[\mathbf{P}_n])+o_{\mathbb{P}}(n^{-1/2})$. Thus,

$$(\det[\mathbf{D}_n]-\det[\mathbf{D}_*])[\mathrm{tr}[\mathbf{R}_{j,n}]] = \det[\mathbf{D}_*](\mathrm{tr}[\mathbf{P}_n])[\mathrm{tr}[\mathbf{R}_{j,*}]]+o_{\mathbb{P}}(n^{-1/2}) \tag{13}$$

by Assumptions 3 and 4. Second, we note that

$$\begin{aligned}\mathbf{B}_n^{-1}\partial_j\mathbf{B}_n-\mathbf{B}_*^{-1}\partial_j\mathbf{B}_* &\equiv (\mathbf{B}_n^{-1}-\mathbf{B}_*^{-1})\partial_j\mathbf{B}_n+\mathbf{B}_*^{-1}\partial_j(\mathbf{B}_n-\mathbf{B}_*)\\ &= -\mathbf{W}_n\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*+\mathbf{B}_*^{-1}\partial_j(\mathbf{B}_n-\mathbf{B}_*)+o_{\mathbb{P}}(n^{-1/2}),\end{aligned} \tag{14}$$

where the last equality holds by Lemma A1($ii$). Third, it follows similarly that

$$\mathbf{A}_n^{-1}\partial_j\mathbf{A}_n-\mathbf{A}_*^{-1}\partial_j\mathbf{A}_* = -\mathbf{U}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*+\mathbf{A}_*^{-1}\partial_j(\mathbf{A}_n-\mathbf{A}_*)+o_{\mathbb{P}}(n^{-1/2}). \tag{15}$$

Finally, we plug Eqs. (13), (14), and (15) into Eq. (12) and obtain

$$\begin{aligned}\nabla_{\boldsymbol{\theta}}(\det[\mathbf{D}_n]-\det[\mathbf{D}_*]) &= \det[\mathbf{D}_*]\{\mathrm{tr}[\mathbf{P}_n][\mathrm{tr}[\mathbf{R}_{j,*}]]+[\mathrm{tr}[\mathbf{A}_*^{-1}(\partial_j\mathbf{A}_*)\mathbf{U}_n-\mathbf{B}_*^{-1}(\partial_j\mathbf{B}_*)\mathbf{W}_n]]\}\\ &+\det[\mathbf{D}_*][\,\mathrm{tr}[\mathbf{B}_*^{-1}\partial_j(\mathbf{B}_n-\mathbf{B}_*)-\mathbf{A}_*^{-1}\partial_j(\mathbf{A}_n-\mathbf{A}_*)]]+o_{\mathbb{P}}(n^{-1/2}).\end{aligned} \tag{16}$$

Here, we note that $\mathbf{B}_*^{-1}\partial_j(\mathbf{B}_n-\mathbf{B}_*)-\mathbf{A}_*^{-1}\partial_j(\mathbf{A}_n-\mathbf{A}_*)=\mathbf{G}_{j,n}-\mathbf{H}_{j,n}=\mathbf{J}_{j,n}$ by the definitions of $\mathbf{G}_{j,n}$, $\mathbf{H}_{j,n}$ and $\mathbf{J}_{j,n}$. This yields the desired result.

($v$) Under $\mathbb{H}_0$, $\det[\mathbf{D}_*]=1$ and $\mathbf{P}_n=\mathbf{M}_n$. Also, $\mathrm{tr}[\mathbf{W}_n\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*-\mathbf{U}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*]=\mathrm{tr}[\mathbf{A}_*^{-1}(\partial_j\mathbf{B}_*)\,\mathbf{W}_n-\mathbf{A}_*^{-1}(\partial_j\mathbf{A}_*)\mathbf{U}_n]$ and $\mathrm{tr}[\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*-\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*]=\mathrm{tr}[\mathbf{A}_*^{-1}(\partial_j\mathbf{B}_*-\partial_j\mathbf{A}_*)]=\mathrm{tr}[\mathbf{S}_{j,*}]$. ■

**Lemma A5.** *Given Assumptions 1 and 2,*

*(i)* $\partial^2_{ji}\mathrm{tr}[\mathbf{D}_*]=\mathrm{tr}[\mathbf{A}_*^{-1}\mathbf{B}_*\{(\mathbf{B}_*^{-1}\partial^2_{ji}\mathbf{B}_*-\mathbf{A}_*^{-1}\partial^2_{ji}\mathbf{A}_*)-(\mathbf{R}_{j,*}\mathbf{A}_*^{-1}\partial_i\mathbf{A}_*+\mathbf{R}_{i,*}\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)\}]$*; and*

*(ii) if* $\mathbb{H}_0$ *further holds,* $\partial^2_{ji}\mathrm{tr}[\mathbf{D}_*]=\mathrm{tr}[[\mathbf{A}_*^{-1}(\partial^2_{ji}\mathbf{B}_*-\partial^2_{ji}\mathbf{A}_*)-(\mathbf{S}_{j,*}\mathbf{A}_*^{-1}\partial_i\mathbf{A}_*+\mathbf{S}_{i,*}\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)]]$. □

**Proof of Lemma A5:** ($i$) Some tedious algebra shows that

$$\begin{aligned}\partial^2_{ji}\mathbf{D}_* &= (\partial^2_{ji}\mathbf{B}_*)\mathbf{A}_*^{-1}-(\partial_i\mathbf{B}_*)\mathbf{A}_*^{-1}(\partial_j\mathbf{A}_*)\mathbf{A}_*^{-1}-(\partial_j\mathbf{B}_*)\mathbf{A}_*^{-1}(\partial_i\mathbf{A}_*)\mathbf{A}_*^{-1}\\ &\quad+\mathbf{B}_*\mathbf{A}_*^{-1}(\partial_j\mathbf{A}_*)\mathbf{A}_*^{-1}(\partial_i\mathbf{A}_*)\mathbf{A}_*^{-1}-\mathbf{B}_*\mathbf{A}_*^{-1}(\partial^2_{ij}\mathbf{A}_*)\mathbf{A}_*^{-1}+\mathbf{B}_*\mathbf{A}_*^{-1}(\partial_i\mathbf{A}_*)\mathbf{A}_*^{-1}(\partial_j\mathbf{A}_*)\mathbf{A}_*^{-1},\end{aligned}$$

so that

$$\begin{aligned}\partial^2_{ji}\mathrm{tr}[\mathbf{D}_*] &= \mathrm{tr}[\mathbf{B}_*\mathbf{A}_*^{-1}[(\partial_j\mathbf{A}_*)\mathbf{A}_*^{-1}(\partial_i\mathbf{A}_*)\mathbf{A}_*^{-1}+(\partial_i\mathbf{A}_*)\mathbf{A}_*^{-1}(\partial_j\mathbf{A}_*)\mathbf{A}_*^{-1}]]\\ &\quad-\mathrm{tr}[(\partial_j\mathbf{A}_*)\mathbf{A}_*^{-1}(\partial_i\mathbf{B}_*)\mathbf{A}_*^{-1}+(\partial_j\mathbf{B}_*)\mathbf{A}_*^{-1}(\partial_i\mathbf{A}_*)\mathbf{A}_*^{-1}]+\mathrm{tr}[[(\partial^2_{ji}\mathbf{B}_*)-(\partial^2_{ij}\mathbf{A}_*)\mathbf{A}_*^{-1}\mathbf{B}_*]\mathbf{A}_*^{-1}].\end{aligned}$$

We now rearrange this using the definitions of $\mathbf{R}_{j,*}$ and $\mathbf{R}_{i,*}$, yielding the desired result.

($ii$) We note that $\mathbf{R}_{j,*}=\mathbf{S}_{j,*}$ and $\mathbf{R}_{i,*}=\mathbf{S}_{i,*}$ under $\mathbb{H}_0$. This completes the proof. ■

**Lemma A6.** *Given Assumptions 1 and 2,*

*(i)* $\partial^2_{ji}\det[\mathbf{D}_*]=\det[\mathbf{D}_*]\{\,\mathrm{tr}[\mathbf{R}_{i,*}]\mathrm{tr}[\mathbf{R}_{j,*}]+\mathrm{tr}[\mathbf{B}_*^{-1}\partial^2_{ji}\mathbf{B}_*-\mathbf{A}_*^{-1}\partial^2_{ji}\mathbf{A}_*-\mathbf{R}_{i,*}\mathbf{B}_*^{-1}(\partial_j\mathbf{B}_*)-\mathbf{R}_{j,*}\mathbf{A}_*^{-1}\,\partial_i\mathbf{A}_*]\}$*; and*

*(ii) if* $\mathbb{H}_0$ *also holds,*

$$\partial^2_{ji}\det[\mathbf{D}_*]=\mathrm{tr}[\mathbf{S}_{i,*}]\mathrm{tr}[\mathbf{S}_{j,*}]+\mathrm{tr}[\mathbf{A}_*^{-1}(\partial^2_{ji}\mathbf{B}_*-\partial^2_{ji}\mathbf{A}_*)-\mathbf{S}_{i,*}\mathbf{B}_*^{-1}(\partial_j\mathbf{B}_*)-\mathbf{S}_{j,*}\mathbf{A}_*^{-1}\partial_i\mathbf{A}_*].$$ □

**Proof of Lemma A6:** ($i$) From the proof of Lemma A4($ii$), $\partial_i\det[\mathbf{D}_*]=\det[\mathbf{D}_*]\mathrm{tr}[\mathbf{R}_{i,*}]$. Therefore,

$$\partial^2_{ji}\det[\mathbf{D}_*]=\mathrm{tr}[\mathbf{R}_{i,*}]\partial_j\det[\mathbf{D}_*]+\det[\mathbf{D}_*]\mathrm{tr}[\partial_j\mathbf{B}_*^{-1}\partial_i\mathbf{B}_*+\mathbf{B}_*^{-1}\partial^2_{ji}\mathbf{B}_*-\partial_j\mathbf{A}_*^{-1}\partial_i\mathbf{A}_*-\mathbf{A}_*^{-1}\partial^2_{ji}\mathbf{A}_*].$$

Now replace $\partial_j\det[\mathbf{D}_*]$, $\partial_j(\mathbf{B}_*^{-1})$, and $\partial_j(\mathbf{A}_*^{-1})$ with $\det[\mathbf{D}_*]\mathrm{tr}[\mathbf{R}_{j,*}]$, $-\mathbf{B}_*^{-1}(\partial_j\mathbf{B}_*)\mathbf{B}_*^{-1}$, and $-\mathbf{A}_*^{-1}(\partial_j\mathbf{A}_*)\mathbf{A}_*^{-1}$, respectively. Using the definitions of $\mathbf{R}_{j,*}$ and $\mathbf{R}_{i,*}$ and rearranging yields the desired result.

($ii$) Under $\mathbb{H}_0$, $\mathbf{A}_*=\mathbf{B}_*$ and $\mathbf{B}_*\mathbf{A}_*^{-1}=\mathbf{I}$. Further, $\mathbf{R}_{j,*}=\mathbf{S}_{j,*}$ and $\mathbf{R}_{i,*}=\mathbf{S}_{i,*}$. ■

**Proof of Lemma 2:** Lemma 3 implies Lemma 2. ■

The following supplementary lemma is useful in obtaining the second-order expansion of $\widehat{S}_n := \mathrm{tr}[\widehat{\mathbf{D}}_n]/k - \det[\widehat{\mathbf{D}}_n]^{1/k}$.

**Lemma A7.** *Given Assumptions 1, 2, and* $\mathbb{H}_0$, $\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*] - \nabla^2_{\boldsymbol{\theta}}\det[\mathbf{D}_*] = [\mathrm{tr}[\mathbf{S}_{j,*}\mathbf{S}_{i,*}] - \mathrm{tr}[\mathbf{S}_{j,*}]\mathrm{tr}[\mathbf{S}_{i,*}]]$. □

**Proof of Lemma A7:** We apply Lemmas A5(*ii*) and A6(*ii*) and obtain

$$\partial^2_{ji}(\mathrm{tr}[\mathbf{D}_*] - \det[\mathbf{D}_*]) = \mathrm{tr}[\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*\mathbf{A}_*^{-1}\partial_i\mathbf{A}_*] - \mathrm{tr}[\mathbf{A}_*^{-1}\partial_j\mathbf{B}_*\mathbf{A}_*^{-1}\partial_i\mathbf{A}_*]$$
$$- \mathrm{tr}[\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*\mathbf{A}_*^{-1}\partial_i\mathbf{B}_*] + \mathrm{tr}[\mathbf{A}_*^{-1}\partial_j\mathbf{B}_*\mathbf{A}_*^{-1}\partial_i\mathbf{B}_*] - \mathrm{tr}[\mathbf{S}_{j,*}]\mathrm{tr}[\mathbf{S}_{i,*}] = \mathrm{tr}[\mathbf{S}_{j,*}\mathbf{S}_{i,*}] - \mathrm{tr}[\mathbf{S}_{j,*}]\mathrm{tr}[\mathbf{S}_{i,*}].$$

Therefore, $\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*] - \nabla^2_{\boldsymbol{\theta}}\det[\mathbf{D}_*] = [\mathrm{tr}[\mathbf{S}_{j,*}\mathbf{S}_{i,*}] - \mathrm{tr}[\mathbf{S}_{j,*}]\mathrm{tr}[\mathbf{S}_{i,*}]]$. This completes the proof. ■

**Proof of Lemma 3:** (*i*) Note that $\mathrm{tr}[\widehat{\mathbf{D}}_n - \mathbf{D}_*] = \mathrm{tr}[\widehat{\mathbf{D}}_n - \mathbf{D}_n] + \mathrm{tr}[\mathbf{D}_n - \mathbf{D}_*]$ and Lemma A1(*iii*) gives

$$\mathrm{tr}[\mathbf{D}_n] - \mathrm{tr}[\mathbf{D}_*] = \mathrm{tr}[\mathbf{P}_n\mathbf{A}_*^{-1}\mathbf{B}_*] - \mathrm{tr}[\mathbf{P}_n\mathbf{U}_n\mathbf{A}_*^{-1}\mathbf{B}_*] + o_{\mathbb{P}}(n^{-1}). \tag{17}$$

We next consider the approximation of $(\widehat{\mathbf{D}}_n - \mathbf{D}_n)$. Applying a Taylor expansion gives

$$\mathrm{tr}[\widehat{\mathbf{D}}_n - \mathbf{D}_n] = \nabla'_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_n]\widetilde{\boldsymbol{\theta}}_n + \frac{1}{2}\widetilde{\boldsymbol{\theta}}'_n\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*]\widetilde{\boldsymbol{\theta}}_n + o_{\mathbb{P}}(n^{-1}), \tag{18}$$

where $\widetilde{\boldsymbol{\theta}}_n := \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*$. We also note that

$$\nabla'_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_n]\widetilde{\boldsymbol{\theta}}_n = \nabla'_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*]\widetilde{\boldsymbol{\theta}}_n + \nabla'_{\boldsymbol{\theta}}(\mathrm{tr}[\mathbf{D}_n] - \mathrm{tr}[\mathbf{D}_*])\widetilde{\boldsymbol{\theta}}_n + o_{\mathbb{P}}(n^{-1}) \quad \text{and} \tag{19}$$

$$\nabla_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*] = [\mathrm{tr}[\mathbf{R}_{j,*}\mathbf{A}_*^{-1}\mathbf{B}_*]] \tag{20}$$

by Lemma A3(*ii*). Further, applying Lemma A3(*i* and *ii*) implies that

$$\begin{aligned}\nabla_{\boldsymbol{\theta}}(\mathrm{tr}[\mathbf{D}_n] - \mathrm{tr}[\mathbf{D}_*]) &= [\mathrm{tr}[(\partial_j\mathbf{B}_n)\mathbf{A}_n^{-1} - (\partial_j\mathbf{B}_*)\mathbf{A}_*^{-1}]] \\ &\quad -[\mathrm{tr}[\mathbf{B}_n\mathbf{A}_n^{-1}(\partial_j\mathbf{A}_n)\mathbf{A}_n^{-1} - \mathbf{B}_*\mathbf{A}_*^{-1}(\partial_j\mathbf{A}_*)\mathbf{A}_*^{-1}]].\end{aligned} \tag{21}$$

We examine each element in Eq. (21) separately. The first term in Eq. (21) is

$$
\begin{aligned}
&(\partial_j \mathbf{B}_n)\mathbf{A}_n^{-1} - (\partial_j \mathbf{B}_*)\mathbf{A}_*^{-1} \\
&\qquad \equiv \partial_j(\mathbf{B}_n - \mathbf{B}_*)\mathbf{A}_*^{-1} + \partial_j(\mathbf{B}_n - \mathbf{B}_*)(\mathbf{A}_n^{-1} - \mathbf{A}_*^{-1}) + (\partial_j \mathbf{B}_*)(\mathbf{A}_n^{-1} - \mathbf{A}_*^{-1}).
\end{aligned}
$$

We now apply Lemma A1($i$) and obtain

$$
(\partial_j \mathbf{B}_n)\mathbf{A}_n^{-1} - (\partial_j \mathbf{B}_*)\mathbf{A}_*^{-1} = \partial_j(\mathbf{B}_n - \mathbf{B}_*)\mathbf{A}_*^{-1} - (\partial_j \mathbf{B}_*)\mathbf{U}_n\mathbf{A}_*^{-1} + o_{\mathbb{P}}(n^{-1/2}). \tag{22}
$$

For the second term in Eq. (21), note that

$$
\begin{aligned}
\mathbf{B}_n\mathbf{A}_n^{-1}(\partial_j \mathbf{A}_n)\mathbf{A}_n^{-1} - \mathbf{B}_*\mathbf{A}_*^{-1}(\partial_j \mathbf{A}_*)\mathbf{A}_*^{-1} &\equiv (\mathbf{B}_n - \mathbf{B}_*)\mathbf{A}_n^{-1}(\partial_j \mathbf{A}_n)\mathbf{A}_n^{-1} \\
&+ \mathbf{B}_*(\mathbf{A}_n^{-1} - \mathbf{A}_*^{-1})(\partial_j \mathbf{A}_n)\mathbf{A}_n^{-1} + \mathbf{B}_*\mathbf{A}_*^{-1}\partial_j(\mathbf{A}_n - \mathbf{A}_*)\mathbf{A}_n^{-1} \\
&+ \mathbf{B}_*\mathbf{A}_*^{-1}(\partial_j \mathbf{A}_*)(\mathbf{A}_n^{-1} - \mathbf{A}_*^{-1}),
\end{aligned}
$$

and apply Lemma A3($i$) and the definitions of $\mathbf{W}_n$ and $\mathbf{H}_{j,n}$ to obtain

$$
\begin{aligned}
&\mathbf{B}_n\mathbf{A}_n^{-1}(\partial_j \mathbf{A}_n)\mathbf{A}_n^{-1} - \mathbf{B}_*\mathbf{A}_*^{-1}(\partial_j \mathbf{A}_*)\mathbf{A}_*^{-1} = \mathbf{B}_*\mathbf{W}_n\mathbf{A}_*^{-1}(\partial_j \mathbf{A}_*)\mathbf{A}_*^{-1} \\
&\quad - \mathbf{B}_*\mathbf{U}_n\mathbf{A}_*^{-1}(\partial_j \mathbf{A}_*)\mathbf{A}_*^{-1} + \mathbf{B}_*\mathbf{H}_{j,n}\mathbf{A}_*^{-1} - \mathbf{B}_*\mathbf{A}_*^{-1}(\partial_j \mathbf{A}_*)\mathbf{U}_n\mathbf{A}_*^{-1} + o_{\mathbb{P}}(n^{-1/2}).
\end{aligned} \tag{23}
$$

We now subtract Eq. (23) from Eq. (22) to obtain the approximation of the RHS in Eq. (21) as follows:

$$
\nabla_{\boldsymbol{\theta}}(\mathrm{tr}[\mathbf{D}_n] - \mathrm{tr}[\mathbf{D}_*]) = [\mathrm{tr}[\{\mathbf{J}_{j,n} - \mathbf{P}_n\mathbf{A}_*^{-1}\partial_j \mathbf{A}_* - \mathbf{R}_{j,*}\mathbf{U}_n\}\mathbf{A}_*^{-1}\mathbf{B}_*]] + o_{\mathbb{P}}(n^{-1/2}), \tag{24}
$$

where letting $\mathbf{J}_{j,n} := \mathbf{G}_{j,n} - \mathbf{H}_{j,n}$, $\mathbf{G}_{j,n} := \mathbf{B}_*^{-1}\partial_j(\mathbf{B}_n - \mathbf{B}_*)$, $\mathbf{P}_n := \mathbf{W}_n - \mathbf{U}_n$, and $\mathbf{R}_{j,*} := \mathbf{B}_*^{-1}\partial_j \mathbf{B}_* - \mathbf{A}_*^{-1}\partial_j \mathbf{A}_*$. We now plug Eqs. (20) and (24) into Eq. (19) and obtain

$$
\nabla'_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_n]\widetilde{\boldsymbol{\theta}}_n = [\mathrm{tr}[\mathbf{R}_{j,*}\mathbf{A}_*^{-1}\mathbf{B}_* + \{\mathbf{J}_{j,n} - \mathbf{P}_n\mathbf{A}_*^{-1}\partial_j \mathbf{A}_* - \mathbf{R}_{j,*}\mathbf{U}_n\}\mathbf{A}_*^{-1}\mathbf{B}_*]]'\widetilde{\boldsymbol{\theta}}_n + o_{\mathbb{P}}(n^{-1}).
$$

Plugging this into Eq. (18) yields

$$
\begin{aligned}
\mathrm{tr}[\widehat{\mathbf{D}}_n - \mathbf{D}_n] =& [\mathrm{tr}[\mathbf{R}_{j,*}\mathbf{A}_*^{-1}\mathbf{B}_* + \{\mathbf{J}_{j,n} - \mathbf{P}_n\mathbf{A}_*^{-1}\partial_j \mathbf{A}_* - \mathbf{R}_{j,*}\mathbf{U}_n\}\mathbf{A}_*^{-1}\mathbf{B}_*]]'\widetilde{\boldsymbol{\theta}}_n \\
&+ \frac{1}{2}\widetilde{\boldsymbol{\theta}}'_n\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*]\widetilde{\boldsymbol{\theta}}_n + o_{\mathbb{P}}(n^{-1}).
\end{aligned} \tag{25}
$$

We now combine Eqs. (17) and (25) and obtain

$$
\begin{aligned}
\mathrm{tr}[\widehat{\mathbf{D}}_n]-\mathrm{tr}[\mathbf{D}_*] =& \mathrm{tr}[\mathbf{L}_n\mathbf{A}_*^{-1}\mathbf{B}_*]-\mathrm{tr}[\mathbf{L}_n\mathbf{U}_n\mathbf{A}_*^{-1}\mathbf{B}_*]+[\mathrm{tr}[(\mathbf{J}_{j,n}-\mathbf{P}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)\mathbf{A}_*^{-1}\mathbf{B}_*]]'\widetilde{\boldsymbol{\theta}}_n \\
&+\frac{1}{2}\widetilde{\boldsymbol{\theta}}_n'\nabla_{\boldsymbol{\theta}}^2\mathrm{tr}[\mathbf{D}_*]\widetilde{\boldsymbol{\theta}}_n+o_{\mathbb{P}}(n^{-1}),
\end{aligned}
\tag{26}
$$

where we let $\mathbf{L}_n := \mathbf{P}_n+\sum_{j=1}^{\ell}\widetilde{\theta}_{j,n}\mathbf{R}_{j,*}$. Finally, imposing $\mathbb{H}_0$ implies that $\mathbf{L}_n=\mathbf{M}_n+\sum_{j=1}^{\ell}\widetilde{\theta}_{j,n}\mathbf{S}_{j,*}$, $\mathbf{A}_*^{-1}\mathbf{B}_*=\mathbf{I}$, and $\mathbf{P}_n=\mathbf{M}_n$, so that the desired result follows.

(*ii*) We note that

$$
\det[\widehat{\mathbf{D}}_n]^{\frac{1}{k}}-\det[\mathbf{D}_*]^{\frac{1}{k}} \equiv (\det[\widehat{\mathbf{D}}_n]^{\frac{1}{k}}-\det[\mathbf{D}_n]^{\frac{1}{k}})+(\det[\mathbf{D}_n]^{\frac{1}{k}}-\det[\mathbf{D}_*]^{\frac{1}{k}}).
\tag{27}
$$

We separately examine the approximation for each term of the RHS and then combine them under the null.

First, we consider $\det[\mathbf{D}_n]^{\frac{1}{k}}-\det[\mathbf{D}_*]^{\frac{1}{k}}$. We note that

$$
\begin{aligned}
\det[\mathbf{D}_n]^{\frac{1}{k}}-\det[\mathbf{D}_*]^{\frac{1}{k}} =& \frac{1}{k}\det[\mathbf{D}_*]^{\frac{1}{k}-1}(\det[\mathbf{D}_n]-\det[\mathbf{D}_*]) \\
&+\frac{1}{2k}\left(\frac{1}{k}-1\right)\det[\mathbf{D}_*]^{\frac{1}{k}-2}(\det[\mathbf{D}_n]-\det[\mathbf{D}_*])^2+o_{\mathbb{P}}(n^{-1})
\end{aligned}
$$

by a second-order Taylor expansion. We apply Lemma A2(*iv*) to obtain

$$
\det[\mathbf{D}_n]^{\frac{1}{k}}-\det[\mathbf{D}_*]^{\frac{1}{k}}=\det[\mathbf{D}_*]^{\frac{1}{k}}\{\frac{1}{k}\mathrm{tr}[\mathbf{P}_n]+\frac{1}{2k^2}\mathrm{tr}[\mathbf{P}_n]^2-\frac{1}{2k}(\mathrm{tr}[\mathbf{W}_n^2]-\mathrm{tr}[\mathbf{U}_n^2])\}+o_{\mathbb{P}}(n^{-1}).
\tag{28}
$$

Next, we approximate $\det[\widehat{\mathbf{D}}_n]^{1/k}-\det[\mathbf{D}_n]^{1/k}$. A second-order Taylor expansion gives

$$
\begin{aligned}
\det[\widehat{\mathbf{D}}_n]^{\frac{1}{k}}-\det[\mathbf{D}_n]^{\frac{1}{k}} =& \frac{1}{k}\det[\mathbf{D}_n]^{\frac{1}{k}-1}\nabla_{\boldsymbol{\theta}}'\det[\mathbf{D}_n]\widetilde{\boldsymbol{\theta}}_n+\frac{1}{2k}\det[\mathbf{D}_n]^{\frac{1}{k}-1}\widetilde{\boldsymbol{\theta}}_n'\nabla_{\boldsymbol{\theta}}^2\det[\mathbf{D}_n]\widetilde{\boldsymbol{\theta}}_n \\
&+\frac{1}{2k}\left(\frac{1}{k}-1\right)\det[\mathbf{D}_n]^{\frac{1}{k}-2}\{\nabla_{\boldsymbol{\theta}}'\det[\mathbf{D}_n]\widetilde{\boldsymbol{\theta}}_n\}^2+o_{\mathbb{P}}(n^{-1}).
\end{aligned}
\tag{29}
$$

Here, each term on the RHS can be approximated as follows: we have

$$
\det[\mathbf{D}_n]^{\frac{1}{k}-2}\{\widetilde{\boldsymbol{\theta}}_n'\nabla_{\boldsymbol{\theta}}\det[\mathbf{D}_n]\}^2=\det[\mathbf{D}_*]^{\frac{1}{k}}\{\widetilde{\boldsymbol{\theta}}_n'[\mathrm{tr}[\mathbf{R}_{j,*}]]\}^2+o_{\mathbb{P}}(n^{-1})
\tag{30}
$$

by Assumption 4 and Lemma A4(*i*,*ii*), and

$$
\det[\mathbf{D}_n]^{\frac{1}{k}-1}\widetilde{\boldsymbol{\theta}}_n'\nabla_{\boldsymbol{\theta}}^2\det[\mathbf{D}_n]\widetilde{\boldsymbol{\theta}}_n=\det[\mathbf{D}_*]^{\frac{1}{k}-1}\widetilde{\boldsymbol{\theta}}_n'\nabla_{\boldsymbol{\theta}}^2\det[\mathbf{D}_*]\widetilde{\boldsymbol{\theta}}_n+o_{\mathbb{P}}(n^{-1}).
\tag{31}
$$

Here, we have used Assumption 3. Also, we can decompose the main part of the first term on the RHS as

$$\det[\mathbf{D}_n]^{\frac{1}{k}-1}\nabla'_{\boldsymbol{\theta}}\det[\mathbf{D}_n]\widetilde{\boldsymbol{\theta}}_n \equiv \det[\mathbf{D}_*]^{\frac{1}{k}-1}\nabla'_{\boldsymbol{\theta}}\det[\mathbf{D}_*]\widetilde{\boldsymbol{\theta}}_n + (\det[\mathbf{D}_n]^{\frac{1}{k}-1} - \det[\mathbf{D}_*]^{\frac{1}{k}-1})\nabla'_{\boldsymbol{\theta}}\det[\mathbf{D}_*]\widetilde{\boldsymbol{\theta}}_n + \det[\mathbf{D}_n]^{\frac{1}{k}-1}(\nabla'_{\boldsymbol{\theta}}\det[\mathbf{D}_n] - \nabla'_{\boldsymbol{\theta}}\det[\mathbf{D}_*])\widetilde{\boldsymbol{\theta}}_n. \tag{32}$$

We examine the approximation of each term on the RHS of Eq. (32): ($a$) the first term is

$$\det[\mathbf{D}_*]^{\frac{1}{k}-1}\nabla'_{\boldsymbol{\theta}}\det[\mathbf{D}_*]\widetilde{\boldsymbol{\theta}}_n = \det[\mathbf{D}_*]^{\frac{1}{k}}[\mathrm{tr}[\mathbf{R}_{j,*}]]'\widetilde{\boldsymbol{\theta}}_n \tag{33}$$

by Lemma A4($ii$); ($b$) the second term on the RHS of Eq. (32) can be written

$$\begin{aligned}(\det[\mathbf{D}_n]^{\frac{1}{k}-1} &- \det[\mathbf{D}_*]^{\frac{1}{k}-1})\nabla'_{\boldsymbol{\theta}}\det[\mathbf{D}_*]\widetilde{\boldsymbol{\theta}}_n \\ &= \left(\frac{1}{k}-1\right)\det[\mathbf{D}_*]^{\frac{1}{k}-2}(\det[\mathbf{D}_n] - \det[\mathbf{D}_*])\nabla'_{\boldsymbol{\theta}}\det[\mathbf{D}_*]\widetilde{\boldsymbol{\theta}}_n + o_{\mathbb{P}}(n^{-1}) \\ &= \left(\frac{1}{k}-1\right)\det[\mathbf{D}_*]^{\frac{1}{k}}\mathrm{tr}[\mathbf{P}_n][\mathrm{tr}[\mathbf{R}_{j,*}]]'\widetilde{\boldsymbol{\theta}}_n + o_{\mathbb{P}}(n^{-1})\end{aligned} \tag{34}$$

by Lemmas A2($iv$) and A4($ii$); ($c$) the final term on the RHS of Eq. (32) can be written as

$$\begin{aligned}\det[\mathbf{D}_n]^{\frac{1}{k}-1}(\nabla'_{\boldsymbol{\theta}}\det[\mathbf{D}_n] - \nabla'_{\boldsymbol{\theta}}\det[\mathbf{D}_*])\widetilde{\boldsymbol{\theta}}_n &= \det[\mathbf{D}_*]^{\frac{1}{k}}\mathrm{tr}[\mathbf{P}_n][\mathrm{tr}[\mathbf{R}_{j,*}]]'\widetilde{\boldsymbol{\theta}}_n \\ &+ \det[\mathbf{D}_*]^{\frac{1}{k}}[\mathrm{tr}[\mathbf{J}_{j,n} + \mathbf{U}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_* - \mathbf{W}_n\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*]]'\widetilde{\boldsymbol{\theta}}_n + o_{\mathbb{P}}(n^{-1})\end{aligned} \tag{35}$$

by Lemma A4($iv$); ($d$) we now plug Eqs. (33), (34), and (35) into Eq. (32) and obtain

$$\begin{aligned}\det[\mathbf{D}_n]^{\frac{1}{k}-1}\nabla'_{\boldsymbol{\theta}}\det[\mathbf{D}_n]\widetilde{\boldsymbol{\theta}}_n &= \det[\mathbf{D}_*]^{\frac{1}{k}}\{1 + k^{-1}\mathrm{tr}[\mathbf{P}_n]\}[\mathrm{tr}[\mathbf{R}_{j,*}]]'\widetilde{\boldsymbol{\theta}}_n \\ &+ \det[\mathbf{D}_*]^{\frac{1}{k}}[\mathrm{tr}[\mathbf{J}_{j,n} + \mathbf{U}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_* - \mathbf{W}_n\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*]]'\widetilde{\boldsymbol{\theta}}_n + o_{\mathbb{P}}(n^{-1}).\end{aligned} \tag{36}$$

Plugging Eqs. (30), (31), and (36) into Eq. (29) gives

$$\begin{aligned}\det[\widehat{\mathbf{D}}_n]^{\frac{1}{k}} - \det[\mathbf{D}_n]^{\frac{1}{k}} &= \frac{1}{k}\det[\mathbf{D}_*]^{\frac{1}{k}}\left\{1 + \frac{1}{k}\mathrm{tr}[\mathbf{P}_n]\right\}[\mathrm{tr}[\mathbf{R}_{j,*}]]'\widetilde{\boldsymbol{\theta}}_n \\ &+ \frac{1}{k}\det[\mathbf{D}_*]^{\frac{1}{k}}[\mathrm{tr}[\mathbf{J}_{j,n} + \mathbf{U}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_* - \mathbf{W}_n\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*]]'\widetilde{\boldsymbol{\theta}}_n \\ &+ \frac{1}{2k}\left(\frac{1}{k}-1\right)\det[\mathbf{D}_*]^{\frac{1}{k}}\{[\mathrm{tr}[\mathbf{R}_{j,*}]]'\widetilde{\boldsymbol{\theta}}_n\}^2 + \frac{1}{2k}\det[\mathbf{D}_*]^{\frac{1}{k}-1}\widetilde{\boldsymbol{\theta}}'_n\nabla^2_{\boldsymbol{\theta}}\det[\mathbf{D}_*]\widetilde{\boldsymbol{\theta}}_n + o_{\mathbb{P}}(n^{-1}).\end{aligned} \tag{37}$$

Summing Eqs. (28) and (37) then gives

$$\det[\mathbf{D}_*]^{-\frac{1}{k}}\left\{\det[\widehat{\mathbf{D}}_n]^{1/k}-\det[\mathbf{D}_*]^{1/k}\right\}=\frac{1}{k}\mathrm{tr}[\mathbf{L}_n]+\frac{1}{2k}\left(\frac{1}{k}-1\right)\mathrm{tr}[\mathbf{L}_n]^2$$
$$+\frac{1}{k}\mathrm{tr}[\mathbf{P}_n]\mathrm{tr}[\sum_{j=1}^{\ell}\widetilde{\theta}_{j,n}\mathbf{R}_{j,*}]+\frac{1}{2k}(\mathrm{tr}[\mathbf{P}_n]^2+\mathrm{tr}[\mathbf{U}_n^2]-\mathrm{tr}[\mathbf{W}_n^2])$$
$$+\frac{1}{k}[\mathrm{tr}[\mathbf{J}_{j,n}+\mathbf{U}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*-\mathbf{W}_n\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*]]'\widetilde{\boldsymbol{\theta}}_n+\frac{1}{2k}\det[\mathbf{D}_*]^{-1}\widetilde{\boldsymbol{\theta}}_n'\nabla_{\boldsymbol{\theta}}^2\det[\mathbf{D}_*]\widetilde{\boldsymbol{\theta}}_n+o(n^{-1}). \tag{38}$$

Here, we let $\mathbf{L}_n:=\mathbf{P}_n+\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n}-\theta_{j,*})\mathbf{R}_{j,*}$. This reduces to the desired result by noting that $\det[\mathbf{D}_*]=1$, $\mathbf{L}_n=\mathbf{K}_n$, $\mathbf{P}_n=\mathbf{M}_n$, and $\mathbf{R}_{j,*}=\mathbf{S}_{j,*}$ under $\mathbb{H}_0$.

(*iii*) By the definitions of $\widehat{T}_n^{\star}$ and $\widehat{D}_n^{\star}$, it also follows that $\widehat{S}_n^{\star}=\widehat{T}_n^{\star}-\widehat{D}_n^{\star}$, so that

$$\widehat{S}_n^{\star}=-\frac{1}{2k}\left(\frac{1}{k}-1\right)\mathrm{tr}[\mathbf{K}_n]^2-\frac{1}{k}\mathrm{tr}[\mathbf{K}_n\mathbf{U}_n]-\frac{1}{2k}(\mathrm{tr}[\mathbf{M}_n]^2+\mathrm{tr}[\mathbf{U}_n^2]-\mathrm{tr}[\mathbf{W}_n^2])$$
$$-\frac{1}{k}[\mathrm{tr}[\mathbf{M}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*+\mathbf{U}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*-\mathbf{W}_n\mathbf{A}_*^{-1}\partial_j\mathbf{B}_*]]'\widetilde{\boldsymbol{\theta}}_n-\frac{1}{k}[\mathrm{tr}[\mathbf{M}_n]\mathrm{tr}[\mathbf{S}_{j,*}]]'\widetilde{\boldsymbol{\theta}}_n$$
$$+\frac{1}{2k}\widetilde{\boldsymbol{\theta}}_n'(\nabla_{\boldsymbol{\theta}}^2\mathrm{tr}[\mathbf{D}_*]-\nabla_{\boldsymbol{\theta}}^2\det[\mathbf{D}_*])\widetilde{\boldsymbol{\theta}}_n.$$

Here, Lemma A7 implies that $(\nabla_{\boldsymbol{\theta}}^2\mathrm{tr}[\mathbf{D}_*]-\nabla_{\boldsymbol{\theta}}^2\det[\mathbf{D}_*])=[\mathrm{tr}[\mathbf{S}_{j,*}\mathbf{S}_{i,*}]-\mathrm{tr}[\mathbf{S}_{j,*}]\mathrm{tr}[\mathbf{S}_{i,*}]]$. Therefore,

$$\frac{1}{2k}\widetilde{\boldsymbol{\theta}}_n'(\nabla_{\boldsymbol{\theta}}^2\mathrm{tr}[\mathbf{D}_*]-\nabla_{\boldsymbol{\theta}}^2\det[\mathbf{D}_*])\widetilde{\boldsymbol{\theta}}_n=\frac{1}{2k}\mathrm{tr}[(\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n}-\theta_{j,*})\mathbf{S}_{j,*})^2]-\frac{1}{2k}\mathrm{tr}[\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n}-\theta_{j,*})\mathbf{S}_{j,*}]^2. \tag{39}$$

Next, we note that $\mathbf{M}_n=\mathbf{W}_n-\mathbf{U}_n$ and $\mathbf{K}_n=\mathbf{M}_n+\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n}-\theta_{j,*})\mathbf{S}_{j,*}$ under $\mathbb{H}_0$, so that

$$-\frac{1}{2k}(\mathrm{tr}[\mathbf{U}_n^2]-\mathrm{tr}[\mathbf{W}_n^2])=\frac{1}{2k}\mathrm{tr}[\mathbf{M}_n^2]+\frac{1}{k}\mathrm{tr}[\mathbf{M}_n\mathbf{U}_n]\ \text{ and} \tag{40}$$

$$-\frac{1}{k}\mathrm{tr}[\mathbf{K}_n\mathbf{U}_n]=-\frac{1}{k}\mathrm{tr}[\mathbf{M}_n\mathbf{U}_n]-\frac{1}{k}\mathrm{tr}[\mathbf{U}_n\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n}-\theta_{j,*})\mathbf{S}_{j,*}]. \tag{41}$$

Finally, we note that under $\mathbb{H}_0$,

$$\frac{1}{k}\mathrm{tr}[\mathbf{M}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*+\mathbf{U}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*-\mathbf{W}_n\mathbf{A}_*^{-1}\partial_j\mathbf{B}_*]=-\frac{1}{k}\mathrm{tr}[\mathbf{M}_n\mathbf{S}_{j,*}+\mathbf{U}_n\mathbf{S}_{j,*}]. \tag{42}$$

We now replace these new representations in (39), (40), (41), and (42) and rearrange the terms to obtain the desired result. ■

**Proof of Lemma 4:** The given claims are already proven by Eqs. (26) and (38). ■

**Proof of Corollary 1:** We let $\mathbf{Q}_n:=(\mathbf{A}_*^{-1}\mathbf{B}_*-\det[\mathbf{D}_*]^{\frac{1}{k}}\mathbf{I})$ for notational simplicity. We also have

$\widehat{S}_n = (T_* - D_*) + \alpha_n + \beta_n + o_{\mathbb{P}}(n^{-1})$, where $\alpha_n := \alpha_n^{(1)} + \alpha_n^{(2)}$, $\alpha_n^{(1)} := \frac{1}{k}\mathrm{tr}[\mathbf{Q}_n\mathbf{L}_n]$,

$$\alpha_n^{(2)} := \frac{1}{k}[\mathrm{tr}[\mathbf{Q}_n\mathbf{J}_{j,n}]]'\widetilde{\boldsymbol{\theta}}_n - \frac{1}{k}\widetilde{\boldsymbol{\theta}}_n'[\mathrm{tr}[\mathbf{Q}_n\mathbf{R}_{j,*}\mathbf{A}_*^{-1}\partial_i\mathbf{A}_* + \mathbf{Q}_n(\mathbf{B}_*^{-1}\partial_{ji}^2\mathbf{B}_* - \mathbf{A}_*^{-1}\partial_{ji}^2\mathbf{A}_*)]]'\,\widetilde{\boldsymbol{\theta}}_n,$$

and

$$\begin{aligned}\beta_n := &-\frac{1}{k}\mathrm{tr}[\mathbf{A}_*^{-1}\mathbf{B}_*\mathbf{L}_n\mathbf{U}_n] - \det[\mathbf{D}_*]^{\frac{1}{k}}\{\frac{1}{2k^2}\mathrm{tr}[\mathbf{L}_n]^2 - \frac{1}{2k}\mathrm{tr}[\mathbf{W}_n^2 - \mathbf{U}_n^2]\}\\ &-\frac{1}{k}[\mathrm{tr}[\mathbf{A}_*^{-1}\mathbf{B}_*\mathbf{P}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*]]'\widetilde{\boldsymbol{\theta}}_n + \frac{1}{k}\det[\mathbf{D}_*]^{\frac{1}{k}}[\mathrm{tr}[\mathbf{U}_n\mathbf{R}_{j,*} + \mathbf{P}_n\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*]]'\widetilde{\boldsymbol{\theta}}_n\\ &+\frac{1}{2k}\det[\mathbf{D}_*]^{\frac{1}{k}}\widetilde{\boldsymbol{\theta}}_n'[\mathrm{tr}[\mathbf{R}_{j,*}\mathbf{R}_{i,*}]]\widetilde{\boldsymbol{\theta}}_n\end{aligned}$$

by Lemmas 4, A5, and A6.

($i$) It is not hard to see that $\alpha_n^{(1)} = O_{\mathbb{P}}(n^{-1/2})$, $\alpha_n^{(2)} = O_{\mathbb{P}}(n^{-1})$, and $\beta_n = O_{\mathbb{P}}(n^{-1})$. Thus, $\widehat{S}_n = (T_* - D_*) + \alpha_n^{(1)} + o_{\mathbb{P}}(n^{-1/2}) = (T_* - D_*) + \frac{1}{k}\mathrm{tr}[(\mathbf{A}_*^{-1}\mathbf{B}_* - \det[\mathbf{D}_*]^{\frac{1}{k}}\mathbf{I})\mathbf{L}_n] + o_{\mathbb{P}}(n^{-1/2})$ by the definition of $\mathbf{Q}_n$.

($ii$) If for some $d_* > 0$, $\mathbf{B}_* = d_*\mathbf{A}_*$, $\det[\mathbf{D}_*]^{\frac{1}{k}} = d_*$, and $\mathbf{Q}_n = \mathbf{0}$. Also, $T_* = D_*$, $\alpha_n = 0$, and

$$\begin{aligned}\beta_n = &-\frac{d_*}{2k^2}\mathrm{tr}[\mathbf{L}_n]^2 - \frac{d_*}{k}\mathrm{tr}[\mathbf{L}_n\mathbf{U}_n] + \frac{d_*}{2k}\mathrm{tr}[\mathbf{W}_n^2 - \mathbf{U}_n^2] + \frac{d_*}{2k}\widetilde{\boldsymbol{\theta}}_n'[\mathrm{tr}[\mathbf{R}_{j,*}\mathbf{R}_{i,*}]]\widetilde{\boldsymbol{\theta}}_n\\ &+\frac{d_*}{k}[\mathrm{tr}[\mathbf{U}_n\mathbf{R}_{j,*} + \mathbf{P}_n(\mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)]]'\widetilde{\boldsymbol{\theta}}_n.\end{aligned}$$

Here, we let $\mathbf{L}_n := \mathbf{W}_n - \mathbf{U}_n + \sum_{j=1}^{\ell}\widetilde{\theta}_{j,n}\mathbf{R}_{j,*}$ and $\mathbf{R}_{j,*} := \mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{A}_*^{-1}\partial_j\mathbf{A}_*$, so that

$$\beta_n = -\frac{d_*}{2k^2}\mathrm{tr}[\mathbf{L}_n]^2 + \frac{d_*}{2k}\mathrm{tr}[(\mathbf{W}_n - \mathbf{U}_n)^2 + 2(\mathbf{W}_n - \mathbf{U}_n)\sum_{j=1}^{\ell}\widetilde{\theta}_{j,n}\mathbf{R}_{j,*} + (\sum_{j=1}^{\ell}\widetilde{\theta}_{j,n}\mathbf{R}_{j,*})^2].$$

Finally, we note that $\mathrm{tr}[\mathbf{L}_n^2] = \mathrm{tr}[\mathbf{W}_n^2 + \mathbf{U}_n^2 - 2\mathbf{W}_n\mathbf{U}_n + 2(\mathbf{W}_n - \mathbf{U}_n)\sum_{j=1}^{\ell}\widetilde{\theta}_{j,n}\mathbf{R}_{j,*} + (\sum_{j=1}^{\ell}\widetilde{\theta}_{j,n}\mathbf{R}_{j,*})^2]$, yielding

$$\beta_n = -\frac{d_*}{2k^2}\mathrm{tr}[\mathbf{L}_n]^2 + \frac{d_*}{2k}\mathrm{tr}[\mathbf{L}_n^2].$$

This completes the proof. ■

The following lemmas are useful in examining the test statistics under the local alternative: $\mathbb{H}_a^{(1)}$.

**Lemma A8.** *Given the same conditions as in Lemma 5 and* $\mathbb{H}_a^{(1)}$*,*

*(i)* $\mathbf{B}_{*,n}^{-1} = \mathbf{B}_*^{-1} - n^{-1/2}\mathbf{N}_*\mathbf{B}_*^{-1} + O(n^{-1})$*;*

*(ii)* $\mathrm{tr}[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}] = k + n^{-1/2}\mathrm{tr}[\mathbf{N}_*]$*;*

*(iii)* $\widetilde{\mathbf{R}}_{j,*,n} = \mathbf{S}_{j,*} + n^{-1/2}\mathbf{C}_{j,*} + O(n^{-1})$*, where* $\widetilde{\mathbf{R}}_{j,*,n} := \mathbf{B}_{*,n}^{-1}\partial_j\mathbf{B}_{*,n} - \mathbf{A}_*^{-1}\partial_j\mathbf{A}_*$*;*

*(iv)* $\widetilde{\mathbf{P}}_n = \mathbf{M}_{o,n} - n^{-1/2}\mathbf{N}_*\mathbf{W}_{o,n} + O_{\mathbb{P}}(n^{-3/2})$*, where* $\widetilde{\mathbf{P}}_n := \mathbf{B}_{*,n}^{-1}(\mathbf{B}_n - \mathbf{B}_{*,n}) - \mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_*)$*;*

*(v)* $\widetilde{\mathbf{L}}_n = \mathbf{K}_{o,n} + n^{-1/2}\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{C}_{j,*} - n^{-1/2}\mathbf{N}_*\mathbf{W}_{o,n} + O_{\mathbb{P}}(n^{-3/2})$*, where* $\widetilde{\mathbf{L}}_n := \widetilde{\mathbf{P}}_n + \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\widetilde{\mathbf{R}}_{j,*,n}$*;*

*(vi)* $\mathrm{tr}[\widetilde{\mathbf{L}}_n\mathbf{A}_*^{-1}\mathbf{B}_{*,n}] = \mathrm{tr}[\mathbf{K}_{o,n}] + n^{-1/2}\mathrm{tr}[\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\{\mathbf{S}_{j,*}\mathbf{N}_* + \mathbf{C}_{j,*}\} - \mathbf{U}_n\mathbf{N}_*] + O_{\mathbb{P}}(n^{-3/2})$*;*

*(vii)* $\mathrm{tr}[\widetilde{\mathbf{L}}_n\mathbf{U}_n\mathbf{A}_*^{-1}\mathbf{B}_{*,n}] = \mathrm{tr}[\mathbf{K}_{o,n}\mathbf{U}_n] + O_{\mathbb{P}}(n^{-3/2})$*; and*

*(viii)* $\widetilde{\mathbf{J}}_{j,n} = \mathbf{J}_{j,o,n} + O_{\mathbb{P}}(n^{-1})$*, where* $\widetilde{\mathbf{J}}_{j,n} := \mathbf{B}_{*,n}^{-1}\partial_j(\mathbf{B}_n - \mathbf{B}_{*,n}) - \mathbf{A}_*^{-1}\partial_j(\mathbf{A}_n - \mathbf{A}_*)$. □

**Proof of Lemma A8:** (*i*) From the definition of $\mathbf{B}_{*,n}$, $\mathbf{B}_{*,n} = \mathbf{B}_*(\mathbf{I} - n^{-1/2}\mathbf{B}_*^{-1}(-\bar{\mathbf{B}}_*))$, so that $\mathbf{B}_{*,n}^{-1} = (\mathbf{I} - n^{-1/2}\mathbf{B}_*^{-1}(-\bar{\mathbf{B}}_*))^{-1}\mathbf{B}_*^{-1}$. We also note that when the sample size $n$ is moderately large, all eigenvalues of $n^{-1/2}\mathbf{B}_*^{-1}(-\bar{\mathbf{B}}_*)$ are less than one in modulus, so that

$$(\mathbf{I} - n^{-1/2}\mathbf{B}_*^{-1}(-\bar{\mathbf{B}}_*))^{-1} = \mathbf{I} - n^{-1/2}\mathbf{B}_*^{-1}\bar{\mathbf{B}}_* + n^{-1}\mathbf{B}_*^{-1}\bar{\mathbf{B}}_*\mathbf{B}_*^{-1}\bar{\mathbf{B}}_* + \ldots.$$

This implies that $\mathbf{B}_{*,n}^{-1} = \mathbf{B}_*^{-1} - n^{-1/2}\mathbf{N}_*\mathbf{B}_*^{-1} + n^{-1}\mathbf{N}_*^2\mathbf{B}_*^{-1} + \ldots$. The desired result follows from this.

(*ii*) Under $\mathbb{H}_a^{(1)}$, $\mathbf{B}_{*,n}\mathbf{A}_*^{-1} = \mathbf{I} + n^{-1/2}\bar{\mathbf{B}}_*\mathbf{B}_*^{-1}$. Thus, $\mathrm{tr}[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}] = k + n^{-1/2}\mathrm{tr}[\mathbf{N}_*]$.

(*iii*) Using Lemma A8(*i*) and $\mathbf{B}_{*,n}$, $\mathbf{B}_{*,n}^{-1}\partial_j\mathbf{B}_{*,n} = (\mathbf{B}_*^{-1} - n^{-1/2}\mathbf{N}_*\mathbf{B}_*^{-1})(\partial_j\mathbf{B}_* + n^{-1/2}\partial_j\bar{\mathbf{B}}_*) + O(n^{-1}) = \mathbf{B}_*^{-1}\partial_j\mathbf{B}_* + n^{-1/2}\mathbf{C}_{j,*} + O(n^{-1})$. Also, under $\mathbb{H}_a^{(1)}$, $\mathbf{A}_* = \mathbf{B}_*$. Thus, the desired result follows by the definition of $\mathbf{S}_{j,*}$.

(*iv*) Using Lemma A8(*i*) and the definition of $\mathbf{W}_{o,n} := \mathbf{B}_*^{-1}(\mathbf{B}_n - \mathbf{B}_{*,n})$, $\widetilde{\mathbf{P}}_n = (\mathbf{B}_*^{-1} - n^{-1/2}\mathbf{N}_*\mathbf{B}_*^{-1} + O(n^{-1}))(\mathbf{B}_n - \mathbf{B}_{*,n}) - \mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_*) = \mathbf{M}_{o,n} - n^{-1/2}\mathbf{N}_*\mathbf{W}_{o,n} + O_{\mathbb{P}}(n^{-3/2})$.

(*v*) From the definition of $\widetilde{\mathbf{L}}_n$, $\widetilde{\mathbf{L}}_n = \widetilde{\mathbf{P}}_n + \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\widetilde{\mathbf{R}}_{j,*,n}$. Lemmas A8(*iii* and *iv*) now imply that $\widetilde{\mathbf{L}}_n = \mathbf{M}_{o,n} - n^{-1/2}\mathbf{N}_*\mathbf{W}_{o,n} + \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{S}_{j,*} + n^{-1/2}\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{C}_{j,*} + O_{\mathbb{P}}(n^{-3/2})$.

(*vi*) From the definition of $\widetilde{\mathbf{L}}_n$, $\widetilde{\mathbf{L}}_n\mathbf{A}_*^{-1}\mathbf{B}_{*,n} = \widetilde{\mathbf{P}}_n\mathbf{A}_*^{-1}\mathbf{B}_{*,n} + \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\widetilde{\mathbf{R}}_{j,*,n}\mathbf{A}_*^{-1}\mathbf{B}_{*,n}$. We examine the traces of the terms in the RHS. First, $\mathrm{tr}[\widetilde{\mathbf{P}}_n\mathbf{A}_*^{-1}\mathbf{B}_{*,n}] = \mathrm{tr}[\mathbf{W}_{o,n} - \mathbf{U}_n] - n^{-1/2}\mathrm{tr}[\mathbf{U}_n\mathbf{N}_*] = \mathrm{tr}[\mathbf{M}_{o,n}] - n^{-1/2}\mathrm{tr}[\mathbf{U}_n\mathbf{N}_*]$ by noting that $\mathrm{tr}[\mathbf{B}_{*,n}^{-1}(\mathbf{B}_n - \mathbf{B}_{*,n})\mathbf{A}_*^{-1}\mathbf{B}_{*,n}] = \mathrm{tr}[\mathbf{A}_*^{-1}(\mathbf{B}_n - \mathbf{B}_{*,n})] = \mathrm{tr}[\mathbf{W}_{o,n}]$. Second, using Lemma A8(*iii*), $\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\widetilde{\mathbf{R}}_{j,*,n}\mathbf{A}_*^{-1}\mathbf{B}_{*,n} = \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{S}_{j,*} + n^{-1/2}\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{S}_{j,*}\mathbf{N}_* + n^{-1/2}\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{C}_{j,*} + O_{\mathbb{P}}(n^{-3/2})$. Therefore, if we add these two terms and note that $\mathbf{K}_{o,n} = \mathbf{M}_{o,n} + \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{S}_{j,*}$, the desired result follows.

(*vii*) From the proof of Lemma A8(*v*), $\widetilde{\mathbf{L}}_n = \mathbf{K}_{o,n} + O_{\mathbb{P}}(n^{-1})$. This implies that $\widetilde{\mathbf{L}}_n\mathbf{U}_n = \mathbf{K}_{o,n}\mathbf{U}_n + O_{\mathbb{P}}(n^{-3/2})$, so that we can conclude that $\mathrm{tr}[\widetilde{\mathbf{L}}_n\mathbf{U}_n\mathbf{A}_*^{-1}\mathbf{B}_{*,n}] = \mathrm{tr}[\mathbf{K}_{o,n}\mathbf{U}_n\mathbf{A}_*^{-1}\mathbf{B}_{*,n}] + O_{\mathbb{P}}(n^{-3/2})$.

(*viii*) By Lemma A8(*i*), it trivially holds. This completes the proof. ■

**Lemma A9.** *Given the same conditions as in Lemma 5 and* $\mathbb{H}_a^{(1)}$*,*

*(i)* $\det[\mathbf{B}_{*,n}] = \det[\mathbf{B}_*] + n^{-1/2}\det[\mathbf{B}_*]\mathrm{tr}[\mathbf{N}_*] + \frac{1}{2n}\det[\mathbf{B}_*]\{\mathrm{tr}[\mathbf{N}_*]^2 - \mathrm{tr}[\mathbf{N}_*^2]\} + O(n^{-3/2})$*;*

*(ii)* $\det[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}] = 1 + n^{-1/2}\mathrm{tr}[\mathbf{N}_*] + \frac{1}{2n}\{\mathrm{tr}[\mathbf{N}_*]^2 - \mathrm{tr}[\mathbf{N}_*^2]\} + O(n^{-3/2})$*; and*

*(iii)* $\det[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}]^{1/k} = 1 + \frac{1}{\sqrt{nk}}\mathrm{tr}[\mathbf{N}_*] + \frac{1}{2nk}\{\mathrm{tr}[\mathbf{N}_*]^2 - \mathrm{tr}[\mathbf{N}_*^2]\} + \frac{1}{2nk}(\frac{1}{k}-1)\mathrm{tr}[\mathbf{N}_*]^2 + O(n^{-3/2})$. □

**Proof of Lemma A9:** (*i*) We can apply Taylor's expansion to the determinant. That is,

$$\det[\mathbf{B}_{*,n}] = \det[\mathbf{B}_*] + \det[\mathbf{B}_*]\mathrm{tr}[\mathbf{B}_*^{-1}(\mathbf{B}_{*,n} - \mathbf{B}_*)] \\ + \frac{\det[\mathbf{B}_*]}{2}\left\{\mathrm{tr}[\mathbf{B}_*^{-1}(\mathbf{B}_{*,n} - \mathbf{B}_*)]^2 - \mathrm{tr}[\mathbf{B}_*^{-1}(\mathbf{B}_{*,n} - \mathbf{B}_*)\mathbf{B}_*^{-1}(\mathbf{B}_{*,n} - \mathbf{B}_*)]\right\} + o_{\mathbb{P}}(n^{-1}).$$

We also note that $\mathbf{B}_{*,n} - \mathbf{B}_* = n^{-1/2}\bar{\mathbf{B}}_*$, and plugging this into the above equation now yields the desired result.

(*ii*) By Lemma A9(*i*) and the facts that $\det[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}] = \det[\mathbf{B}_{*,n}]/\det[\mathbf{A}_*]$ and $\det[\mathbf{B}_*] = \det[\mathbf{A}_*]$, the desired result follows by dividing the equation in Lemma A9(*i*) with $\det[\mathbf{A}_*]$.

(*iii*) If we expand the power function $x^{1/k}$ around 1,

$$x^{1/k} = 1 + \frac{1}{k}(x-1) + \frac{1}{2k}\left(\frac{1}{k}-1\right)(x-1)^2 + \dots.$$

We now let $x$ be $\det[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}]$ and instead plug its expansion in Lemma A9(*ii*) into this equation. Next, we arrange the terms in the RHS according to their convergence rates. The desired result follows from this. ■

**Proof of Lemma 5:** (*i*) As the key parameter is now $\mathbf{B}_{*,n}$, we approximate $\widehat{T}_n$ around $\mathbf{A}_*$ and $\mathbf{B}_{*,n}$ using Lemma 4(*i*). Then,

$$\widehat{T}_n = \frac{1}{k}\mathrm{tr}[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}] - 1 + \frac{1}{k}\mathrm{tr}[\widetilde{\mathbf{L}}_n\mathbf{A}_*^{-1}\mathbf{B}_{*,n}] - \frac{1}{k}\mathrm{tr}[\widetilde{\mathbf{L}}_n\mathbf{U}_n\mathbf{A}_*^{-1}\mathbf{B}_{*,n}] \\ + \frac{1}{k}[\mathrm{tr}[(\widetilde{\mathbf{J}}_{j,n} - \widetilde{\mathbf{P}}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)\mathbf{A}_*^{-1}\mathbf{B}_{*,n}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + \frac{1}{2k}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1}).$$

We note that the asymptotic approximation of the first line in this equation is provided in Lemma A8(*iii*, *vi*, and *vii*). In addition, Lemma A8 provides the asymptotic expansions of $\widetilde{\mathbf{P}}_n$ and $\widetilde{\mathbf{K}}_{j,n}$, so that

$$\frac{1}{k}[\mathrm{tr}[(\widetilde{\mathbf{J}}_{j,n} - \widetilde{\mathbf{P}}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)\mathbf{A}_*^{-1}\mathbf{B}_{*,n}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + \frac{1}{2k}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \\ = \frac{1}{k}[\mathrm{tr}[(\widetilde{\mathbf{J}}_{j,o,n} - \mathbf{M}_{o,n}\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)\mathbf{A}_*^{-1}\mathbf{B}_{*,n}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + \frac{1}{2k}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*^{-1}](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1}).$$

Therefore, if we collect all these results, it is not hard to obtain that

$$\widehat{T}_n = \widehat{T}_{o,n} + \frac{1}{\sqrt{n}k}\left\{\mathrm{tr}[\mathbf{N}_*] - \mathrm{tr}[\mathbf{N}_*\mathbf{U}_n] + [\mathrm{tr}[\mathbf{A}_*^{-1}(\partial_j\bar{\mathbf{B}}_*) - \mathbf{N}_*\mathbf{A}_*^{-1}(\partial_j\bar{\mathbf{A}}_*)]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)\right\} + o_{\mathbb{P}}(n^{-1})$$

using the definition of $\widehat{T}_{o,n}$. This is the desired result.

(*ii*) The proof is almost identical. If we approximate $\widehat{D}_n$ around $\mathbf{A}_*$ and $\mathbf{B}_{*,n}$ using Lemma 4(*ii*),

$$\begin{aligned}\widehat{D}_n =& \det[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}]^{\frac{1}{k}} - 1 + \frac{1}{k}\det[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}]^{\frac{1}{k}}\mathrm{tr}[\widetilde{\mathbf{L}}_n] + \frac{1}{2k}\left(\frac{1}{k}-1\right)\det[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}]^{\frac{1}{k}}\mathrm{tr}[\widetilde{\mathbf{L}}_n]^2\\ &+ \frac{1}{k}\det[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}]^{\frac{1}{k}}\mathrm{tr}[\widetilde{\mathbf{P}}_n][\mathrm{tr}[\widetilde{\mathbf{R}}_{j,*,n}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + \frac{1}{2k}\det[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}]^{\frac{1}{k}}(\mathrm{tr}[\widetilde{\mathbf{P}}_n]^2 + \mathrm{tr}[\mathbf{U}_n^2] - \mathrm{tr}[\widetilde{\mathbf{W}}_n^2])\\ &+ \frac{1}{k}\det[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}]^{\frac{1}{k}}[\mathrm{tr}[\widetilde{\mathbf{J}}_{j,n} + \mathbf{U}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_* - \widetilde{\mathbf{W}}_n\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)\\ &+ \frac{1}{2k}\det[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}]^{\frac{1}{k}-1}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla^2_{\boldsymbol{\theta}}\det[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1}),\end{aligned}$$

where $\widetilde{\mathbf{W}}_n := \mathbf{B}_{*,n}^{-1}(\mathbf{B}_n - \mathbf{B}_{*,n})$. We note that the last three lines of this equation is $O_{\mathbb{P}}(n^{-1})$, so that they can be reformulated into

$$\begin{aligned}&\frac{1}{k}\mathrm{tr}[\mathbf{M}_{o,n}][\mathrm{tr}[\mathbf{S}_{j,*}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + \frac{1}{2k}(\mathrm{tr}[\mathbf{M}_{o,n}]^2 + \mathrm{tr}[\mathbf{U}_n^2] - \mathrm{tr}[\mathbf{W}_{o,n}^2])\\ &+ \frac{1}{k}[\mathrm{tr}[\mathbf{J}_{j,o,n} + \mathbf{U}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_* - \mathbf{W}_{o,n}\mathbf{A}_*^{-1}\partial_j\mathbf{B}_*]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + \frac{1}{2k}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla^2_{\boldsymbol{\theta}}\det[\mathbf{D}_*](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + O_{\mathbb{P}}(n^{-3/2})\end{aligned}$$

under $\mathbb{H}_a^{(1)}$. On the other hand, the first line of the equation is formed by $\det[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}]$ and $\mathrm{tr}[\widetilde{\mathbf{L}}_n]$, and their asymptotic expansions are given in Lemmas A8($v$) and A9(*iii*), respectively. Using these, we obtain that

$$\begin{aligned}&\det[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}]^{\frac{1}{k}} - 1 + \frac{1}{k}\det[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}]^{\frac{1}{k}}\mathrm{tr}[\widetilde{\mathbf{L}}_n] + \frac{1}{2k}\left(\frac{1}{k}-1\right)\det[\mathbf{B}_{*,n}\mathbf{A}_*^{-1}]^{\frac{1}{k}}\mathrm{tr}[\widetilde{\mathbf{L}}_n]^2\\ =& \frac{1}{k}\mathrm{tr}[\widetilde{\mathbf{K}}_{o,n}] + \frac{1}{\sqrt{n}k}\mathrm{tr}[\mathbf{N}_*] + \frac{1}{2k}\left(\frac{1}{k}-1\right)\mathrm{tr}[\widetilde{\mathbf{K}}_{o,n}]^2 - \frac{1}{\sqrt{n}k}\mathrm{tr}[\mathbf{N}_*\mathbf{W}_{o,n}] + \frac{1}{\sqrt{n}k}[\mathrm{tr}[\mathbf{C}_{j,*}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)\\ &+ \frac{1}{\sqrt{n}k^2}\mathrm{tr}[\mathbf{N}_*]\mathrm{tr}[\mathbf{K}_{o,n}] + \frac{1}{2nk^2}\mathrm{tr}[\mathbf{N}_*]^2 - \frac{1}{2nk}\mathrm{tr}[\mathbf{N}_*^2] + O_{\mathbb{P}}(n^{-3/2}).\end{aligned}$$

We now combine them and note the definition of $\widehat{D}_{o,n}$ to obtain the desired result.

(*iii*) Using the definition of $\widehat{S}_n := \widehat{T}_n - \widehat{D}_n$, we rearrange the terms according to their convergence rate. Using the fact that $\mathbf{S}_{j,*} := \mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{A}_*^{-1}\partial\mathbf{A}_*$ under $\mathbb{H}_a^{(1)}$, we can derive the desired result. This completes the proof. ∎

**Proof of Theorem 5:** (*i*) From Lemma 5, we note that $k\sqrt{n}\widehat{T}_n = \mathrm{tr}[\mathbf{N}_*] + k\sqrt{n}\widehat{T}_{o,n} + O_{\mathbb{P}}(n^{-1/2})$ and

$k\sqrt{n}\widehat{D}_n = \mathrm{tr}[\mathbf{N}_*] + k\sqrt{n}\widehat{D}_{o,n} + O_{\mathbb{P}}(n^{-1/2})$, so that

$$\frac{nk^2}{2}(\widehat{T}_n^2 + \widehat{D}_n^2) = \mathrm{tr}[\mathbf{N}_*]^2 + k\sqrt{n}(\widehat{T}_{o,n} + \widehat{D}_{o,n})\mathrm{tr}[\mathbf{N}_*] + \frac{k^2 n}{2}(\widehat{T}_{o,n}^2 + \widehat{D}_{o,n}^2) + O_{\mathbb{P}}(n^{-1/2}).$$

Also, $\widehat{T}_{o,n} = \frac{1}{k}\mathrm{tr}[\mathbf{K}_{o,n}] + O_{\mathbb{P}}(n^{-1})$ and $\widehat{D}_{o,n} = \frac{1}{k}\mathrm{tr}[\mathbf{K}_{o,n}] + O_{\mathbb{P}}(n^{-1})$ by the definitions of $\widehat{T}_{o,n}$ and $\widehat{D}_{o,n}$ and the given conditions. Thus, it follows that

$$\frac{nk^2}{2}(\widehat{T}_n^2 + \widehat{D}_n^2) = \mathrm{tr}[\mathbf{N}_*]^2 + 2\sqrt{n}\mathrm{tr}[\mathbf{K}_{o,n}]\mathrm{tr}[\mathbf{N}_*] + n\mathrm{tr}[\mathbf{K}_{o,n}]^2 + O_{\mathbb{P}}(n^{-1/2}),$$

and we note that $\widehat{\mathscr{B}}_n^{(1)} := \frac{nk^2}{2}(\widehat{T}_n^2 + \widehat{D}_n^2)$, implying that $\widehat{\mathscr{B}}_n^{(1)} = (\mathrm{tr}[\mathbf{N}_*] + \sqrt{n}\mathrm{tr}[\mathbf{K}_{o,n}])^2 + O_{\mathbb{P}}(n^{-1/2}) = (\mathrm{tr}[\mathbf{N}_* + \sqrt{n}\mathbf{K}_{o,n}])^2 + O_{\mathbb{P}}(n^{-1/2})$. This is the desired result.

(*ii*) By Lemma 5(*i*), $\sqrt{n}\widehat{T}_n = \frac{1}{k}\mathrm{tr}[\mathbf{N}_*] + \frac{\sqrt{n}}{k}\mathrm{tr}[\mathbf{K}_{o,n}] + O_{\mathbb{P}}(n^{-1/2})$, so that

$$nk\widehat{T}_n^2 = \frac{1}{k}\mathrm{tr}[\mathbf{N}_*]^2 + \frac{n}{k}\mathrm{tr}[\mathbf{K}_{o,n}]^2 + \frac{2\sqrt{n}}{k}\mathrm{tr}[\mathbf{N}_*]\mathrm{tr}[\mathbf{K}_{o,n}] + O_{\mathbb{P}}(n^{-1/2}).$$

Next, Lemma 5(*iii*) implies that

$$\begin{aligned}2nk\widehat{S}_n =& \mathrm{tr}[\mathbf{N}_*^2] + n\mathrm{tr}[\mathbf{K}_{o,n}^2] - \frac{1}{k}\mathrm{tr}[\mathbf{N}_*]^2 - \frac{n}{k}\mathrm{tr}[\mathbf{K}_{o,n}]^2 - \frac{2\sqrt{n}}{k}\mathrm{tr}[\mathbf{N}_*]\mathrm{tr}[\mathbf{K}_{o,n}] \\ &+ 2\sqrt{n}\mathrm{tr}[\mathbf{N}_*\mathbf{M}_{o,n}] + 2\sqrt{n}[\mathrm{tr}[\mathbf{N}_*\mathbf{A}_*^{-1}(\partial_j\mathbf{B}_* - \partial_j\mathbf{A}_*)]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + O_{\mathbb{P}}(n^{-1/2}).\end{aligned}$$

Thus, if we add $nk\widehat{T}_n^2$ to $2nk\widehat{S}_n$,

$$\begin{aligned}nk\widehat{T}_n^2 + 2nk\widehat{S}_n =& \mathrm{tr}[\mathbf{N}_*^2] + n\mathrm{tr}[\mathbf{K}_{o,n}^2] \\ &+ 2\sqrt{n}\left\{\mathrm{tr}[\mathbf{N}_*\mathbf{M}_{o,n}] + [\mathrm{tr}[\mathbf{N}_*\mathbf{A}_*^{-1}(\partial_j\mathbf{B}_* - \partial_j\mathbf{A}_*)]]'(\widehat{\theta}_{j,n} - \theta_{j,*})\right\} + O_{\mathbb{P}}(n^{-1/2}).\end{aligned}$$

We now note that $\mathrm{tr}[\mathbf{N}_*\mathbf{M}_{o,n}] + [\mathrm{tr}[\mathbf{N}_*\mathbf{A}_*^{-1}(\partial_j\mathbf{B}_* - \partial_j\mathbf{A}_*)]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) = \mathrm{tr}[\mathbf{N}_*\mathbf{K}_{o,n}]$, so that $\widehat{\mathscr{B}}_n^{(2)} = nk\widehat{T}_n^2 + 2nk\widehat{S}_n = \mathrm{tr}[\mathbf{N}_*^2] + 2\sqrt{n}\mathrm{tr}[\mathbf{N}_*\mathbf{K}_{o,n}] + n\mathrm{tr}[\mathbf{K}_{o,n}^2] + O_{\mathbb{P}}(n^{-1/2})$, implying that $\widehat{\mathscr{B}}_n^{(2)} = \mathrm{tr}[(\mathbf{N}_* + \sqrt{n}\mathbf{K}_{o,n})^2] + o_{\mathbb{P}}(1)$, as desired.

(*iii*) By Lemma 5(*ii*), $\sqrt{n}\widehat{D}_n = \frac{1}{k}\mathrm{tr}[\mathbf{N}_*] + \frac{\sqrt{n}}{k}\mathrm{tr}[\mathbf{K}_{o,n}] + O_{\mathbb{P}}(n^{-1/2})$, and this expansion is the same as that of $\sqrt{n}\widehat{T}_n$. We can thus conclude that $\widehat{\mathscr{B}}_n^{(3)}$ has the same asymptotic expansion as $\widehat{\mathscr{B}}_n^{(2)}$. This completes the proof. ∎

**Proof of Lemma 6**: (*i*) We note that $\widehat{L}_n = \ln[\widehat{D}_n + 1]$ and $\widehat{D}_n = O_{\mathbb{P}}(n^{-1/2})$. We now approximate $\widehat{L}_n$

around 1 by using Taylor's expansion, which yields that

$$\widehat{L}_n = \widehat{D}_n - \frac{1}{2}\widehat{D}_n^2 + o_{\mathbb{P}}(n^{-1}). \tag{43}$$

Given this, Lemma 5(*ii*) provides the asymptotic expansion of $\widehat{D}_n$, so that it follows that

$$\frac{1}{2}\widehat{D}_n^2 = \frac{1}{2nk^2}\mathrm{tr}[\mathbf{N}_*]^2 + \frac{1}{2k^2}\mathrm{tr}[\mathbf{K}_{o,n}]^2 + \frac{1}{\sqrt{n}k^2}\mathrm{tr}[\mathbf{N}_*]\mathrm{tr}[\mathbf{K}_{o,n}] + o_{\mathbb{P}}(n^{-1}). \tag{44}$$

We now combine the results in Lemma 5(*ii*) and Eq. (44) with Eq. (43), and this yields the desired result.

(*ii*) We note that $\widehat{Q}_n = \ln[\widehat{T}_n + 1]$ and $\widehat{T}_n = O_{\mathbb{P}}(n^{-1/2})$. We now approximate $\widehat{Q}_n$ around 1 by Taylor's expansion, and we obtain that

$$\widehat{Q}_n = \widehat{T}_n - \frac{1}{2}\widehat{T}_n^2 + o_{\mathbb{P}}(n^{-1}). \tag{45}$$

We further note that

$$\widehat{T}_n^2 = \frac{1}{nk^2}\mathrm{tr}[\mathbf{N}_*]^2 + \frac{1}{k^2}\mathrm{tr}[\mathbf{K}_{o,n}]^2 + \frac{2}{\sqrt{n}k^2}\mathrm{tr}[\mathbf{N}_*]\mathrm{tr}[\mathbf{K}_{o,n}] + o_{\mathbb{P}}(n^{-1}) \tag{46}$$

and combine this with Lemma 5(*i*) according to Eq. (45). This gives the desired result.

(*iii*) We simply combine the results in Lemmas 5(*i*) and 6(*ii*) by using the definition of $\widehat{M}_n$. This shows that

$$\widehat{M}_n = \frac{1}{2kn}\mathrm{tr}[(\mathbf{N}_* + \sqrt{n}\mathbf{K}_{o,n})^2] + \frac{1}{2k^2}\mathrm{tr}[\mathbf{K}_{o,n}]^2 - \frac{1}{2k}\mathrm{tr}[\mathbf{K}_{o,n}^2] + \widehat{S}_{o,n} + o_{\mathbb{P}}(n^{-1}).$$

We next substitute $\widehat{S}_{o,n}$ in this equation with its definition. Most terms are canceled by others, and the only surviving term is $\mathrm{tr}[(\mathbf{N}_* + \sqrt{n}\mathbf{K}_{o,n})^2]/(2kn) + o_{\mathbb{P}}(n^{-1})$.

(*iv*) We combine the results in Lemma 6(*i* and *ii*) by using the definition of $\widehat{W}_n$. Then, it follows that

$$\begin{aligned}\widehat{W}_n &= \widehat{Q}_n - \widehat{L}_n \\ &= \frac{1}{2nk}\mathrm{tr}[(\mathbf{N}_* + \sqrt{n}\mathbf{K}_{o,n})^2] - \frac{1}{2nk^2}\mathrm{tr}[\mathbf{N}_* + \sqrt{n}\mathbf{K}_{o,n}]^2 + \frac{1}{2k^2}\mathrm{tr}[\mathbf{K}_{o,n}]^2 - \frac{1}{2k}\mathrm{tr}[\mathbf{K}_{o,n}^2] + \widehat{S}_{o,n} + o_{\mathbb{P}}(n^{-1}).\end{aligned}$$

We now use the definition of $\widehat{S}_{o,n}$ to obtain the desired result. This completes the proof. ∎

**Proof of Theorem 6:** (*i*) We note that $\mathscr{LR}_n^{(1)} = 2\{\ln[L_n(\widehat{\boldsymbol{\theta}}_n, \widehat{\mathbf{B}}_n)] - \ln[L_n(\widehat{\boldsymbol{\theta}}_n, \widehat{\mathbf{A}}_n)]\} = nk(\widehat{T}_n - \widehat{L}_n) =$

$nk\widehat{M}_n = \frac{1}{2}\text{tr}[(\mathbf{N}_* + \sqrt{n}\mathbf{K}_{o,n})^2] + o_{\mathbb{P}}(n^{-1})$ by Lemma 6(*ii*).

(*ii* and *iii*) The desired results follow from Theorem 6(*i*) and Theorem 5(*ii* and *iii*). ■

**Proof of Theorem 7:** (*i*) We note that $\mathscr{LR}_n^{(2)} = -n\ln[\det(\widehat{\mathbf{B}}_n)] + nk\ln[\text{tr}[\widehat{\mathbf{A}}_n^{-1}\widehat{\mathbf{B}}_n]/k] + n\ln[\det(\widehat{\mathbf{A}}_n)] = n(k\ln[\text{tr}[\widehat{\mathbf{A}}_n^{-1}\widehat{\mathbf{B}}_n]/k] - \ln[\det(\widehat{\mathbf{A}}_n^{-1}\widehat{\mathbf{B}}_n)]) = n(k\widehat{Q}_n - k\widehat{L}_n) = nk\widehat{W}_n$. Given this, the desired result holds by Lemma 6(*iv*).

(*ii* and *iii*) We can combine Theorem 7(*i*) with Theorem 5 to complete the proof. ■

## B.2 Additional Assumptions for Theorem 2

**Assumption A 1.** $(\Omega, \mathscr{F}, \mathbb{P})$ *is a complete probability space.* □

**Assumption A 2.** (*i*) *For* $k \in \mathbb{N}$*, the symmetric matrix* $\mathbf{A}_* := E[\mathbf{a}(Y_t, \mathbf{X}_t)\mathbf{a}(Y_t, \mathbf{X}_t)'] - E[\mathbf{a}(Y_t, \mathbf{X}_t)]E[\mathbf{a}(Y_t, \mathbf{X}_t)']$ *is positive definite; and*

(*ii*) *the symmetric matrix* $\mathbf{B}_* := E[\mathbf{b}(Y_t, \mathbf{X}_t)\mathbf{b}(Y_t, \mathbf{X}_t)'] - E[\mathbf{b}(Y_t, \mathbf{X}_t)]E[\mathbf{b}(Y_t, \mathbf{X}_t)']$ *is positive definite.*

□

**Assumption A 3.** (*i*) *The estimators* $\widetilde{\mathbf{A}}_n$ *and* $\widetilde{\mathbf{B}}_n$ *are consistent for* $\mathbf{A}_*$ *and* $\mathbf{B}_*$*, respectively; and*

(*ii*) $\sqrt{n}[\text{vech}[(\widehat{\Sigma}_{\mathbf{a},n} - \Sigma_{\mathbf{a},*})]', \text{vech}[(\widehat{\Sigma}_{\mathbf{b},n} - \Sigma_{\mathbf{b},*})]', (\widehat{\boldsymbol{\mu}}_{\mathbf{a},n} - \boldsymbol{\mu}_{\mathbf{a},*})', (\widehat{\boldsymbol{\mu}}_{\mathbf{b},n} - \boldsymbol{\mu}_{\mathbf{b},*})']' \overset{A}{\sim} N(\mathbf{0}, \Upsilon_*)$*, where* $\Upsilon_*$ *is a* $(k^2+3k)\times(k^2+3k)$ *positive semi-definite matrix.* □

**Assumption A 4.** (*i*) *The estimators* $\widetilde{\mathbf{A}}_n$ *and* $\widetilde{\mathbf{B}}_n$ *are consistent for* $\mathbf{A}_*$ *and* $\mathbf{B}_*$*, respectively; and*

(*ii*) $\sqrt{n}[\text{vech}[(\widehat{\Sigma}_{\mathbf{a},n} - \Sigma_{\mathbf{a},*})]', \text{vech}[(\widehat{\Sigma}_{\mathbf{b},n} - \Sigma_{\mathbf{b},*,n})]', (\widehat{\boldsymbol{\mu}}_{\mathbf{a},n} - \boldsymbol{\mu}_{\mathbf{a},*})', (\widehat{\boldsymbol{\mu}}_{\mathbf{b},n} - \boldsymbol{\mu}_{\mathbf{b},*,n})']' \overset{A}{\sim} N(\mathbf{0}, \Upsilon_*)$*, where* $\Upsilon_*$ *is a* $(k^2+3k)\times(k^2+3k)$ *positive semi-definite matrix.* □

**Assumption A 5.** *The symmetric mapping* $\bar{\mathbf{B}}_*$ *is positive definite.* □

# References


ABADIR, K. AND MAGNUS, J. (2005): *Matrix Algebra*. NY: Cambridge University Press.

ANDERSON, T. (2003): *An Introduction to Multivariate Statistical Analysis*. Hoboken, NJ: John Wiley & Sons, Inc.

BERA, A. (1986): "Model Specification Test Through Eigenvalues," Department of Economics, University of Illinois, mimeo.

BERA, A. AND HALL, A. (1991): "An Eignevalue Based Test for Heteroskedasticity," Department of Economics, University of Illinois, mimeo.

CHESHER, A. (1983): "Information Matrix Test," *Economics Letters*, 13, 45–48.

CHESHER, A. AND SPADY, R. (1991): "Asymptotic Expansions of the Information Matrix Test Statistic," *Econometrica*, 59, 787–815.

CHO, J. S. AND WHITE, H. (2007): "Testing for Regime Switching," *Econometrica*, 75, 1671–1720.

CHO, J. S. AND WHITE, H. (2010): "Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models," *Journal of Econometrics*, 157, 458–480.

CHO, J. S., CHEONG, T. U., AND WHITE, H. (2011): "Experience with the Weighted Bootstrap in Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models," *Journal of Economic Theory and Econometrics*, 22:2, 60–91

DAS GUPTA, S. AND GIRI, N. (1973): "Properties of Tests Concerning Covariance Matrices of Normal Distributions," *Annals of Statistics*, 1, 1222-1224.

DHAENE, G. AND HOORELBEKE, D. (2004): "The Information Matrix Test with Bootstrap-Based Covariance Matrix Estimation," *Economics Letters*, 82, 341–347.

FISHER, R. (1922): "On the Mathematical Foundations of Theoretical Statistics," *Philosophical Transactions of the Royal Society, A*, 222, 309–368.

FISHER, R. (1925): "Theory of Statistical Estimation," *Proceedings of the Cambridge Philosophical Society*, 22, 700–725.

GOLDEN, R., HENLEY, S., WHITE, H., AND KASHNER, T. (2013): "New Directions in Information Matrix Testing: Eigenspectrum Tests," in *Causality, Prediction, and Specification Analysis: Recent Advances and Future Directions Essays in Honour of Halbert L. White, Jr.*, eds. Norman Rasmus Swanson and Xiaohong Chen. New York: Springer, pp. 145–177.

HALL, A. (1987): "The Information Matrix Test for the Linear Model," *Review of Economic Studies*, 54, 257–263.

HECKMAN, J. AND SINGER, B. (1984): "A Method of Minimizing the Impact of Distributional Assumptions in Econometric Models of Duration Data," *Econometrica*, 52, 271–320.

HOROWITZ, J. (1994): "Bootstrap-based Critical Values for the Information Matrix Test," *Journal of Econometrics*, 61, 395–411.

JARQUE, C. AND BERA, A. (1987): "A Test for Normality of Observations and Regression Residuals," *International Statistical Review*, 55, 163–218.

KING, M., ZHANG, X., AND AKRAM, M. (2011): "A New Procedure for Multiple Testing of Econometric Models," Discussion Paper, Department of Econometrics and Business Statistics, Monash University.

LANCASTER, T. (1979): "Econometric Methods for the Duration of Unemployment," *Econometrica*, 47, 939–956.

LANCASTER, T. (1984): "The Covariance Matrix of the Information Matrix Test," *Econometrica*, 52, 1052–1053.

MAGNUS, J. AND NEUDECKER, H. (1999): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester: John Wiley & Sons.

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*. New York: John Wiley & Sons, Inc.

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.

NAGAO, H. (1973): "On Some Test Criteria for Covariance Matrix," *Annals of Statistics*, 1, 700–709.

NAGAO, H. (1974): "Asymptotic Non-null Distributions of Two Test Criteria for Equality of Covariance Matrices under Local Alternatives," *Annals of the Institute of Statistical Mathematics*, 26, 395–402.

NAGARSENKER, B. N. AND PILLAI, K. C. S. (1973): "The Distribution of the Sphericity Test Criterion," *Journal of Multivariate Analysis*, 3, 226-235.

ORME, C. (1988): "The Calculation of the Information Matrix Test for Binary Data Models," *The Manchester School*, 56, 370–376.

ORME, C. (1990): "The Small-Sample Performance of the Information-Matrix Test," *Journal of Econometrics*, 46, 309–331.

PILLAI, K. AND JAYACHANDRAN, K. (1968): "Power Comparison of Tests of Equality of Two Covariance Matrices Based on Four Criteria," *Biometrika*, 55, 335–342.

PILLAI, K. AND NAGARSENKER, B. (1972): "On the Distribution of a Class of Statistics in Multivariate Analysis," *Journal of Multivariate Analysis*, 2, 96–114.

STRAWDERMAN, R. (1994): "A Note on Necessary and Sufficient Conditions for Proving that a Random Symmeetric Matrix Converges to a Given Limit," *Statiistics & Probability Letters*, 21, 367–370.

TAYLOR, L. (1987): "The Size Bias of White's Information Matrix Test," *Economics Letters*, 24, 63–67.

WALD, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," *Transactions of American Mathematical Society*, 54, 426–482.

ROY, S. (1953): "On a Heuristic Method of Test Construction and Its Use in Multivariate Analysis," *Annals of Mathematical Statistics*, 24, 220–238.

VAN DEN BERG, G. AND RIDDER, G. (1998): "An Empirical Equilibrium Search Model of the Labor Market," *Econometrica*, 66, 1183–1221.

VAN DER VAART, A. (2000): *Asymptotic Statistics*. Cambridge University Press.

WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–26.

WHITE, H. (1987): "Specification Testing in Dynamic Models," in T. Bewley ed., *Advances in Econometrics – Fifth World Congress*. Vol. 1. New York: Cambridge University Press, pp. 1–58.

WHITE, H. (1994): *Estimation, Inference and Specification Analysis*. New York: Cambridge University Press.

WILKS, S. (1935): "On the Independence of $k$ sets of Normally Distributed Statistical Variables," *Econometrica*, 3, 309–326.

| Conditions for $(T_*, D_*)$ | Relationships |
|---|---|
| A | $\widehat{\mathscr{B}}_n^{(2)} \gtrsim \widehat{\mathscr{B}}_n^{(1)} \gtrsim \widehat{\mathscr{B}}_n^{(3)}$ |
| B | $\widehat{\mathscr{B}}_n^{(1)} \gtrsim \widehat{\mathscr{B}}_n^{(3)} \gtrsim \widehat{\mathscr{B}}_n^{(2)}$ |
| C | $\widehat{\mathscr{B}}_n^{(2)} \gtrsim \widehat{\mathscr{B}}_n^{(3)} \gtrsim \widehat{\mathscr{B}}_n^{(1)}$ |
| D | $\widehat{\mathscr{B}}_n^{(3)} \gtrsim \widehat{\mathscr{B}}_n^{(2)} \gtrsim \widehat{\mathscr{B}}_n^{(1)}$ |
| E | $\widehat{\mathscr{B}}_n^{(3)} \gtrsim \widehat{\mathscr{B}}_n^{(1)} \gtrsim \widehat{\mathscr{B}}_n^{(2)}$ |
| F | $\widehat{\mathscr{B}}_n^{(1)} \gtrsim \widehat{\mathscr{B}}_n^{(2)} \gtrsim \widehat{\mathscr{B}}_n^{(3)}$ |

Table 1: GLOBAL POWER RELATIONSHIPS. This table shows the power orders of $\widehat{\mathscr{B}}_n^{(1)}$, $\widehat{\mathscr{B}}_n^{(2)}$, and $\widehat{\mathscr{B}}_n^{(3)}$ under Conditions A to F. These orders are obtained by the divergence rates of the test statistics given by Eqs. (6), (7), and (8). This implies that if the sample size is fairly large and $(T_*, D_*)$ belongs to the region indexed by F, say, then $\widehat{\mathscr{B}}_n^{(1)}$ is more powerful than $\widehat{\mathscr{B}}_n^{(2)}$, and $\widehat{\mathscr{B}}_n^{(2)}$ is more powerful than $\widehat{\mathscr{B}}_n^{(3)}$.

| Statistics | Levels \ $n$ | 50 | 100 | 200 | 400 | 600 | 800 | 1,000 | 2,000 |
|---|---|---|---|---|---|---|---|---|---|
| | 1% | 1.07 | 1.02 | 1.03 | 1.05 | 1.21 | 1.18 | 1.06 | 1.01 |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 5% | 5.32 | 5.42 | 4.99 | 5.12 | 5.25 | 5.33 | 4.98 | 5.01 |
| | 10% | 10.48 | 10.23 | 9.82 | 10.06 | 10.06 | 10.44 | 9.86 | 10.14 |
| | 1% | 1.02 | 1.12 | 1.13 | 1.12 | 1.08 | 1.13 | 1.07 | 1.06 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 5% | 5.00 | 5.10 | 4.95 | 5.29 | 5.28 | 5.22 | 4.86 | 5.20 |
| | 10% | 9.81 | 10.23 | 10.16 | 10.19 | 10.46 | 10.30 | 9.83 | 10.23 |
| | 1% | 1.06 | 1.08 | 1.14 | 1.15 | 1.01 | 1.14 | 1.04 | 1.02 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 5% | 5.02 | 5.15 | 4.84 | 5.25 | 5.33 | 5.21 | 4.89 | 5.09 |
| | 10% | 9.90 | 10.10 | 9.97 | 10.30 | 10.51 | 10.24 | 9.85 | 10.17 |
| | 1% | 36.93 | 25.91 | 16.68 | 10.40 | 7.98 | 6.51 | 5.73 | 3.43 |
| $IM_n$ | 5% | 53.28 | 40.14 | 28.45 | 20.22 | 16.45 | 14.15 | 13.22 | 9.60 |
| | 10% | 62.79 | 49.01 | 36.45 | 27.38 | 23.51 | 20.86 | 19.61 | 15.56 |
| | 1% | 1.62 | 1.90 | 1.76 | 1.78 | 1.37 | 1.54 | 1.33 | 1.13 |
| $JB_n$ | 5% | 3.51 | 4.21 | 4.48 | 4.87 | 5.04 | 4.96 | 4.92 | 4.88 |
| | 10% | 5.22 | 6.70 | 7.61 | 8.58 | 9.15 | 9.40 | 9.28 | 9.68 |
| | 1% | 1.11 | 1.05 | 0.97 | 1.08 | 1.11 | 1.03 | 1.08 | 1.01 |
| $IM_n^p$ | 5% | 5.14 | 5.15 | 4.97 | 4.85 | 5.06 | 5.15 | 5.27 | 5.03 |
| | 10% | 10.37 | 10.00 | 9.86 | 9.80 | 10.16 | 9.98 | 10.35 | 9.89 |
| | 1% | 1.09 | 1.15 | 1.06 | 1.21 | 1.05 | 1.25 | 1.03 | 0.98 |
| $JB_n^p$ | 5% | 4.73 | 5.09 | 4.98 | 5.30 | 5.35 | 5.24 | 5.05 | 4.96 |
| | 10% | 9.91 | 10.16 | 9.94 | 9.98 | 10.23 | 10.45 | 10.12 | 10.06 |

Table 2: LEVELS OF THE TEST STATISTICS. Number of Repetitions: 20,000. Number of Bootstrap Repetitions: 1,000. MODEL: $Y_t = \mathbf{X}_t'\boldsymbol{\beta} + U_t$ and $U_t \sim N(0, \sigma^2)$. DGP: $Y_t = \mathbf{X}_t'\boldsymbol{\beta}_* + U_t$ and $U_t \sim N(0, \sigma_*^2)$, $\mathbf{X}_t = (1, X_t)'$, and $X_t \sim N(0,1)$.

| Statistics \ $n$ | 50 | 100 | 200 | 400 | 600 | 800 | 1,000 | 2,000 |
|---|---|---|---|---|---|---|---|---|
| | | | | $U_t\vert\mathbf{X}_t \sim N(0,\exp(\mathbf{X}_t'\boldsymbol{\beta}_*))$ | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 53.68 | 80.76 | 97.48 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 87.62 | 99.64 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 89.29 | 99.74 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $IM_n^p$ | 22.22 | 88.64 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $JB_n^p$ | 50.82 | 78.94 | 96.98 | 99.98 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | | $U_t\vert\mathbf{X}_t \sim N(0,\exp(2\mathbf{X}_t'\boldsymbol{\beta}_*))$ | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 96.18 | 99.96 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 99.94 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 99.98 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $IM_n^p$ | 80.72 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $JB_n^p$ | 94.44 | 99.92 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | | $U_t\vert\mathbf{X}_t \sim 0.5\cdot N(-1,1)+0.5\cdot N(1,1)$ | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 9.42 | 14.90 | 25.32 | 44.92 | 63.18 | 76.76 | 85.14 | 99.32 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 1.04 | 1.32 | 3.98 | 23.78 | 48.18 | 70.44 | 84.76 | 99.70 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 1.60 | 2.64 | 8.94 | 32.16 | 55.24 | 75.56 | 87.38 | 99.76 |
| $IM_n^p$ | 16.10 | 25.90 | 43.72 | 67.08 | 81.20 | 90.20 | 95.20 | 99.86 |
| $JB_n^p$ | 0.88 | 1.04 | 7.80 | 36.84 | 65.28 | 83.28 | 92.80 | 99.94 |
| | | | | $U_t\vert\mathbf{X}_t \sim 0.5\cdot N(-1.5,1)+0.5\cdot N(1.5,1)$ | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 33.46 | 59.90 | 88.34 | 99.62 | 99.98 | 100.0 | 100.0 | 100.0 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 1.84 | 15.16 | 74.88 | 99.76 | 100.0 | 100.0 | 100.0 | 100.0 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 7.40 | 34.86 | 85.60 | 99.88 | 100.0 | 100.0 | 100.0 | 100.0 |
| $IM_n^p$ | 51.86 | 82.42 | 97.42 | 99.98 | 100.0 | 100.0 | 100.0 | 100.0 |
| $JB_n^p$ | 0.58 | 20.40 | 86.90 | 99.96 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | | $U_t\vert\mathbf{X}_t \sim t_{30}$ | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 5.40 | 7.40 | 8.94 | 11.96 | 16.44 | 18.00 | 20.86 | 34.70 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 8.14 | 10.08 | 12.24 | 15.54 | 20.58 | 22.26 | 24.90 | 39.98 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 8.18 | 9.50 | 11.66 | 14.80 | 19.44 | 21.34 | 23.92 | 38.86 |
| $IM_n^p$ | 4.14 | 3.32 | 2.72 | 2.28 | 1.86 | 2.24 | 2.84 | 5.78 |
| $JB_n^p$ | 8.68 | 10.76 | 14.02 | 18.04 | 23.14 | 25.34 | 28.46 | 45.82 |
| | | | | $U_t\vert\mathbf{X}_t \sim t_{20}$ | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 2.32 | 6.40 | 9.14 | 13.80 | 17.70 | 22.92 | 26.66 | 48.82 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 0.08 | 0.20 | 1.04 | 4.20 | 8.80 | 14.70 | 20.44 | 45.00 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 0.06 | 0.16 | 0.56 | 2.66 | 6.30 | 11.86 | 17.24 | 42.10 |
| $IM_n^p$ | 4.00 | 2.76 | 2.12 | 1.62 | 1.70 | 2.58 | 3.46 | 14.86 |
| $JB_n^p$ | 10.68 | 15.34 | 19.90 | 29.78 | 37.64 | 44.18 | 51.84 | 74.14 |

Table 3: POWERS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). NUMBER OF REPETITIONS: 5,000. NUMBER OF BOOTSTRAP REPETITIONS: 1,000. MODEL: $Y_t = \mathbf{X}_t'\boldsymbol{\beta} + U_t$ and $U_t \sim N(0,\sigma^2)$. DGP: $Y_t = \mathbf{X}_t'\boldsymbol{\beta}_* + U_t$, $\mathbf{X}_t = (1, X_t)'$, and $X_t \sim N(0,1)$.

| $\mathbf{X}_t$ | Alternatives \ $n$ | 50 | 100 | 200 | 400 | 600 | 800 | 1,000 | 2,000 |
|---|---|---|---|---|---|---|---|---|---|
| $(1,X_t)'$ | ALT 1 | ● | ● | ○ | △ | △ | △ | △ | △ |
| | ALT 2 | ● | ● | △ | △ | △ | △ | △ | △ |
| | ALT 3 | ● | ● | ● | ● | ● | ● | ○ | ○ |
| | ALT 4 | ● | ● | ● | ○ | ○ | △ | △ | △ |
| | ALT 5 | ○ | ● | ○ | ○ | ○ | ○ | ○ | ○ |
| | ALT 6 | ● | ● | ● | ● | ● | ● | ● | ● |

Table 4: PREDICTION OF THE MOST POWERFUL TEST. MODEL: $Y_t = \mathbf{X}_t'\boldsymbol{\beta} + U_t$ and $U_t \sim N(0,\sigma^2)$. This table shows the results of the most powerful tests predicted by Table 1. "○" and "●" indicate that the prediction is correct and wrong, respectively. "△" indicates that the prediction of the most powerful test cannot be properly made because $\widehat{\mathscr{B}}_n^{(1,p)}$, $\widehat{\mathscr{B}}_n^{(2,p)}$, and $\widehat{\mathscr{B}}_n^{(3,p)}$ all reject the null.

| Statistics | Levels \ $n$ | 50 | 100 | 200 | 400 | 600 | 800 | 1,000 | 2,000 |
|---|---|---|---|---|---|---|---|---|---|
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 1% | 1.15 | 1.02 | 1.01 | 0.90 | 1.10 | 1.15 | 1.05 | 1.21 |
| | 5% | 5.12 | 5.00 | 4.89 | 5.45 | 5.16 | 5.38 | 4.99 | 5.35 |
| | 10% | 10.53 | 10.25 | 9.77 | 10.30 | 10.50 | 10.67 | 9.97 | 10.25 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 1% | 1.13 | 1.05 | 1.04 | 0.90 | 1.13 | 1.11 | 1.05 | 1.17 |
| | 5% | 5.15 | 4.92 | 4.88 | 5.40 | 5.11 | 5.44 | 4.96 | 5.17 |
| | 10% | 10.20 | 9.93 | 9.96 | 10.00 | 10.35 | 10.74 | 10.25 | 10.24 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 1% | 1.11 | 1.05 | 1.06 | 0.95 | 1.17 | 1.09 | 1.05 | 1.15 |
| | 5% | 5.20 | 5.03 | 4.82 | 5.45 | 5.19 | 5.53 | 4.93 | 5.20 |
| | 10% | 10.20 | 10.12 | 9.89 | 9.75 | 10.30 | 10.84 | 10.12 | 10.27 |
| $IM_n$ | 1% | 16.47 | 16.08 | 12.76 | 10.90 | 10.16 | 8.94 | 8.30 | 5.89 |
| | 5% | 33.79 | 31.29 | 26.67 | 22.55 | 20.09 | 18.05 | 17.01 | 13.26 |
| | 10% | 45.36 | 41.41 | 35.53 | 30.30 | 27.28 | 25.45 | 23.77 | 19.49 |
| $LM_n$ | 1% | 30.65 | 19.77 | 12.14 | 7.98 | 5.70 | 4.96 | 3.92 | 2.79 |
| | 5% | 35.54 | 24.99 | 17.24 | 12.64 | 9.84 | 9.24 | 8.08 | 7.01 |
| | 10% | 39.01 | 28.26 | 20.90 | 16.22 | 13.61 | 13.12 | 12.02 | 11.11 |
| $IM_n^p$ | 1% | 1.09 | 1.11 | 1.09 | 1.05 | 1.16 | 1.18 | 1.16 | 1.12 |
| | 5% | 5.04 | 4.96 | 4.95 | 4.95 | 5.11 | 5.37 | 5.33 | 5.33 |
| | 10% | 9.96 | 10.12 | 9.83 | 9.05 | 10.26 | 10.26 | 10.52 | 10.31 |
| $LM_n^p$ | 1% | 1.03 | 1.06 | 1.11 | 1.11 | 1.03 | 1.20 | 1.02 | 1.04 |
| | 5% | 5.38 | 4.91 | 5.09 | 5.53 | 5.17 | 5.21 | 4.98 | 5.33 |
| | 10% | 10.34 | 9.92 | 9.81 | 10.41 | 9.74 | 10.38 | 9.95 | 10.48 |

Table 5: LEVELS OF THE TEST STATISTICS. NUMBER OF REPETITIONS: 20,000. NUMBER OF BOOTSTRAP REPETITIONS: 1,000. MODEL: $Y_t|\mathbf{X}_t \sim \text{Exp}(\alpha \exp(\mathbf{X}_t'\boldsymbol{\beta}))$. DGP: $Y_t|\mathbf{X}_t \sim \text{Exp}(\alpha_* \exp(\mathbf{X}_t'\boldsymbol{\beta}_*))$, $\mathbf{X}_t = X_t$, and $X_t \sim N(0,1)$.

| Statistics \ $n$ | 50 | 100 | 200 | 400 | 600 | 800 | 1,000 | 2,000 |
|---|---|---|---|---|---|---|---|---|
| | | | | $\delta_t \sim$ DM(0.7370, 1.9296; 0.5) | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 15.52 | 28.12 | 48.46 | 74.44 | 88.94 | 95.72 | 98.16 | 99.98 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 20.84 | 30.38 | 47.60 | 71.60 | 86.86 | 94.40 | 97.56 | 100.0 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 17.72 | 28.00 | 45.88 | 76.48 | 86.34 | 94.20 | 97.52 | 100.0 |
| $IM_n^p$ | 0.92 | 0.54 | 0.72 | 6.44 | 24.98 | 53.44 | 75.84 | 99.96 |
| $LM_n^p$ | 0.80 | 0.20 | 0.02 | 0.02 | 1.24 | 17.42 | 60.20 | 100.0 |
| | | | | $\delta_t \sim$ Gamma(5,5) | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 21.14 | 37.32 | 62.04 | 87.14 | 96.42 | 98.64 | 99.48 | 100.0 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 27.10 | 42.38 | 63.82 | 86.72 | 95.94 | 98.72 | 99.58 | 100.0 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 24.44 | 39.80 | 62.20 | 86.20 | 95.76 | 98.62 | 99.54 | 100.0 |
| $IM_n^p$ | 1.74 | 1.32 | 1.80 | 8.88 | 26.56 | 51.54 | 73.38 | 99.20 |
| $LM_n^p$ | 0.68 | 0.26 | 0.02 | 0.00 | 0.08 | 2.44 | 16.22 | 87.58 |
| | | | | $\delta_t \sim$ LN($-$ln(1.2)/2, ln(1.2)) | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 14.06 | 14.06 | 41.68 | 67.20 | 84.46 | 90.58 | 95.72 | 99.98 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 19.46 | 19.46 | 43.82 | 66.54 | 83.12 | 89.62 | 94.90 | 99.90 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 17.08 | 17.08 | 42.36 | 65.10 | 82.44 | 89.26 | 94.72 | 99.90 |
| $IM_n^p$ | 1.68 | 1.68 | 0.94 | 2.66 | 10.80 | 24.68 | 44.64 | 96.73 |
| $LM_n^p$ | 0.96 | 0.42 | 0.18 | 0.00 | 0.08 | 1.98 | 16.24 | 98.68 |
| | | | | $\delta_t \sim U(0.30053, 2.3661)$ | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 32.96 | 53.62 | 82.44 | 97.34 | 99.50 | 99.96 | 99.98 | 100.0 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 40.72 | 57.30 | 82.76 | 97.38 | 99.60 | 99.98 | 100.0 | 100.0 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 36.08 | 54.70 | 81.54 | 97.14 | 99.58 | 99.98 | 100.0 | 100.0 |
| $IM_n^p$ | 1.12 | 1.94 | 5.72 | 32.98 | 71.00 | 91.02 | 97.90 | 100.0 |
| $LM_n^p$ | 0.02 | 0.06 | 0.00 | 0.00 | 0.38 | 6.68 | 31.44 | 97.76 |
| | | | | $\delta_t \sim U(1, 5/3)$ | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 4.78 | 5.54 | 6.36 | 7.56 | 8.44 | 10.00 | 9.62 | 13.90 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 5.62 | 7.06 | 7.52 | 8.20 | 8.74 | 10.20 | 10.02 | 12.84 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 5.32 | 6.68 | 7.20 | 7.92 | 8.52 | 9.88 | 9.68 | 12.50 |
| $IM_n^p$ | 4.32 | 3.76 | 3.34 | 3.44 | 2.84 | 2.78 | 3.00 | 2.58 |
| $LM_n^p$ | 4.10 | 4.00 | 2.84 | 2.20 | 1.82 | 1.48 | 1.08 | 2.54 |

Table 6: POWERS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). NUMBER OF REPETITIONS: 5,000. NUMBER OF BOOTSTRAP REPETITIONS: 1,000. MODEL: $Y_t|\mathbf{X}_t \sim \text{Exp}(\alpha \exp(\mathbf{X}_t'\boldsymbol{\beta}))$. DGP: $Y_t|(\delta_t, \mathbf{X}_t) \sim \text{Exp}(\delta_t \exp(\mathbf{X}_t'\boldsymbol{\beta}_*))$, $\mathbf{X}_t = X_t$, and $X_t \sim N(0,1)$.

| $\mathbf{X}_t$ | Alternatives \ $n$ | 50 | 100 | 200 | 400 | 600 | 800 | 1,000 | 2,000 |
|---|---|---|---|---|---|---|---|---|---|
| $(1, X_t)'$ | ALT 1 | ○ | ○ | ● | ● | ● | ● | ● | ○ |
| | ALT 2 | ○ | ● | ○ | ● | ● | ○ | ○ | △ |
| | ALT 3 | ○ | ○ | ○ | ● | ● | ● | ● | ○ |
| | ALT 4 | ○ | ○ | ○ | ○ | ○ | ○ | ○ | △ |
| | ALT 5 | ● | ● | ○ | ○ | ○ | ○ | ○ | ● |

Table 7: PREDICTION OF THE MOST POWERFUL TEST. MODEL: $Y_t|\mathbf{X}_t \sim \text{Exp}(\alpha \exp(\mathbf{X}_t'\boldsymbol{\beta}))$. This table shows the results of the most powerful tests predicted by Table 1. "○" and "●" indicate that the prediction is correct and wrong, respectively. "△" indicates that the prediction of the most powerful test cannot be properly made because $\widehat{\mathscr{B}}_n^{(1,p)}$, $\widehat{\mathscr{B}}_n^{(2,p)}$, and $\widehat{\mathscr{B}}_n^{(3,p)}$ all reject the null.

| Statistics | Levels \ $n$ | 50 | 100 | 200 | 300 | 400 | 500 | 600 | 700 |
|---|---|---|---|---|---|---|---|---|---|
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 1% | 0.99 | 1.00 | 1.26 | 1.29 | 1.20 | 1.09 | 1.16 | 1.38 |
| | 5% | 2.95 | 4.20 | 5.07 | 4.98 | 5.20 | 5.13 | 4.87 | 5.11 |
| | 10% | 6.85 | 8.61 | 9.98 | 9.80 | 10.40 | 9.64 | 9.79 | 10.20 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 1% | 0.96 | 1.07 | 1.20 | 1.14 | 1.28 | 1.18 | 1.08 | 1.27 |
| | 5% | 3.89 | 4.36 | 4.96 | 4.87 | 5.10 | 5.01 | 4.82 | 5.11 |
| | 10% | 7.21 | 8.67 | 9.62 | 9.66 | 10.41 | 9.52 | 9.42 | 10.10 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 1% | 0.79 | 0.92 | 1.19 | 1.16 | 1.31 | 1.14 | 1.10 | 1.25 |
| | 5% | 3.04 | 4.20 | 5.02 | 4.74 | 5.11 | 4.84 | 4.76 | 5.06 |
| | 10% | 6.51 | 8.47 | 9.49 | 9.53 | 10.38 | 9.68 | 9.36 | 10.06 |
| $IM_n$ | 1% | 35.98 | 28.89 | 21.72 | 1.68 | 1.52 | 12.13 | 11.34 | 10.51 |
| | 5% | 53.38 | 42.48 | 31.94 | 27.03 | 25.30 | 20.87 | 19.50 | 18.51 |
| | 10% | 61.85 | 49.87 | 39.17 | 33.91 | 31.63 | 26.94 | 25.66 | 25.12 |
| $IM_n^p$ | 1% | 0.58 | 0.80 | 1.10 | 1.10 | 1.06 | 1.01 | 1.26 | 1.14 |
| | 5% | 3.54 | 4.52 | 5.00 | 5.19 | 5.37 | 4.89 | 5.35 | 5.27 |
| | 10% | 8.01 | 9.40 | 9.96 | 9.88 | 10.60 | 9.50 | 9.94 | 10.41 |

Table 8: LEVELS OF THE TEST STATISTICS. NUMBER OF REPETITIONS: 10,000. NUMBER OF BOOTSTRAP REPETITIONS: 500. MODEL: $Y_t|\mathbf{X}_t \sim \text{Probit}(\mathbf{X}_t'\boldsymbol{\beta})$. DGP: $Y_t|\mathbf{X}_t \sim \text{Probit}(\mathbf{X}_t'\boldsymbol{\beta}_*)$, $\mathbf{X}_t = (1, X_t)'$, and $X_t \sim N(0,1)$.

| Statistics \ $n$ | 50 | 100 | 200 | 300 | 400 | 500 | 600 | 700 |
|---|---|---|---|---|---|---|---|---|
| | | | | $Y_t\|\mathbf{X}_t \sim \text{Probit}(\mathbf{X}_t'\boldsymbol{\beta}_*/\exp(0.5\mathbf{X}_t'\boldsymbol{\beta}_*))$ | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 9.03 | 15.40 | 22.13 | 27.10 | 34.93 | 38.83 | 43.93 | 47.73 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 15.86 | 38.90 | 71.23 | 89.86 | 97.06 | 99.16 | 99.90 | 99.96 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 14.03 | 37.93 | 71.16 | 90.06 | 97.10 | 99.30 | 99.90 | 99.96 |
| $IM_n^p$ | 6.93 | 27.80 | 67.93 | 90.00 | 97.40 | 99.63 | 99.93 | 100.0 |
| | | | | $Y_t\|\mathbf{X}_t \sim \text{Probit}((\mathbf{X}_t'\boldsymbol{\beta}_*)^2)$ | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 25.56 | 70.00 | 92.83 | 97.96 | 99.40 | 99.70 | 99.93 | 99.96 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 36.46 | 93.50 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 45.53 | 95.43 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $IM_n^p$ | 63.96 | 97.96 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | | | | $Y_t\|\mathbf{X}_t \sim \text{Logit}(\mathbf{X}_t'\boldsymbol{\beta}_*)$ | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 3.36 | 4.40 | 5.20 | 6.03 | 7.13 | 7.70 | 7.80 | 7.60 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 3.70 | 4.30 | 5.20 | 6.50 | 6.93 | 7.20 | 7.73 | 8.26 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 2.96 | 3.93 | 4.90 | 6.56 | 6.90 | 7.06 | 7.63 | 7.90 |
| $IM_n^p$ | 1.53 | 3.43 | 3.60 | 4.33 | 3.73 | 3.33 | 4.13 | 3.80 |

Table 9: POWERS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). NUMBER OF REPETITIONS: 3,000. NUMBER OF BOOTSTRAP REPETITIONS: 500. MODEL: $Y_t|\mathbf{X}_t \sim \text{Probit}(\mathbf{X}_t'\boldsymbol{\beta})$. DGP: $\mathbf{X}_t = (1, X_t)'$ and $X_t \sim N(0,1)$.

| $\mathbf{X}_t$ | Alternatives \ $n$ | 50 | 100 | 200 | 300 | 400 | 500 | 600 | 700 |
|---|---|---|---|---|---|---|---|---|---|
| $(1, X_t)'$ | ALT 1 | ○ | ○ | ○ | ● | ● | ● | ○ | ○ |
| | ALT 2 | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| | ALT 3 | ○ | ○ | ○ | ● | ● | ● | ● | ○ |

Table 10: PREDICTION OF THE MOST POWERFUL TEST. MODEL: $Y_t|\mathbf{X}_t \sim \text{Probit}(\mathbf{X}_t'\boldsymbol{\beta})$. This table shows the results of the most powerful tests predicted by Table 1. "○" and "●" indicate that the prediction is correct and wrong, respectively.

| Statistics | Levels \ $n$ | 50 | 100 | 200 | 300 | 400 |
|---|---|---|---|---|---|---|
| | 1% | 0.92 | 1.32 | 1.15 | 1.20 | 1.04 |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 5% | 4.90 | 5.25 | 4.99 | 5.45 | 4.97 |
| | 10% | 10.07 | 10.49 | 10.18 | 10.32 | 10.05 |
| | 1% | 0.99 | 1.17 | 1.03 | 1.22 | 0.98 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 5% | 4.79 | 4.93 | 4.87 | 4.68 | 5.06 |
| | 10% | 9.31 | 9.57 | 9.75 | 9.62 | 10.37 |
| | 1% | 0.95 | 1.27 | 1.07 | 1.18 | 1.03 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 5% | 4.79 | 4.84 | 4.92 | 4.83 | 5.10 |
| | 10% | 9.31 | 9.73 | 9.65 | 9.80 | 10.30 |
| | 1% | 52.43 | 41.72 | 30.54 | 24.60 | 20.66 |
| $IM_n$ | 5% | 69.66 | 56.86 | 44.81 | 37.67 | 32.94 |
| | 10% | 77.01 | 65.59 | 52.92 | 45.76 | 41.25 |
| | 1% | 1.09 | 1.16 | 1.21 | 1.39 | 1.20 |
| $IM_n^p$ | 5% | 4.65 | 5.84 | 6.00 | 5.88 | 5.21 |
| | 10% | 10.15 | 11.86 | 11.35 | 11.29 | 9.53 |

Table 11: LEVELS OF THE TEST STATISTICS. NUMBER OF REPETITIONS: 10,000. NUMBER OF BOOTSTRAP REPETITIONS: 500. MODEL: $Y_t = \max[0, \mathbf{X}_t'\boldsymbol{\beta} + U_t]$ and $U_t \sim N(0, \sigma^2)$. DGP: $Y_t = \max[0, \mathbf{X}_t'\boldsymbol{\beta}_* + U_t]$, $\mathbf{X}_t = (1, X_t)'$, and $(X_t, U_t)' \sim N(\mathbf{0}, \mathbf{I}_2)$.

| Statistics \ $n$ | 50 | 100 | 200 | 300 | 400 |
|---|---|---|---|---|---|
| $Y_t = \max[0, \mathbf{X}_t'\boldsymbol{\beta}_* + U_t]$ and $U_t \vert \mathbf{X}_t \sim N(0, \exp(0.5\mathbf{X}_t'\boldsymbol{\beta}_*))$ | | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 32.30 | 50.63 | 75.26 | 88.73 | 94.46 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 62.96 | 89.66 | 99.73 | 99.93 | 100.0 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 64.80 | 91.43 | 99.73 | 100.0 | 100.0 |
| $IM_n^p$ | 8.33 | 24.00 | 92.76 | 99.90 | 100.0 |
| $Y_t = \max[0, (\mathbf{X}_t'\boldsymbol{\beta}_*)^2 + U_t]$ and $U_t \vert \mathbf{X}_t \sim N(0, 1)$ | | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 81.76 | 98.36 | 99.96 | 100.0 | 100.0 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 80.03 | 98.56 | 99.96 | 100.0 | 100.0 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 76.13 | 97.90 | 99.96 | 100.0 | 100.0 |
| $IM_n^p$ | 37.43 | 62.23 | 87.63 | 96.10 | 98.53 |
| $Y_t = \max[0, \mathbf{X}_t'\boldsymbol{\beta}_* + U_t]$ and $U_t \vert \mathbf{X}_t \sim t_{30}$ | | | | | |
| $\widehat{\mathscr{B}}_n^{(1,p)}$ | 5.00 | 6.33 | 7.76 | 8.63 | 9.63 |
| $\widehat{\mathscr{B}}_n^{(2,p)}$ | 7.26 | 8.16 | 9.86 | 10.70 | 13.20 |
| $\widehat{\mathscr{B}}_n^{(3,p)}$ | 6.96 | 7.73 | 9.26 | 10.56 | 12.63 |
| $IM_n^p$ | 4.36 | 4.80 | 4.30 | 4.20 | 2.66 |

Table 12: POWERS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). NUMBER OF REPETITIONS: 3,000. NUMBER OF BOOTSTRAP REPETITIONS: 500. MODEL: $Y_t = \max[0, \mathbf{X}_t'\boldsymbol{\beta} + U_t]$ and $U_t \sim N(0, \sigma^2)$. DGP: $\mathbf{X}_t = (1, X_t)'$ and $X_t \sim N(0, 1)$.

| $\mathbf{X}_t$ | Alternatives \ $n$ | 50 | 100 | 200 | 300 | 400 |
|---|---|---|---|---|---|---|
| $(1, X_t)'$ | ALT 1 | ● | ● | ● | ● | ○ |
| | ALT 2 | ● | ○ | ○ | △ | △ |
| | ALT 3 | ● | ● | ○ | ○ | ○ |

Table 13: PREDICTION OF THE MOST POWERFUL TEST. MODEL: $Y_t = \max[0, \mathbf{X}_t'\boldsymbol{\beta} + U_t]$ and $U_t \sim N(0, \sigma^2)$. This table shows the results of the most powerful tests predicted by Table 1. "○" and "●" indicate that the prediction is correct and wrong, respectively. "△" indicates that the prediction of the most powerful test cannot be properly made because $\widehat{\mathscr{B}}_n^{(1,p)}$, $\widehat{\mathscr{B}}_n^{(2,p)}$, and $\widehat{\mathscr{B}}_n^{(3,p)}$ all reject the null.

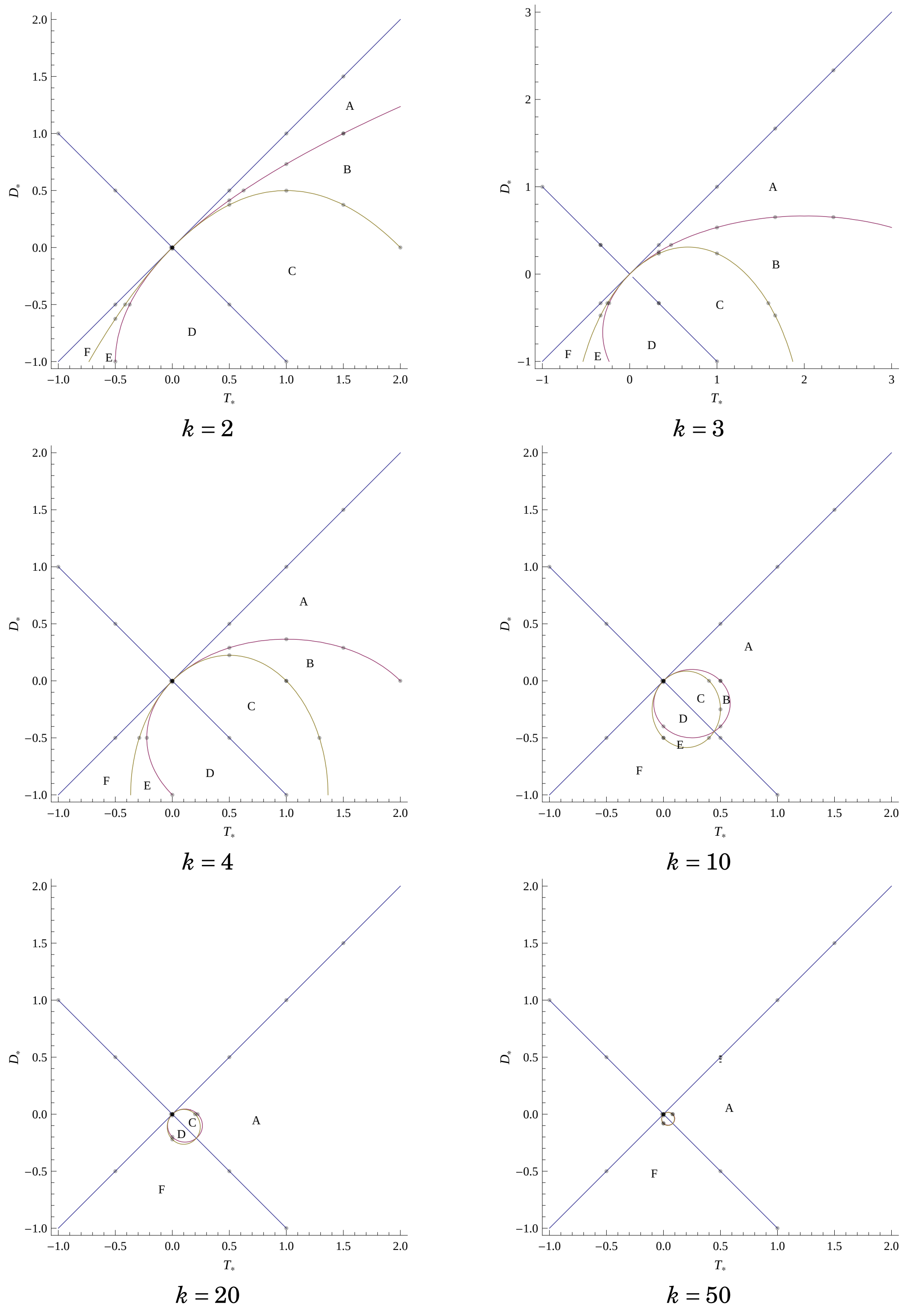


Figure 1: PARTITIONED SPACE OF $(T_*, D_*)$ . For $k = 2$, 3, 4, 10, 20, and 50, this figure shows the space of $(T_*, D_*)$ partitioned by the conditions A to F. For example, if $(T_*, D_*)$ belongs to the region indexed by A, we have that $\widehat{\mathscr{B}}_n^{(2)} \gtrsim \widehat{\mathscr{B}}_n^{(1)} \gtrsim \widehat{\mathscr{B}}_n^{(3)}$ in probability by Table 1. As another case, if $(T_*, D_*)$ belongs to the region indexed by B, we have that $\widehat{\mathscr{B}}_n^{(1)} \gtrsim \widehat{\mathscr{B}}_n^{(3)} \gtrsim \widehat{\mathscr{B}}_n^{(2)}$ in probability by Table 1. We also observe that the number of parameters ($k$) matters in partitioning the space of $(T_*, D_*)$. As $k$ tends to infinity, most space of $(T_*, D_*)$ is indexed by A and F.