Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers | Cowles Foundation

12-1-2014

Testing Equality of Covariance Matrices via Pythagorean Means

Jin Seo Cho

Peter C.B. Phillips

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

 Part of the Economics Commons

Recommended Citation

Cho, Jin Seo and Phillips, Peter C.B., "Testing Equality of Covariance Matrices via Pythagorean Means" (2014). *Cowles Foundation Discussion Papers*. 2384.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/2384

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

TESTING EQUALITY OF COVARIANCE MATRICES VIA PYTHAGOREAN MEANS

By

Jin Seo Cho and Peter C. B. Phillips

December 2014

COWLES FOUNDATION DISCUSSION PAPER NO. 1970



COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Testing Equality of Covariance Matrices via Pythagorean Means

JIN SEO CHO

School of Economics

Yonsei University

50 Yonsei-ro, , Seodaemun-gu, Seoul, Korea

Email: jinseocho@yonsei.com

PETER C.B. PHILLIPS

Yale University

University of Auckland

Singapore Management University &

University of Southampton

October 7, 2014

Abstract

We provide a new test for equality of covariance matrices that leads to a convenient mechanism for testing specification using the information matrix equality. The test relies on a new characterization of equality between two k dimensional positive-definite matrices $\mathbf{A}$ and $\mathbf{B}$: the traces of $\mathbf{AB}^{-1}$ and $\mathbf{BA}^{-1}$ are equal to k if and only if $\mathbf{A} = \mathbf{B}$. Using this criterion, we introduce a class of omnibus test statistics for equality of covariance matrices and examine their null, local, and global approximations under some mild regularity conditions. Monte Carlo experiments are conducted to explore the performance characteristics of the test criteria and provide comparisons with existing tests under the null hypothesis and local and global alternatives. The tests are applied to the classic empirical models for voting turnout investigated by Wolfinger and Rosenstone (1980) and Nagler (1991, 1994). Our tests show that all classic models for the 1984 presidential voting turnout are misspecified in the sense that the information matrix equality fails.

Key Words: Matrix equality; Trace; Determinant; Arithmetic mean; Geometric mean; Harmonic mean; Information matrix; Eigenvalues; Parametric bootstrap.

Subject Classification: C01, C12, C52

Acknowledgements: The authors acknowledge helpful discussions with Jin Lee, Yoon-Jae Whang, and other participants at the KEA conference (Korea University). Phillips acknowledges research support from the NSF under grant No. SES 12-58258.

1 Introduction

Comparing covariance matrices and testing the equivalence of two positive definite matrices have attracted substantial past attention in multivariate analysis (e.g., Muirhead, 1982). In statistics, the earliest work was by Wilks (1932) on independent multivariate normal samples and used a likelihood ratio test, which led to a substantial subsequent literature. In econometrics, the problem of covariance matrix equality arises naturally in several modeling contexts that are important in applications. For example, the asymptotic distribution of the maximum likelihood (ML) estimator is characterized by the usual information matrix equality. On the other hand, the information matrix equality does not hold for the qausi-ML (QML) estimator. As another example, least squares (LS) and generalized method of moments (GMM) estimators have relatively simple covariance matrix structures except when heteroskedasticity or autocorrelation is present. The simple covariance matrix structure is then delivered by the proportional equality of two positive-definite matrices (viz., $X'X$ and $X'\Sigma X$ in the usual regression notation).

These material econometric interests have led to much literature on covariance matrix equality testing, with special attention being given to the information matrix equality (*e.g.*, White, 1982; 1994; Bera, 1986; Hall, 1987; Orme, 1988, 1990; Chesher and Spady, 1991; Horowitz 1994; Dhaene and Hoorelbeke, 2004; and Golden, Henley, White, and Kashner, 2013), although work is not limited to that setting alone (*e.g.*, Bera, 1991). Much of this past work arises from the desire for an omnibus test without level distortion and with high power. The problem in size control is simply stated. For two general $k \times k$ positive-definite matrices $\mathbf{A}$ and $\mathbf{B}$ say, testing *every* pair of corresponding elements in $\mathbf{A}$ and $\mathbf{B}$ generates enormous level distortions for the tests even with moderately sized k.

The goal of the current study is to develop simple and straightforward omnibus tests for the equality of two positive-definite matrices. The approach that we use here has an antecedent in Cho and White (2014; CW, henceforth). CW provided omnibus tests of matrix equality by using the fact that the conditions $\mathrm{tr}[\mathbf{B}\mathbf{A}^{-1}] = k$ and $\det[\mathbf{B}\mathbf{A}^{-1}] = 1$ are necessary and sufficient for $\mathbf{A} = \mathbf{B}$. Our starting point is to extend this condition with another, even simpler, characterization of equality that enables a new class of omnibus tests for equality that have little size distortion and comparable powers to other tests. We also seek to clarify the interrelationships among the many tests that are now available and examined in the current study.

Our test statistics are developed to supplement the tests in CW in the following ways. First, we provide another characterization for the equality of two positive-definite matrices. We show that the simple dual conditions $\mathrm{tr}[\mathbf{A}\mathbf{B}^{-1}] = k$ and $\mathrm{tr}[\mathbf{B}\mathbf{A}^{-1}] = k$ are also necessary and sufficient for $\mathbf{A} = \mathbf{B}$. This characterization is made by noting that $k^{-1}\mathrm{tr}[\mathbf{B}\mathbf{A}^{-1}]$ and $k\mathrm{tr}[\mathbf{A}\mathbf{B}^{-1}]^{-1}$ are the arithmetic and harmonic

means of the eigenvalues of $\mathbf{BA}^{-1}$, respectively, and these means are equal if and only if all eigenvalues are identical. Under the given conditions, all eigenvalues are unity, implying that $\mathbf{BA}^{-1} = \mathbf{I}$. Furthermore, two equal positive-definite matrices can also be characterized by combining this characterization and that given in CW, viz. $\mathrm{tr}[\mathbf{AB}^{-1}] = k$ and $\det[\mathbf{BA}^{-1}] = 1$ if and only if $\mathbf{A} = \mathbf{B}$. Using these new characterizations, we introduce several useful omnibus tests.

Second, our tests are more powerful than those in CW in alternative directions that are different from those for the tests in CW. For this analysis, we examine the approximations of the tests under global alternatives and single out the factors that lead to the consistency property of the tests. Third, we compare the local power properties of the tests and examine them for equivalence. By this process, we can group the tests according to their local powers.

Finally, we apply our tests to empirical data on voting turnout. In the political economy and political science literature, an important research question involves identifying factors that determine presidential voting turnouts (*e.g.*, Wolfinger and Rosenstone, 1980; Feddersen and Pesendorfer, 1996; Nagler, 1991, 1994; Bénabou 2000; Besley and Case, 2003; Berry, DeMeritt, and Esarey, 2010, among many others). In particular, Wolfinger and Rosenstone's (1980) classic study examines the interaction effects between education and registry requirement to the voting turnout by estimating a probit model under ML. Nagler (1994) further extends their results by estimating a scobit model. We reexamine their models and empirically test whether their models are correctly specified for ML estimation. For this purpose, we use the 1984 presidential election data of the US that are provided by Altman and McDonald (2003).

The plan of this paper is as follows. Section 2 provides a basic lemma characterizing equality between two positive-definite matrices. Section 3 motivates and defines the test statistics employed, and develops asymptotic theory under the null and alternative hypotheses. Simulation results are reported in Section 4. We focus on linear normal and linear probit regression models and test the information matrix equality in these two frameworks. The empirical application is given in Section 5. Mathematical proofs are collected in the Appendix.

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

2 A Basic Lemma and Its Testing Implications

We proceed first withh the following lemma that characterizes the equality of two positive-definite matrices.

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*
(i) $\mathrm{tr}[\mathbf{D}] = k$ *and* $\mathrm{tr}[\mathbf{D}^{-1}] = k$*, where* $\mathbf{D} := \mathbf{B}\mathbf{A}^{-1}$*; or*
(ii) $\det[\mathbf{D}] = 1$ *and* $\mathrm{tr}[\mathbf{D}^{-1}] = k$*.* □

To our knowledge and somewhat surprisingly given its simplicity, Lemma 1 is new to the literature and is proved in the Appendix. Briefly, part (i) follows because the arithmetic mean of positive numbers is identical to their harmonic mean, if and only if all of the positive numbers are identical. Since $k^{-1}\mathrm{tr}[\mathbf{D}]$ is the arithmetic mean of the eigenvalues of $\mathbf{D}$, and $k^{-1}\mathrm{tr}[\mathbf{D}^{-1}]$ is the inverse of the harmonic mean of the same eigenvalues, we have $\mathbf{D} = \mathbf{I}$, if and only if all the eigenvalues are identical to unity, which implies that $\mathbf{A} = \mathbf{B}$.

The characterization in Lemma 1(*i*) can also be associated with a convexity property of the trace operator. Note that $\phi(\cdot) := \mathrm{tr}[(\cdot)^{-1}] + \mathrm{tr}[\cdot]$ is a convex function on the space of $k \times k$ positive-definite matrices (*e.g.*, Bernstein, 2005, p. 283) and is also bounded from below by $2k$ (*e.g.*, Abadir and Magnus, 2005, p.338). The lower bound is achieved if and only if the argument of $\phi(\cdot)$ is $\mathbf{I}$.

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

Both Lemma 1 and the characterization in CW rely on fundamental properties of the Pythagorean (harmonic, geometric, and arithmetic) means of positive numbers:

$$\text{Harmonic Mean} \;\leq\; \text{Geometric mean} \;\leq\; \text{Arithmetic mean.} \tag{1}$$

All three means are identical if the positive numbers are identical. Lemma 1(*i*) is obtained by interrelating the harmonic mean with the arithmetic mean, and CW links the geometric mean to the arithmetic mean for their characterization. Lemma 1(*ii*) also associates the harmonic and geometric means for the equality.

We now exploit Lemma 1 to test the equality of two positive-definite matrices. Lemma 1(*i*) is our first focus. Let $T := k^{-1}\mathrm{tr}[\mathbf{D}] - 1$, $H := k\mathrm{tr}[\mathbf{D}^{-1}]^{-1} - 1$, and $C := k^{-1}\mathrm{tr}[\mathbf{D}] - k\mathrm{tr}[\mathbf{D}^{-1}]^{-1}$ for notational

[1]The prior literature has separately examined the determinant- and trace-based tests. For example, Pillai and Nagarsenker (1972) and Das Gupta and Giri (1973) investigated determinant ratio tests for equal covariance matrices, and Roy (1953), Pillai and Jayachandran (1968), and Nagao (1973, 1974) developed trace-based tests and compared their performance to that of determinant-based tests.

simplicity. Note that if any two of T, H, and C equal to zero, the remaining one is also zero. Therefore, Lemma 1(*i*) holds if and only if any two of T, H, and C equal to zero. This implies that the equality of two positive-definite matrices can be tested by testing one of the following hypotheses:

$$\begin{aligned}
\mathcal{H}_0^{(1)}: &\quad T=0 \text{ and } H=0 \quad \text{vs.} \quad \mathcal{H}_1^{(1)}: T\neq 0 \text{ or } H\neq 0;\\
\mathcal{H}_0^{(2)}: &\quad T=0 \text{ and } C=0 \quad \text{vs.} \quad \mathcal{H}_1^{(2)}: T\neq 0 \text{ or } C\neq 0;\\
\mathcal{H}_0^{(3)}: &\quad H=0 \text{ and } C=0 \quad \text{vs.} \quad \mathcal{H}_1^{(3)}: H\neq 0 \text{ or } C\neq 0.
\end{aligned}$$

Similarly, we can exploit Lemma 1(*ii*) and for this, let $D := \det[\mathbf{D}]^{1/k} - 1$ and $G := \det[\mathbf{D}]^{1/k} - k\mathrm{tr}[\mathbf{D}^{-1}]^{-1}$. If any two of D, H, and G are zero, the remaining one is zero, so that Lemma 1(*ii*) holds if and only if any two of them are zero. Hence, we construct the corresponding hypotheses as

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

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

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

All these 9 hypothesis systems are equivalent systems of hypotheses to the simple null $\mathcal{H}_0 : \mathbf{A} = \mathbf{B}$ versus the alternative $\mathcal{H}_1 : \mathbf{A} \neq \mathbf{B}$.

3 Test Statistics and Their Asymptotic Expansions

This section introduces the test statistics and examine their asymptotic expansions under the null, alternative, and local alternative. We also supplement the test statistics considered in CW.

3.1 Definitions of Test Statistics and Asymptotic Approximations

We introduce testing environments by supposing that the previously defined $\mathbf{A}$ and $\mathbf{B}$ are in fact parameterized as $\mathbf{A} \equiv \mathbf{A}(\boldsymbol{\theta}_*)$ and $\mathbf{B} \equiv \mathbf{B}(\boldsymbol{\theta}_*)$, respectively, where both $\mathbf{A}(\cdot)$ and $\mathbf{B}(\cdot)$ are defined on $\boldsymbol{\Theta} \in \mathbb{R}^{\ell\times\ell}$, and

$\boldsymbol{\theta}_* \in \boldsymbol{\Theta}$ is an unknown parameter. We further suppose that $\mathbf{A}_n := \mathbf{A}_n(\boldsymbol{\theta}_*)$ and $\mathbf{B}_n := \mathbf{B}_n(\boldsymbol{\theta}_*)$ estimate $\mathbf{A}(\boldsymbol{\theta}_*)$ and $\mathbf{B}(\boldsymbol{\theta}_*)$ consistently, where $\mathbf{A}_n(\cdot)$ and $\mathbf{B}_n(\cdot)$ are consistent for $\mathbf{A}(\cdot)$ and $\mathbf{B}(\cdot)$ uniformly on $\boldsymbol{\Theta}$ and uniformly positive definite almost surely on $\boldsymbol{\Theta}$ for large enough n. Therefore, $\mathbf{D}_n := \mathbf{B}_n\mathbf{A}_n^{-1}$ and $\mathbf{D}_n^{-1}$ consistently estimate $\mathbf{D}$ and $\mathbf{D}^{-1}$, respectively. Here, $\mathbf{D}$ is estimated by a two-step estimation procedure. Specifically, the unknown parameter $\boldsymbol{\theta}_*$ is consistently estimated by an estimator $\widehat{\boldsymbol{\theta}}_n$, so that $\widehat{\mathbf{A}}_n := \mathbf{A}_n(\widehat{\boldsymbol{\theta}}_n)$ and $\widehat{\mathbf{B}}_n := \mathbf{B}_n(\widehat{\boldsymbol{\theta}}_n)$ are consistent for $\mathbf{A}(\boldsymbol{\theta}_*)$ and $\mathbf{B}(\boldsymbol{\theta}_*)$, respectively. Therefore, $\widehat{\mathbf{D}}_n := \widehat{\mathbf{B}}_n\widehat{\mathbf{A}}_n^{-1}$ and $\widehat{\mathbf{D}}_n^{-1} = \widehat{\mathbf{A}}_n\widehat{\mathbf{B}}_n^{-1}$ are also consistent for $\mathbf{D}$ and $\mathbf{D}^{-1}$, respectively. To reduce notational clutter, we simply indicate the influence of $\boldsymbol{\theta}_*$ on these matrices by letting

$$\mathbf{A}_* := \mathbf{A}(\boldsymbol{\theta}_*), \quad \mathbf{B}_* := \mathbf{B}(\boldsymbol{\theta}_*), \quad \mathbf{D}_* := \mathbf{B}_*\mathbf{A}_*^{-1}.$$

Similarly, let $T_* := k^{-1}\mathrm{tr}[\mathbf{D}_*] - 1$, $H_* := k/\mathrm{tr}[\mathbf{D}_*^{-1}] - 1$, $D_* := \det[\mathbf{D}_*]^{1/k} - 1$, $C_* := T_* - H_*$, $G_* := D_* - H_*$, and $S_* := T_* - D_*$. When these matrices are estimated using $\widehat{\mathbf{A}}_n$ and $\widehat{\mathbf{B}}_n$, we denote the resulting statistics as $\widehat{T}_n := k^{-1}\mathrm{tr}[\widehat{\mathbf{D}}_n] - 1$, $\widehat{H}_n := k/\mathrm{tr}[\widehat{\mathbf{D}}_n^{-1}] - 1$, $\widehat{D}_n := \det[\widehat{\mathbf{D}}_n]^{1/k} - 1$, $\widehat{C}_n := \widehat{T}_n - \widehat{H}_n$, $\widehat{G}_n := \widehat{D}_n - \widehat{H}_n$, and $\widehat{S}_n := \widehat{T}_n - \widehat{D}_n$. All these statistics, which form the base elements of the tests given below, are dependent upon $\widehat{\boldsymbol{\theta}}_n$.

Before defining the tests, we provide the following regularity conditions.

Assumption A (Cho and White, 2014)**.** *(i)* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a complete probability space;*
(ii) $\boldsymbol{\Theta} \subset \mathbb{R}^\ell$ *is a compact convex set with non-empty interior and* $\ell \in \mathbb{N}$;
(iii) a sequence of measurable mappings $\{\widehat{\boldsymbol{\theta}}_n : \Omega \mapsto \boldsymbol{\Theta}\}$ *is consistent for a unique* $\boldsymbol{\theta}_* \in int(\boldsymbol{\Theta})$*;*
(iv) $\mathbf{A} : \boldsymbol{\Theta} \mapsto \mathbb{R}^{k\times k}$ *and* $\mathbf{B} : \boldsymbol{\Theta} \mapsto \mathbb{R}^{k\times k}$ *are in* $\mathcal{C}^{(2)}(\boldsymbol{\Theta})$*, and* $\mathbf{A}_*$ *and* $\mathbf{B}_*$ *are positive-definite;*
(v) $\mathbf{A}_n(\cdot)$ *and* $\mathbf{B}_n(\cdot)$ *are consistent for* $\mathbf{A}(\cdot)$ *and* $\mathbf{B}(\cdot)$*, respectively, uniformly on* $\boldsymbol{\Theta}$*;*
(vi) $\sqrt{n}[(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)', \mathrm{vech}[\mathbf{A}_n - \mathbf{A}_*]', \mathrm{vech}[\mathbf{B}_n - \mathbf{B}_*]']' = O_{\mathbb{P}}(1)$*;*
(vii) 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* $\boldsymbol{\Theta}$*; and*
(viii) for $j = 1, \ldots, \ell$*,* $\mathbf{H}_{j,n} = O_{\mathbb{P}}(n^{-1/2})$ *and* $\mathbf{G}_{j,n} = O_{\mathbb{P}}(n^{-1/2})$*, where* $\mathbf{H}_{j,n} := \mathbf{A}_*^{-1}\partial_j(\mathbf{A}_n - \mathbf{A}_*)$ *and* $\mathbf{G}_{j,n} := \mathbf{B}_*^{-1}\partial_j(\mathbf{B}_n - \mathbf{B}_*)$. □

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

Our omnibus tests are motivated by testing whether the critical quantities T_*, D_*, H_*, S_* C_*, and G_*, which we call the test base elements, equal zero. We first examine stochastic asymptotic representations of

consistent estimates of these quantities. For this purpose, we let

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

for notational simplicity, where $\mathbf{P}_n := \mathbf{W}_n - \mathbf{U}_n := \mathbf{B}_*^{-1}(\mathbf{B}_n - \mathbf{B}_*) - \mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_*)$, and for $j = 1, 2, \ldots, \ell$, $\mathbf{R}_{j,*} := \mathbf{B}_*^{-1}\partial_j \mathbf{B}_* - \mathbf{A}_*^{-1}\partial_j \mathbf{A}_*$. Note under Assumption A we have $\mathbf{L}_n = O_{\mathbb{P}}(n^{-1/2})$, $\mathbf{P}_n = O_{\mathbb{P}}(n^{-1/2})$, and for $j = 1, 2, \ldots, \ell$, $\mathbf{R}_{j,*} = O(1)$. These correspond with the definitions in CW. The following lemma provides the explicit asymptotic expansions.

Lemma 2. *Given Assumption A,*
(i) $\widehat{T}_n = T_* + k^{-1}\mathrm{tr}[\mathbf{L}_n \mathbf{A}_*^{-1}\mathbf{B}_*] + O_{\mathbb{P}}(n^{-1})$*;*
(ii) $\widehat{D}_n = D_* + k^{-1}\det[\mathbf{D}_*]^{1/k}\mathrm{tr}[\mathbf{L}_n] + O_{\mathbb{P}}(n^{-1})$*;*
(iii) $\widehat{H}_n = H_* + k^{-1}\mathrm{tr}[\mathbf{L}_n \mathbf{B}_*^{-1}\mathbf{A}_*]/(k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}])^2 + O_{\mathbb{P}}(n^{-1})$*;*
(iv) $\widehat{S}_n = S_* + k^{-1}\mathrm{tr}[\mathbf{L}_n \mathbf{A}_*^{-1}\mathbf{B}_*] - k^{-1}\det[\mathbf{D}_*]^{1/k}\mathrm{tr}[\mathbf{L}_n] + O_{\mathbb{P}}(n^{-1})$*;*
(v) $\widehat{C}_n = C_* + k^{-1}\mathrm{tr}[\mathbf{L}_n \mathbf{A}_*^{-1}\mathbf{B}_*] - k^{-1}\mathrm{tr}[\mathbf{L}_n \mathbf{B}_*^{-1}\mathbf{A}_*]/(k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}])^2 + O_{\mathbb{P}}(n^{-1})$*; and*
(vi) $\widehat{G}_n = G_* + k^{-1}\det[\mathbf{D}_*]^{1/k}\mathrm{tr}[\mathbf{L}_n] - k^{-1}\mathrm{tr}[\mathbf{L}_n \mathbf{B}_*^{-1}\mathbf{A}_*]/(k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}])^2 + O_{\mathbb{P}}(n^{-1})$. □

Several remarks are in order. First, we do not prove Lemma 2 in the Appendix. Lemma 2(*i*, *ii*, and *iv*) are already established in lemma 4 of CW, and Lemma 2(*iii*) holds as a corollary of Lemma 3 below. The other results trivially follow from Lemma 2(*i*, *ii*, and *iii*). Second, some signs of the statistics are predetermined by the interrelationships between the Pythagorean means. That is, we know $\widehat{T}_n \geq \widehat{D}_n \geq \widehat{H}_n$, so that $\widehat{S}_n$, $\widehat{C}_n$, and $\widehat{G}_n$ are always greater than zero. Third, the asymptotic approximations of the statistics in Lemma 2 have different forms under $\mathcal{H}_0$ and $\mathcal{H}_1$. If $\mathcal{H}_0$ holds, $T_* = D_* = H_* = S_* = C_* = G_* = 0$, so that $\widehat{T}_n$, $\widehat{D}_n$, and $\widehat{H}_n$ are $O_{\mathbb{P}}(n^{-1/2})$, and $\widehat{S}_n$, $\widehat{C}_n$, and $\widehat{G}_n$ are $O_{\mathbb{P}}(n^{-1})$. On the other hand, $\widehat{T}_n$, $\widehat{D}_n$, $\widehat{H}_n$, $\widehat{S}_n$, $\widehat{C}_n$, and $\widehat{G}_n$ are $O_{\mathbb{P}}(1)$ under $\mathcal{H}_1$. These different forms make the test base element quantities useful in distinguishing $\mathcal{H}_0$ and $\mathcal{H}_1$.

We now define the first group of tests

$$\widehat{\mathfrak{B}}_n^{(1)} := \frac{nk}{4}\left(\widehat{T}_n^2 + \widehat{D}_n^2\right), \quad \widehat{\mathfrak{B}}_n^{(2)} := \frac{nk}{2}\left(\widehat{T}_n^2 + 2\widehat{S}_n\right), \quad \text{and} \quad \widehat{\mathfrak{B}}_n^{(3)} := \frac{nk}{2}\left(\widehat{D}_n^2 + 2\widehat{S}_n\right),$$

which modify the tests in CW. These tests exploit the discriminatory properties of the statistics $\widehat{T}_n$ and $\widehat{D}_n$, which embody elements of the Wald (1943) test principle. The coefficients of the statistics differ from those in CW: specifically, $\widehat{\mathfrak{B}}_n^{(1)}$ is (respectively, $\widehat{\mathfrak{B}}_n^{(2)}$ and $\widehat{\mathfrak{B}}_n^{(3)}$ are) defined by dividing the corresponding test in CW by $2k$ (respectively, 2). As detailed below, this modification makes their power comparisons more fair.

We define a second group of tests as follows:

$$\widehat{\mathfrak{D}}_n^{(1)} := \frac{nk}{4}\left(\widehat{T}_n^2 + \widehat{H}_n^2\right), \quad \widehat{\mathfrak{D}}_n^{(2)} := \frac{nk}{2}\left(\widehat{T}_n^2 + \widehat{C}_n\right), \quad \text{and} \quad \widehat{\mathfrak{D}}_n^{(3)} := \frac{nk}{2}\left(\widehat{H}_n^2 + \widehat{C}_n\right);$$

and

$$\widehat{\mathfrak{S}}_n^{(1)} := \frac{nk}{4}\left(\widehat{D}_n^2 + \widehat{H}_n^2\right), \quad \widehat{\mathfrak{S}}_n^{(2)} := \frac{nk}{2}\left(\widehat{D}_n^2 + 2\widehat{G}_n\right), \quad \text{and} \quad \widehat{\mathfrak{S}}_n^{(3)} := \frac{nk}{2}\left(\widehat{H}_n^2 + 2\widehat{G}_n\right).$$

Note that $\widehat{\mathfrak{D}}_n^{(1)}$, $\widehat{\mathfrak{D}}_n^{(2)}$, and $\widehat{\mathfrak{D}}_n^{(3)}$ are defined by associating the arithmetic mean with the harmonic mean, whereas $\widehat{\mathfrak{S}}_n^{(1)}$, $\widehat{\mathfrak{S}}_n^{(2)}$, and $\widehat{\mathfrak{S}}_n^{(3)}$ are defined by associating the geometric mean with the harmonic mean. As before, $\widehat{T}_n$, $\widehat{D}_n$, and $\widehat{H}_n$ are empowered with discriminatory capability.

3.2 Asymptotic Null Approximations of the Test Statistics

We now develop null approximations for each of the tests and start the development with corresponding null approximations of the test base elements. For notational simplicity, let

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

which follows by imposing the null $\mathbf{A}_* = \mathbf{B}_*$ on the linearization $\mathbf{L}_n$. The following result is derived from Lemma 2.

Corollary 1. *Given Assumption A and $\mathcal{H}_0$,*
(i) $\widehat{T}_n = k^{-1}\mathrm{tr}[\mathbf{K}_n] + O_{\mathbb{P}}(n^{-1})$*;*
(ii) $\widehat{D}_n = k^{-1}\mathrm{tr}[\mathbf{K}_n] + O_{\mathbb{P}}(n^{-1})$*;*
(iii) $\widehat{H}_n = k^{-1}\mathrm{tr}[\mathbf{K}_n] + O_{\mathbb{P}}(n^{-1})$*;*
(iv) $\widehat{S}_n = O_{\mathbb{P}}(n^{-1})$*;*
(v) $\widehat{C}_n = O_{\mathbb{P}}(n^{-1})$*; and*
(vi) $\widehat{G}_n = O_{\mathbb{P}}(n^{-1})$. □

Items (*i*), (*ii*), and (*iv*) of Corollary 1 are already available in CW.

The main implication of Corollary 1 is that $\widehat{T}_n$, $\widehat{D}_n$, and $\widehat{H}_n$ are asymptotically equivalent under the null, so that $\widehat{S}_n$, $\widehat{C}_n$, and $\widehat{G}_n$ have a convergence rate n^{-1} that is slower than $\widehat{T}_n$, $\widehat{D}_n$ and $\widehat{H}_n$. This aspect was noticed by CW in the case of $\widehat{S}_n$. Corollary 1 extends their results by showing that the same properties apply for $\widehat{C}_n$ and $\widehat{G}_n$. In consequence, the desired asymptotic null approximations involve study of higher

order approximations of $\widehat{T}_n$, $\widehat{D}_n$, and $\widehat{H}_n$. Lemma 4 of CW provides these for $\widehat{T}_n$ and $\widehat{D}_n$, and we present them here for completeness to make this study self-contained:

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

$$\widehat{D}_n = D_* + k^{-1}\det[\mathbf{D}_*]^{\frac{1}{k}}\{\mathrm{tr}[\mathbf{L}_n] + 2^{-1}\left(k^{-1}-1\right)\mathrm{tr}[\mathbf{L}_n]^2 + 2^{-1}(\mathrm{tr}[\mathbf{P}_n]^2 + \mathrm{tr}[\mathbf{U}_n^2] - \mathrm{tr}[\mathbf{W}_n^2])\}$$
$$+ k^{-1}\det[\mathbf{D}_*]^{\frac{1}{k}}[\mathrm{tr}[\mathbf{P}_n]\mathrm{tr}[\mathbf{R}_{j,*}] + \mathrm{tr}[\mathbf{J}_{j,n} + \mathbf{U}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_* - \mathbf{W}_n\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)$$
$$+ (2k)^{-1}\det[\mathbf{D}_*]^{\frac{1}{k}-1}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla^2_{\boldsymbol{\theta}}\det[\mathbf{D}_*](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1}). \quad (3)$$

The following lemma gives the next-order expansion of $\widehat{H}_n$.

Lemma 3. *(i) Given Assumption A,*

$$\widehat{H}_n = H_* + k^{-1}\mathrm{tr}[\mathbf{L}_n\mathbf{B}_*^{-1}\mathbf{A}_*]/(k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}])^2$$
$$+ (k^{-1}\mathrm{tr}[\mathbf{L}_n\mathbf{B}_*^{-1}\mathbf{A}_*])^2/(k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}])^3 - (k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}])^{-2}\{k^{-1}\mathrm{tr}[\mathbf{L}_n\mathbf{W}_n\mathbf{B}_*^{-1}\mathbf{A}_*]\}$$
$$- (k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}])^{-2}\{k^{-1}[\mathrm{tr}[(-\mathbf{J}_{j,n} + \mathbf{P}_n\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*)\mathbf{B}_*^{-1}\mathbf{A}_*]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)\}$$
$$- (2k)^{-1}(k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}])^{-2}\{(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*^{-1}](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)\} + o_{\mathbb{P}}(n^{-1}),$$

where $\mathbf{J}_{j,n} := \mathbf{G}_{j,n} - \mathbf{H}_{j,n}$*;*
(ii) If $\mathcal{H}_0$ *holds,* $\widehat{H}_n = \widehat{H}_n^\star + o_{\mathbb{P}}(n^{-1})$*, where*

$$\widehat{H}_n^\star := k^{-1}\mathrm{tr}[\mathbf{K}_n] + (k^{-1}\mathrm{tr}[\mathbf{K}_n])^2 - k^{-1}\mathrm{tr}[\mathbf{K}_n\mathbf{W}_n]$$
$$-k^{-1}[\mathrm{tr}[(-\mathbf{J}_{j,n} + \mathbf{M}_n\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*)]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) - (2k)^{-1}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*^{-1}](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*). \quad \square$$

Lemma 3 is proved in the Appendix. Note that Lemma 2(*iii*) and Corollary 1(*iii*) follow from Lemma 3. Here, $\widehat{H}_n^\star$ of Lemma 3(*ii*) is the second-order expansion of $\widehat{H}_n$ under $\mathcal{H}_0$, and it is not hard to see that $\widehat{H}_n^\star = k^{-1}\mathrm{tr}[\mathbf{K}_n] + O_{\mathbb{P}}(n^{-1})$ under Assumption A.

Lemma 3 can also be used to obtain the next-order asymptotic expansions of $\widehat{C}_n$ and $\widehat{G}_n$ in Lemma 2, and these are provided in the following result.

Lemma 4. *Given Assumption A, we have:*
(i) if for all $d > 0$*,* $\mathbf{B}_* \neq d\mathbf{A}_*$*,*

(i.a) $\widehat{C}_n = C_* + k^{-1}\mathrm{tr}[\mathbf{L}_n\mathbf{A}_*^{-1}\mathbf{B}_*] - (k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}])^{-2}k^{-1}\mathrm{tr}[\mathbf{L}_n\mathbf{B}_*^{-1}\mathbf{A}_*] + O_{\mathbb{P}}(n^{-1})$*;*

(i.b) $\widehat{G}_n = G_* + k^{-1}\det[\mathbf{D}_*]^{1/k}\mathrm{tr}[\mathbf{L}_n] - k^{-1}\mathrm{tr}[\mathbf{L}_n\mathbf{B}_*^{-1}\mathbf{A}_*]/(k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}])^2 + O_{\mathbb{P}}(n^{-1})$*;*

(ii) if for some $d_* > 0$, $\mathbf{B}_* = d_*\mathbf{A}_*$,

(ii.a) $\widehat{C}_n = d_*\left\{k^{-1}\mathrm{tr}[\mathbf{L}_n^2] - k^{-2}\mathrm{tr}[\mathbf{L}_n]^2\right\} + o_{\mathbb{P}}(n^{-1})$*;*

(ii.b) $\widehat{G}_n = 2^{-1}d_*\left\{k^{-1}\mathrm{tr}[\mathbf{L}_n^2] - k^{-2}\mathrm{tr}[\mathbf{L}_n]^2\right\} + o_{\mathbb{P}}(n^{-1})$*;*

(iii) If in addition $\mathcal{H}_0$ *holds,*

(iii.a) $\widehat{C}_n = k^{-1}\mathrm{tr}[\mathbf{K}_n^2] - k^{-2}\mathrm{tr}[\mathbf{K}_n]^2 + o_{\mathbb{P}}(n^{-1})$*;*

(iii.b) $\widehat{G}_n = 2^{-1}\{k^{-1}\mathrm{tr}[\mathbf{K}_n^2] - k^{-2}\mathrm{tr}[\mathbf{K}_n]^2\} + o_{\mathbb{P}}(n^{-1})$. □

We prove Lemma 4 by combining Lemma 3 and lemma 4 of CW. The point of Lemma 4(*i.a*) is that the asymptotic expansions for $\widehat{C}_n$ and $\widehat{G}_n$ in Lemma 2 are useful when $\mathbf{B}_*$ is not proportional to $\mathbf{A}_*$. If $\mathbf{B}_*$ is proportional to $\mathbf{A}_*$, they are not usefully exploited in the approximations. Further higher-order expansions are needed, and they are given in Lemma 4(*ii*). CW's corollary 5 observes the same feature for $\widehat{S}_n$. For convenience, we state their result here: if for all $d > 0$, $\mathbf{B}_* \neq d\mathbf{A}_*$, $\widehat{S}_n = S_* + k^{-1}\mathrm{tr}[(\mathbf{A}_*^{-1}\mathbf{B}_* - \det[\mathbf{D}_*]^{\frac{1}{k}}\mathbf{I})\mathbf{L}_n] + o_{\mathbb{P}}(n^{-1/2})$; and if for some $d_* > 0$, $\mathbf{B}_* = d_*\mathbf{A}_*$, $\widehat{S}_n = -d_*(2k^2)^{-1}\mathrm{tr}[\mathbf{L}_n]^2 + d_*(2k)^{-1}\mathrm{tr}[\mathbf{L}_n^2] + o_{\mathbb{P}}(n^{-1})$. Note the same property holds as those for $\widehat{C}_n$ and $\widehat{G}_n$: the asymptotic expansion order of $\widehat{S}_n$ depends on whether $\mathbf{B}_*$ is proportional to $\mathbf{A}_*$ or not.

Lemmas 3 and 4 now straightforwardly deliver the asymptotic null approximations of the tests. We collect these together in the following theorem which characterizes the relationships between the test statistics.

Theorem 1. *Given Assumption A and* $\mathcal{H}_0$*,*

(i) $\widehat{\mathfrak{B}}_n^{(1)} = \frac{n}{2k}\mathrm{tr}[\mathbf{K}_n]^2 + o_{\mathbb{P}}(1)$, $\widehat{\mathfrak{B}}_n^{(2)} = \frac{n}{2}\mathrm{tr}[\mathbf{K}_n^2] + o_{\mathbb{P}}(1)$, *and* $\widehat{\mathfrak{B}}_n^{(3)} = \frac{n}{2}\mathrm{tr}[\mathbf{K}_n^2] + o_{\mathbb{P}}(1)$*;*

(ii) $\widehat{\mathfrak{D}}_n^{(1)} = \frac{n}{2k}\mathrm{tr}[\mathbf{K}_n]^2 + o_{\mathbb{P}}(1)$, $\widehat{\mathfrak{D}}_n^{(2)} = \frac{n}{2}\mathrm{tr}[\mathbf{K}_n^2] + o_{\mathbb{P}}(1)$, *and* $\widehat{\mathfrak{D}}_n^{(3)} = \frac{n}{2}\mathrm{tr}[\mathbf{K}_n^2] + o_{\mathbb{P}}(1)$*; and*

(iii) $\widehat{\mathfrak{S}}_n^{(1)} = \frac{n}{2k}\mathrm{tr}[\mathbf{K}_n]^2 + o_{\mathbb{P}}(1)$, $\widehat{\mathfrak{S}}_n^{(2)} = \frac{n}{2}\mathrm{tr}[\mathbf{K}_n^2] + o_{\mathbb{P}}(1)$, *and* $\widehat{\mathfrak{S}}_n^{(3)} = \frac{n}{2}\mathrm{tr}[\mathbf{K}_n^2] + o_{\mathbb{P}}(1)$. □

Theorem 1(*i*) corresponds to theorem 1 of CW. Applying their theorem yields Theorem 1(*i*) as a corollary. Theorems 1(*ii* and *iii*) also hold as corollaries of Corollary 1 and Lemma 4(*iii*). From Theorem 1, it follows that $\widehat{\mathfrak{B}}_n^{(1)}$, $\widehat{\mathfrak{D}}_n^{(1)}$, and $\widehat{\mathfrak{S}}_n^{(1)}$ are asymptotically equivalent under $\mathcal{H}_0$. Furthermore, $\widehat{\mathfrak{B}}_n^{(2)}$, $\widehat{\mathfrak{B}}_n^{(3)}$, $\widehat{\mathfrak{D}}_n^{(2)}$, $\widehat{\mathfrak{D}}_n^{(3)}$, $\widehat{\mathfrak{S}}_n^{(2)}$, and $\widehat{\mathfrak{S}}_n^{(3)}$ are also asymptotically equivalent.

3.3 Asymptotic Alternative Approximations of the Test Statistics

We now examine asymptotic approximations of the tests under the alternative. As before, we first examine asymptotic approximations of the test base elements. They are easily obtained by combining Lemmas 3, 4, and lemma 4 of CW. The following corollary collects them together.

Corollary 2. *Given Assumption A,*

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

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

(i.b) $\widehat{\mathfrak{B}}_n^{(2)} = \frac{nk}{2}(T_*^2 + 2S_*) + n\mathrm{tr}[[(T_*+1)\mathbf{A}_*^{-1}\mathbf{B}_* - \det[\mathbf{D}_*]^{\frac{1}{k}}\mathbf{I}]\mathbf{L}_n] + O_{\mathbb{P}}(1)$;

(i.c) $\widehat{\mathfrak{B}}_n^{(3)} = \frac{nk}{2}(D_*^2 + 2S_*) + n\{(D_*^2-1)\mathrm{tr}[\mathbf{L}_n] + \mathrm{tr}[\mathbf{A}_*^{-1}\mathbf{B}_*\mathbf{L}_n]\} + O_{\mathbb{P}}(1)$;

(i.d) $\widehat{\mathfrak{D}}_n^{(1)} = \frac{nk}{2}(\frac{1}{2}T_*^2 + \frac{1}{2}H_*^2) + \frac{n}{2}\left\{T_*\mathrm{tr}[\mathbf{L}_n\mathbf{A}_*^{-1}\mathbf{B}_*] + H_*\mathrm{tr}[\mathbf{L}_n\mathbf{B}_*^{-1}\mathbf{A}_*]/(k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}])^2\right\} + O_{\mathbb{P}}(1)$;

(i.e) $\widehat{\mathfrak{D}}_n^{(2)} = \frac{nk}{2}(T_*^2 + C_*) + \frac{n}{2}\left\{(2T_*+1)\mathrm{tr}[\mathbf{L}_n\mathbf{A}_*^{-1}\mathbf{B}_*] - \mathrm{tr}[\mathbf{L}_n\mathbf{B}_*^{-1}\mathbf{A}_*]/(k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}])^2\right\} + O_{\mathbb{P}}(1)$;

(i.f) $\widehat{\mathfrak{D}}_n^{(3)} = \frac{nk}{2}(H_*^2 + C_*) + \frac{n}{2}\left\{\mathrm{tr}[\mathbf{L}_n\mathbf{A}_*^{-1}\mathbf{B}_*] + (2H_*-1)\mathrm{tr}[\mathbf{L}_n\mathbf{B}_*^{-1}\mathbf{A}_*]/(k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}])^2\right\} + O_{\mathbb{P}}(1)$;

(i.g) $\widehat{\mathfrak{S}}_n^{(1)} = \frac{nk}{2}(\frac{1}{2}D_*^2 + \frac{1}{2}H_*^2) + \frac{n}{2}\left\{H_*\mathrm{tr}[\mathbf{L}_n\mathbf{B}_*^{-1}\mathbf{A}_*]/(k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}])^2 + D_*\det[\mathbf{D}_*]^{1/k}\mathrm{tr}[\mathbf{L}_n]\right\} + O_{\mathbb{P}}(1)$;

(i.h) $\widehat{\mathfrak{S}}_n^{(2)} = \frac{nk}{2}(D_*^2 + 2G_*) + n\{(kD_*+1)\det[\mathbf{D}_*]^{1/k}\mathrm{tr}[\mathbf{L}_n] - \mathrm{tr}[\mathbf{L}_n\mathbf{B}_*^{-1}\mathbf{A}_*]/(k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}])^2\} + O_{\mathbb{P}}(1)$;

(i.i) $\widehat{\mathfrak{S}}_n^{(3)} = \frac{nk}{2}(H_*^2 + 2G_*) + n\{(H_*-1)\mathrm{tr}[\mathbf{L}_n\mathbf{B}_*^{-1}\mathbf{A}_*]/(k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}])^2 + \det[\mathbf{D}_*]^{1/k}\mathrm{tr}[\mathbf{L}_n]\} + O_{\mathbb{P}}(1)$;

(ii) if for some $d_* > 0$, $\mathbf{B}_* = d_*\mathbf{A}_*$, *for* $j = 1, 2, 3$, $\widehat{\mathfrak{B}}_n^{(j)}$, $\widehat{\mathfrak{D}}_n^{(j)}$, *and* $\widehat{\mathfrak{S}}_n^{(j)}$ *are equal to* $\frac{nk}{2}(d_*-1)^2 + nd_*(d_*-1)\mathrm{tr}[\mathbf{L}_n] + O_{\mathbb{P}}(1)$. □

Corollary 2 is implied by Lemmas 3, 4, and lemma 4 of CW so its proof is omitted.

Some remarks are warranted. First, the leading term of each test is the first term of the right sides in (*i*, *ii*), and all of them are $O(n)$. This divergence implies that all tests are consistent. Second, if $\mathbf{B}_*$ is proportional to $\mathbf{A}_*$, every test is equivalent even under the alternative. If $d_* = 1$, the given approximations are nothing but the null approximations in Theorem 1. Third, if $\mathbf{B}_*$ is not proportional to $\mathbf{A}_*$, the global powers of the tests are determined by the leading terms of the right sides in (*i.a–i.i*). In particular, the leading terms of the CW tests are also determined by k, but k affects the leading terms of the CW tests in different ways. Consequently, the global power of each test is differently affected by k, and this aspect makes power comparisons difficult. On the other hand, the leading terms of all of our tests are affected by k in the same manner, so that their global powers can be compared without considering the effects of k. Fourth, one of the main determinants for a test to be globally most powerful (GMP) is the leading term. For example, if the leading term $\frac{1}{2}(T_*^2 + D_*^2)$ of $\widehat{\mathfrak{B}}_n^{(1)}$ is greater than the other leading terms in (*i.b–i.i*), $\widehat{\mathfrak{B}}_n^{(1)}$ is likely to be GMP. We compare them and determine conditions for each term to be greater than the others. The results are collected in the following theorem.

Theorem 2. *Given Assumption A and* $\mathcal{H}_1$, *if for all* $d > 0$, $\mathbf{B}_* \neq d\mathbf{A}_*$,

(i) the leading terms of $\widehat{\mathfrak{B}}_n^{(1)}$, $\widehat{\mathfrak{D}}_n^{(1)}$, *and* $\widehat{\mathfrak{S}}_n^{(1)}$ *cannot be largest among those in Corollary 2(i.a–i.i);*

(ii) the leading term of $\widehat{\mathfrak{B}}_n^{(2)}$ *is greater than the others if and only if* $T_*^2 \geq D_*^2$, $T_*^2 + S_* \geq H_*^2 + G_*$, *and* $S_* \geq G_*$;

(iii) the leading term of $\widehat{\mathfrak{B}}_n^{(3)}$ *is greater than the others and is GMP if and only if* $D_*^2 \geq T_*^2$, $D_*^2 + S_* \geq H_*^2 + G_*$, *and* $S_* \geq G_*$;
(iv) the leading term of $\widehat{\mathfrak{D}}_n^{(2)}$ *is greater than the others if and only if* $T_*^2 + S_* \geq D_*^2 + G_*$, $T_*^2 + S_* \geq H_*^2 + G_*$, *and* $G_* \geq S_*$;
(v) the leading term of $\widehat{\mathfrak{D}}_n^{(3)}$ *is greater than the others if and only if* $H_*^2 + G_* \geq T_*^2 + S_*$, $H_*^2 + G_* \geq D_*^2 + S_*$, *and* $S_* \geq G_*$;
(vi) the leading term of $\widehat{\mathfrak{S}}_n^{(2)}$ *is greater than the others if and only if* $D_*^2 \geq H_*^2$, $D_*^2 + G_* \geq T_*^2 + S_*$, *and* $G_* \geq S_*$; *and*
(vii) the leading term of $\widehat{\mathfrak{S}}_n^{(3)}$ *is greater than the others if and only if* $H_*^2 \geq D_*^2$, $H_*^2 + G_* \geq T_*^2 + S_*$, *and* $G_* \geq S_*$. □

Some remarks are warranted. First, by Theorem 2(*i*), some caution is needed in testing $\mathbf{A}_* = \mathbf{B}_*$ using $\widehat{\mathfrak{B}}_n^{(1)}$, $\widehat{\mathfrak{D}}_n^{(1)}$, and $\widehat{\mathfrak{S}}_n^{(1)}$ for their leading terms cannot be greater than the others, although their local alternative properties may be different. The other tests are likely to yield better inferential discrimination in terms of global power. Second, the necessary and sufficient conditions in Theorems 2(*i–vii*) can be consistently selected by estimating T_*, D_*, S_* and by comparing the conditions in Theorem 2. For example, if $\widehat{T}_n^2 \geq \widehat{D}_n^2$, $\widehat{T}_n^2 + \widehat{S}_n \geq \widehat{H}_n^2 + \widehat{G}_n$, and $\widehat{S}_n \geq \widehat{G}_n$ and the sample size is reasonably large, testing the hypotheses by relying on $\widehat{\mathfrak{B}}_n^{(2)}$ can be better than the other tests. Finally, if $k = 2$, it follows that $S_* \geq G_*$ from the fact that $(D_* + 1)^2 = (T_* + 1)(H_* + 1)$. This implies that $\widehat{\mathfrak{B}}_n^{(2)}$, $\widehat{\mathfrak{B}}_n^{(3)}$, and $\widehat{\mathfrak{D}}_n^{(3)}$ are likely GMP.

3.4 Asymptotic Local Alternative Approximations of the Test Statistics

We now examine asymptotic approximations of the tests under local alternatives. We consider the following local alternative: for some positive-definite $\bar{\mathbf{A}}_*$ and $\bar{\mathbf{B}}_*$ such that $\bar{\mathbf{A}}_* \neq \bar{\mathbf{B}}_*$,

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

Here, as the sample size n tends to infinity, $\mathbf{A}_{*,n}$ and $\mathbf{B}_{*,n}$ converge to $\mathbf{A}_*$ and $\mathbf{B}_*$, respectively, at the rate $n^{-1/2}$. Note that $\mathcal{H}_\ell$ reduces to $\mathcal{H}_0$ if $\bar{\mathbf{A}}_* = \bar{\mathbf{B}}_*$. The local alternative differs from the null by requiring that $\bar{\mathbf{A}}_* \neq \bar{\mathbf{B}}_*$. This local alternative generalizes that used in CW, where it is assumed that $\bar{\mathbf{A}}_* = \mathbf{0}$.

The following separate conditions are imposed for the local alternative approximations.

Assumption B (Local Alternative)**.** *(i)* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a complete probability space;*
(ii) $\boldsymbol{\Theta} \subset \mathbb{R}^\ell$ *is a compact convex set with non-empty interior and* $k \in \mathbb{N}$;

(iii) a sequence of measurable mappings $\{\widehat{\boldsymbol{\theta}}_n : \Omega \mapsto \boldsymbol{\Theta}\}$ *is consistent for a unique* $\boldsymbol{\theta}_* \in int(\boldsymbol{\Theta})$*;*
(iv) $\mathbf{A} : \boldsymbol{\Theta} \mapsto \mathbb{R}^{k\times k}$ *and* $\mathbf{B} : \boldsymbol{\Theta} \mapsto \mathbb{R}^{k\times k}$ *are in* $\mathcal{C}^{(1)}(\boldsymbol{\Theta})$*, and* $\mathbf{A}_*$ *and* $\mathbf{B}_*$ *are positive definite;*
(v) The symmetric mappings $\bar{\mathbf{A}} : \boldsymbol{\Theta} \mapsto \mathbb{R}^{k\times k}$ *and* $\bar{\mathbf{B}} : \boldsymbol{\Theta} \mapsto \mathbb{R}^{k\times k}$ *are in* $\mathcal{C}^{(2)}(\boldsymbol{\Theta})$ *and such that* $\bar{\mathbf{A}}_* := \bar{\mathbf{A}}(\boldsymbol{\theta}_*)$ *and* $\bar{\mathbf{B}}_* := \bar{\mathbf{B}}(\boldsymbol{\theta}_*)$ *are positive definite, and* $\bar{\mathbf{A}}_* \neq \bar{\mathbf{B}}_*$*;*
(vi) $\mathbf{A}_n(\cdot)$ *and* $\mathbf{B}_n(\cdot)$ *are consistent for* $\mathbf{A}(\cdot)$ *and* $\mathbf{B}(\cdot)$*, respectively, uniformly on* $\boldsymbol{\Theta}$*;*
(vii) $\sqrt{n}[(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)', \text{vech}[\mathbf{A}_n - \mathbf{A}_{n,*}]', \text{vech}[\mathbf{B}_n - \mathbf{B}_{n,*}]']' = O_{\mathbb{P}}(1)$*;*
(viii) 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* $\boldsymbol{\Theta}$*; and*
(ix) for $j = 1, \ldots, \ell$*,* $\mathbf{H}_{j,o,n} = O_{\mathbb{P}}(n^{-1/2})$ *and* $\mathbf{G}_{j,o,n} = O_{\mathbb{P}}(n^{-1/2})$*, where* $\mathbf{H}_{j,o,n} := \mathbf{A}_*^{-1}\partial_j(\mathbf{A}_n - \mathbf{A}_*)$ *and* $\mathbf{G}_{j,o,n} := \mathbf{B}_*^{-1}\partial_j(\mathbf{B}_n - \mathbf{B}_*)$*.* □

The major differences between Assumptions A and B are in Assumptions B(*v*, *vii*, and *ix*). The localizing matrix parameters $\bar{\mathbf{A}}_*$ and $\bar{\mathbf{B}}_*$ are formally introduced in Assumption B(*v*), and the other two conditions modify the corresponding conditions in Assumption A to accommodate the presence of these localizing parameters.

Before examining the local asymptotic approximations, we provide notations relevant to the main claims of this section. We define

$$\mathbf{W}_{o,n} := \mathbf{B}_*^{-1}(\mathbf{B}_n - \mathbf{B}_{*,n}); \quad \mathbf{W}_{a,n} := \mathbf{B}_{*,n}^{-1}(\mathbf{B}_n - \mathbf{B}_{*,n});$$
$$\mathbf{U}_{o,n} := \mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_{*,n}); \quad \mathbf{U}_{a,n} := \mathbf{A}_{*,n}^{-1}(\mathbf{A}_n - \mathbf{A}_{*,n});$$
$$\mathbf{P}_{o,n} := \mathbf{W}_{o,n} - \mathbf{U}_{o,n}; \quad \mathbf{P}_{a,n} := \mathbf{W}_{a,n} - \mathbf{U}_{a,n};$$
$$\mathbf{M}_{o,n} := \mathbf{B}_*^{-1}(\mathbf{B}_n - \mathbf{A}_n - \mathbf{B}_{*,n} + \mathbf{A}_{*,n}); \quad \mathbf{M}_{a,n} := \mathbf{B}_{*,n}^{-1}(\mathbf{B}_n - \mathbf{A}_n - \mathbf{B}_{*,n} + \mathbf{A}_{*,n});$$
$$\mathbf{L}_{o,n} := \mathbf{P}_{o,n} + \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{R}_{j,*}; \quad \mathbf{K}_{o,n} := \mathbf{M}_{o,n} + \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{S}_{j,*};$$
$$\mathbf{L}_{a,n} := \mathbf{P}_{a,n} + \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})(\mathbf{B}_{*,n}^{-1}\partial_j\mathbf{B}_{*,n} - \mathbf{A}_{*,n}^{-1}\partial_j\mathbf{A}_{*,n});$$
$$\mathbf{K}_{a,n} := \mathbf{M}_{a,n} + \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{B}_{*,n}^{-1}(\partial_j\mathbf{B}_{*,n} - \partial_j\mathbf{A}_{*,n}),$$

where $\mathbf{S}_{j*} := \mathbf{A}_*^{-1}(\partial_j\mathbf{B}_* - \partial_j\mathbf{A}_*)$. These statistics are defined to highlight the asymptotic roles of the localizing parameters. The statistics indexed by the subscript "o" correspond to those in previous sections, in which the localizing parameters are absent (zero). On the other hand, the statistics indexed by the subscript "a" are defined to explicitly consider the asymptotic effects of the locality parameters. Specifically, the inverse matrices in $\mathbf{W}_{o,n}$, $\mathbf{U}_{o,n}$, and $\mathbf{M}_{o,n}$ are difference from those in $\mathbf{W}_{a,n}$, $\mathbf{U}_{a,n}$, and $\mathbf{M}_{a,n}$, respectively.

If the localizing parameters are zero matrices in the inverse matrices, $\mathbf{W}_{o,n}$, $\mathbf{U}_{o,n}$, and $\mathbf{M}_{o,n}$ are reduced versions of $\mathbf{W}_{a,n}$, $\mathbf{U}_{a,n}$, and $\mathbf{M}_{a,n}$. Further note that $\mathbf{A}_* = \mathbf{B}_*$ under $\mathcal{H}_\ell$, so that $\mathbf{P}_{o,n} = \mathbf{M}_{o,n}$, $\mathbf{R}_{j,*} = \mathbf{S}_{j,*}$, and $\mathbf{L}_{o,n} = \mathbf{K}_{o,n}$. Using this fact, we let

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

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

$$\begin{aligned}\widehat{H}_{o,n} &:= k^{-1}\mathrm{tr}[\mathbf{K}_{o,n}] + (k^{-1}\mathrm{tr}[\mathbf{K}_{o,n}])^2 - k^{-1}\mathrm{tr}[\mathbf{K}_{o,n}\mathbf{W}_{o,n}] \\ &\quad - k^{-1}[\mathrm{tr}[-\mathbf{J}_{j,o,n} + \mathbf{M}_{o,n}\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) - (2k)^{-1}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\mathrm{tr}[\mathbf{D}_*^{-1}](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*),\end{aligned}$$

and define $\widehat{S}_{o,n} := \widehat{T}_{o,n} - \widehat{D}_{o,n}$, $\widehat{C}_{o,n} := \widehat{T}_{o,n} - \widehat{H}_{o,n}$, and $\widehat{G}_{o,n} := \widehat{D}_{o,n} - \widehat{H}_{o,n}$, where

$$\begin{aligned}\mathbf{J}_{j,o,n} &:= \mathbf{G}_{j,o,n} - \mathbf{H}_{j,o,n} := \mathbf{B}_*^{-1}\partial_j(\mathbf{B}_n - \mathbf{B}_{*,n}) - \mathbf{A}_*^{-1}\partial_j(\mathbf{A}_n - \mathbf{A}_{*,n}); \quad \text{and} \\ \mathbf{J}_{j,a,n} &:= \mathbf{G}_{j,a,n} - \mathbf{H}_{j,a,n} := \mathbf{B}_{*,n}^{-1}\partial_j(\mathbf{B}_n - \mathbf{B}_{*,n}) - \mathbf{A}_{*,n}^{-1}\partial_j(\mathbf{A}_n - \mathbf{A}_{*,n}).\end{aligned}$$

These are the second-order approximations of the test base elements that are obtained by imposing $\mathbf{A}_* = \mathbf{B}_*$ and by letting the localizing parameters be zero in the inverse matrices. These definitions are obtained by reformulating (2), (3), and Lemma 3(*i*) to fit the current context. In particular, $\widehat{S}_{o,n}$ is simplified to $\widehat{S}_{o,n} := -\frac{1}{2k^2}\mathrm{tr}[\mathbf{K}_{o,n}]^2 + \frac{1}{2k}\mathrm{tr}[\mathbf{K}_{o,n}^2]$ after some tedious algebra. All these statistics are $O_{\mathbb{P}}(n^{-1})$.

Due to the effects of ignoring the asymptotic impact of $\bar{\mathbf{A}}_*$ and $\bar{\mathbf{B}}_*$, these statistics poorly approximate the test base elements under $\mathcal{H}_\ell$. Their differences from the second-order approximations of the tests are not asymptotically negligible and affect their asymptotic approximations under $\mathcal{H}_\ell$. The following lemma explicitly shows their differences:

Lemma 5. *Given Assumption B and $\mathcal{H}_\ell$,*

$$(i)\ \ \widehat{T}_n - \widehat{T}_{o,n} = n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{V}_*] - n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{F}_*\mathbf{W}_{o,n} - \mathbf{C}_*\mathbf{U}_{o,n}] + n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{K}_{o,n}\mathbf{V}_*] - (nk)^{-1}\mathrm{tr}[\mathbf{C}_*\mathbf{V}_*] + n^{-1/2}k^{-1}[\mathrm{tr}[\mathbf{Q}_{j,*} - (\mathbf{F}_*\mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{C}_*\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1}),$$

where $\mathbf{F}_* := \mathbf{B}_*^{-1}\bar{\mathbf{B}}_*$, $\mathbf{C}_* := \mathbf{A}_*^{-1}\bar{\mathbf{A}}_*$, $\mathbf{V}_* := \mathbf{F}_* - \mathbf{C}_*$, *and* $\mathbf{Q}_{j,*} := \mathbf{B}_*^{-1}\partial_j\bar{\mathbf{B}}_* - \mathbf{A}_*^{-1}\partial_j\bar{\mathbf{A}}_*$;

$$(ii)\ \ \widehat{D}_n - \widehat{D}_{o,n} = n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{V}_*] - n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{F}_*\mathbf{W}_{o,n} - \mathbf{C}_*\mathbf{U}_{o,n}] + n^{-1/2}k^{-2}\mathrm{tr}[\mathbf{V}_*]\mathrm{tr}[\mathbf{K}_{o,n}] + (2nk^2)^{-1}\mathrm{tr}[\mathbf{V}_*]^2 + (2nk)^{-1}(\mathrm{tr}[\mathbf{C}_*^2] - \mathrm{tr}[\mathbf{F}_*^2]) + n^{-1/2}k^{-1}[\mathrm{tr}[\mathbf{Q}_{j,*} - (\mathbf{F}_*\mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{C}_*\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1});$$

$$(iii)\ \ \widehat{H}_n - \widehat{H}_{o,n} = n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{V}_*] - (nk)^{-1}\mathrm{tr}[\mathbf{F}_*\mathbf{V}_*] - n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{K}_{o,n}\mathbf{V}_*] + 2(n^{1/2}k^2)^{-1}\mathrm{tr}[\mathbf{V}_*]\mathrm{tr}[\mathbf{K}_{o,n}] + (nk^2)^{-1}\mathrm{tr}[\mathbf{V}_*]^2 - n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{F}_*\mathbf{W}_{o,n} - \mathbf{C}_*\mathbf{U}_{o,n}] + n^{-1/2}k^{-1}[\mathrm{tr}[\mathbf{Q}_{j,*} - (\mathbf{F}_*\mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{C}_*\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1});$$

$$(iv)\ \ \widehat{S}_n - \widehat{S}_{o,n} = (2k)^{-1}\mathrm{tr}[(\mathbf{K}_{o,n} + n^{-1/2}\mathbf{V}_*)^2] - (2k^2)^{-1}\mathrm{tr}[\mathbf{K}_{o,n} + n^{-1/2}\mathbf{V}_*]^2 + (2k^2)^{-1}\mathrm{tr}[\mathbf{K}_{a,n}]^2 - (2k)^{-1}\mathrm{tr}[\mathbf{K}_{o,n}^2] + o_{\mathbb{P}}(n^{-1});$$

$$(v)\ \ \widehat{C}_n - \widehat{C}_{o,n} = (nk)^{-1}\mathrm{tr}[\mathbf{V}_*^2] - 2(n^{1/2}k^2)^{-1}\mathrm{tr}[\mathbf{V}_*]\mathrm{tr}[\mathbf{K}_{o,n}] + 2(n^{1/2}k)^{-1}\mathrm{tr}[\mathbf{K}_{o,n}\mathbf{V}_*] - (nk^2)^{-1}\mathrm{tr}[\mathbf{V}_*]^2 + o_{\mathbb{P}}(n^{-1});$$

and

$$(vi)\ \ \widehat{G}_n - \widehat{G}_{o,n} = (2nk)^{-1}\mathrm{tr}[\mathbf{V}_*^2] - (n^{1/2}k^2)^{-1}\mathrm{tr}[\mathbf{V}_*]\mathrm{tr}[\mathbf{K}_{o,n}] - (2nk^2)^{-1}\mathrm{tr}[\mathbf{V}_*]^2 + (n^{1/2}k)^{-1}\mathrm{tr}[\mathbf{K}_{o,n}\mathbf{V}_*] + o_{\mathbb{P}}(n^{-1}).\ \square$$

Several remarks are warranted. First, note that if $\bar{\mathbf{A}}_* = \bar{\mathbf{B}}_* \equiv \mathbf{0}$, $\mathbf{F}_* = \mathbf{C}_* = \mathbf{V}_* = \mathbf{0}$ and for each $j = 1, 2, \ldots, \ell$, $\mathbf{Q}_j = \mathbf{0}$, so that all the leading terms in Lemma 5(*i–vi*) are zero matrices. This implies that $\widehat{T}_n - \widehat{T}_{o,n} = o_{\mathbb{P}}(n^{-1})$, $\widehat{D}_n - \widehat{D}_{o,n} = o_{\mathbb{P}}(n^{-1})$, $\widehat{H}_n - \widehat{H}_{o,n} = o_{\mathbb{P}}(n^{-1})$, $\widehat{S}_n - \widehat{S}_{o,n} = o_{\mathbb{P}}(n^{-1})$,

$\widehat{C}_n - \widehat{C}_{o,n} = o_{\mathbb{P}}(n^{-1})$, and $\widehat{G}_n - \widehat{G}_{o,n} = o_{\mathbb{P}}(n^{-1})$. Therefore, $\widehat{T}_{o,n}$, $\widehat{D}_{o,n}$, $\widehat{H}_{o,n}$, $\widehat{S}_{o,n}$, $\widehat{C}_{o,n}$, and $\widehat{G}_{o,n}$ are the second-order approximations under the condition that $\bar{\mathbf{A}}_* = \bar{\mathbf{B}}_* \equiv \mathbf{0}$. This is the desired aspect of these definitions. Second, if $\bar{\mathbf{A}}_* = \mathbf{0}$, Lemmas 5(*i*, *ii*, and *iv*) reduce to lemma 5 of CW. Thus, Lemma 5 generalizes those results. Third, using these differences, the asymptotic approximations of the tests can also be derived, as they depend on the second-order approximations of the test base elements under $\mathcal{H}_\ell$. We provide them in the following theorem. In the statement of the result, and elsewhere in the paper we use $\mathrm{tr}\,[F]^2$ to represent $(\mathrm{tr}\,[F])^2$.

Theorem 3. *Given Assumption B and $\mathcal{H}_\ell$,*
(i) $\widehat{\mathfrak{B}}_n^{(1)} = \frac{1}{2k}\mathrm{tr}[\mathbf{V}_* + \sqrt{n}\mathbf{K}_{o,n}]^2 + o_{\mathbb{P}}(1)$*, and for* $i = 2, 3$, $\widehat{\mathfrak{B}}_n^{(i)} = \frac{1}{2}\mathrm{tr}[(\mathbf{V}_* + \sqrt{n}\mathbf{K}_{o,n})^2] + o_{\mathbb{P}}(1)$*;*
(ii) $\widehat{\mathfrak{D}}_n^{(1)} = \frac{1}{2k}\mathrm{tr}[\mathbf{V}_* + \sqrt{n}\mathbf{K}_{o,n}]^2 + o_{\mathbb{P}}(1)$*, and for* $i = 2, 3$, $\widehat{\mathfrak{D}}_n^{(i)} = \frac{1}{2}\mathrm{tr}[(\mathbf{V}_* + \sqrt{n}\mathbf{K}_{o,n})^2] + o_{\mathbb{P}}(1)$*; and*
(iii) $\widehat{\mathfrak{S}}_n^{(1)} = \frac{1}{2k}\mathrm{tr}[\mathbf{V}_* + \sqrt{n}\mathbf{K}_{o,n}]^2 + o_{\mathbb{P}}(1)$*, and for* $i = 2, 3$, $\widehat{\mathfrak{S}}_n^{(i)} = \frac{1}{2}\mathrm{tr}[(\mathbf{V}_* + \sqrt{n}\mathbf{K}_{o,n})^2] + o_{\mathbb{P}}(1)$. □

Therefore, the asymptotic approximations of the tests are obtained by shifting the location parameter of $\mathrm{tr}[\mathbf{K}_{o,n}]$ by $n^{-1/2}\mathrm{tr}[\mathbf{V}_*]$, from which the local power of the tests is derived. We also note that if $\mathrm{tr}[\mathbf{V}_*] = 0$, $\widehat{\mathfrak{B}}_n^{(1)}$, $\widehat{\mathfrak{D}}_n^{(1)}$, and $\widehat{\mathfrak{S}}_n^{(1)}$ do not have local power different from size. Similarly, if $\mathrm{tr}[\mathbf{V}_*^2] = 0$, then $\widehat{\mathfrak{B}}_n^{(2)}$, $\widehat{\mathfrak{B}}_n^{(3)}$, $\widehat{\mathfrak{D}}_n^{(2)}$, $\widehat{\mathfrak{D}}_n^{(3)}$, $\widehat{\mathfrak{S}}_n^{(2)}$, and $\widehat{\mathfrak{S}}_n^{(3)}$ have local power equal to size. Thus, $\mathrm{tr}[\mathbf{V}_*] \neq 0$ and $\mathrm{tr}[\mathbf{V}_*^2] \neq 0$ are necessary for these tests to have non-trivial local powers, respectively.

Before moving to the next section, we note that the local asymptotic approximations of the tests are equivalent to that of the likelihood ratio test under certain conditions. First, Nagao (1967), Nagarsenker and Pillai (1973), Muirhead (1982), and Anderson (2003) examine the LR test that a covariance matrix is equal to a certain matrix. CW apply this test and suppose that $\mathbf{X}_t \sim N(\boldsymbol{\theta}_*, \mathbf{B}_*)$ and that there exists is a consistent estimator $\widehat{\mathbf{A}}_n \to_p \mathbf{A}_*$. They then test for the equivalence $\mathbf{A}_* = \mathbf{B}_*$ using the LR test, and their theorem 6 implies that the LR test is locally equivalent to the tests indexed by (2) or (3). From the local optimality property of the LR test, the tests indexed by (2) or (3) are also locally optimal. Second, Mauchly (1940), Muirhead (1982), and Anderson (2003) test the sphericity condition: for some d_*, $\mathbf{B}_* = d_*\mathbf{A}_*$, and theorem 7 of CW implies that the LR test that is obtained under the same distributional condition as the above is locally equivalent to the difference between the tests indexed by (2) or (3) and those indexed by (1).

4 Monte Carlo Experiments

This section reports Monte Carlo experiments examining the performance of the tests analyzed in the previous section. As the structures parallel those given in CW and may be used in the same context, we also

consider applications of our procedures to information matrix testing. This helps to corroborate the relevance of the asymptotic theory. We consider the linear regression and probit models and test for correct distributional assumptions underlying ML estimation.

4.1 Linear Regression

We examine the finite sample properties of the tests by estimating the unknown parameters that are present in linear Gaussian regression models. Specifically, we start by assuming the model

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

with $U_t \mid \mathbf{X}_t \sim$ IID $N(0, \sigma^2)$ and where the unknown parameters $\boldsymbol{\beta}$ and σ^2 are estimated by ML. We consider three DGP classes. First, we let

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

where $U_t \mid \mathbf{X}_t \sim$ IID $N(0, \sigma_*^2)$ and $(\boldsymbol{\beta}_*', \sigma_*^2)' = (0, 1, 1)'$ with $\mathbf{X}_t = (1, X_t)'$ and $X_t \sim$ IID $N(0, 1)$. This DGP is correctly specified by the model, and the information matrix equality holds in ML estimation. Second, we consider four different DGPs for examining the global power properties. These are as follows:

- ALT1: $\mathbb{E}[Y_t|X_t] = X_t$, $U_t|X_t \sim$ independent $N(0, \exp(X_t))$, and $X_t \sim$ IID $N(1, 1)$;
- ALT2: $\mathbb{E}[Y_t|X_t] = X_t$, $U_t|X_t \sim$ IID $MN(-1, 1; 1, 1; 0.5)$, and $X_t \sim$ IID $N(0, 1)$;
- ALT3: $\mathbb{E}[Y_t|\mathbf{X}_t] = X_t$, $U_t|X_t \sim$ IID t_3, and $X_t \sim$ IID $N(0, 1)$; and
- ALT4: $\mathbb{E}[Y_t|\mathbf{X}_t] = X_t + \frac{1}{2}X_t^2$, $U_t|X_t \sim$ IID $N(0, 1)$, and $X_t \sim$ IID $N(0, 1)$.

Here, $Z \sim MN(a, b; c, d; p)$ denotes a finite mixture of normal distributions: $Z \sim N(a, b)$ with probability p, and $Z \sim N(c, d)$ with probability $1 - p$, and t_3 denotes the t-distribution with 3 degrees of freedom. The first alternative exhibits conditional heteroskedasticity. Although the conditional mean is correctly specified by the model, the error distribution is misspecified by the presence of the conditional heteroskedasticity. The next alternative has a PDF with two peaks and dispersed distributions, and the third alternative has heavy tails. These two DGPs are included for examining the effects of distributional misspecification with a particular focus on heavy tails. The final DGP has misspecification in the conditional mean, and this affects the asymptotic distribution of the ML estimator. Under these four DGPs, the model is misspecified, so that the information matrix equality does not hold. We use these alternatives for examining the global powers of the tests.

Third, we consider another four DGPs for examining the local power properties. These are

- LOC1: $E[Y_t|X_t] = X_t$, $U_t|X_t \sim$ indepent $N(0, \exp(2n^{-1/2}X_t))$, and $X_t \sim$ IID $N(1,1)$;
- LOC2: $E[Y_t|X_t] = X_t$, $U_t|X_t \sim$ IID *MN*$(-(1-p), p; 1, 1; p)$, $p = 10n^{-1/2}$, and $X_t \sim$ IID $N(0,1)$;
- LOC3: $E[Y_t|\mathbf{X}_t] = X_t$, $U_t|X_t \sim$ IID $\sim N(0,1)/\{1 + n^{-1/2}(\mathcal{X}_3^2 - 1)\}^{1/2}$, and $X_t \sim$ IID $N(0,1)$;
- LOC4: $E[Y_t|\mathbf{X}_t] = X_t + 5n^{-1/2}X_t^2$, $U_t|X_t \sim$ IID $N(0,1)$, and $X_t \sim$ IID $N(0,1)$.

Here, $\mathcal{X}_3^2$ denotes a chi-squared variate with 3 degrees of freedom that is independent of the standard normal variate in the numerator. These DGPs are obtained by modifying the DGPs in the second group. Note that as the sample size tends to infinity, they approach the first DGP at the rate $n^{-1/2}$. If the sample size is finite and small, they are also similarly distributed to the DGPs in the second group. These DGPs are used to determine the local power properties of the tests.

There is a caveat for the local DGPs. The distribution of X_t in LOC1 is different from the others. The non-zero mean condition of X_t is required for satisfying Assumption B. If $\mathbb{E}[X_t] = 0$ or $\mathbb{E}[X_t^3] = 0$ as in the other DGPs, Assumption B(*v*) does not hold, and it approaches the first DGP at the rate of $n^{-1/4}$. Although its local power is not negligible, the theory in the previous section is not applicable for this case. Further higher-order approximations are required for the local asymptotic approximations of the tests. We thus let $X_t \sim N(1,1)$ for ALT1 and LOC1 DGPs so that $\mathbb{E}[X_t] \neq 0$ and $\mathbb{E}[X_t^3] \neq 0$, thereby ensuring the relevance of the theory in the previous section.

Testing is implemented by the following two-step approach. First, we work on the given model assumptions and estimate $\widehat{\mathbf{D}}_n := \widehat{\mathbf{B}}_n\widehat{\mathbf{A}}_n^{-1}$ by letting $\widehat{\mathbf{A}}_n$ be the consistent negative Hessian matrix and $\widehat{\mathbf{B}}_n$ the covariance matrix of the scores, viz.,

$$\widehat{\mathbf{A}}_n := \frac{1}{n}\sum_{t=1}^{n}\begin{bmatrix} \frac{1}{\widehat{\sigma}_n^2}\mathbf{X}_t\mathbf{X}_t' & \frac{1}{\widehat{\sigma}_n^4}\widehat{U}_t\mathbf{X}_t \\ \frac{1}{\widehat{\sigma}_n^4}\widehat{U}_t\mathbf{X}_t' & \frac{1}{2\widehat{\sigma}_n^6}\widehat{\sigma}_n^2 \end{bmatrix} \text{ and } \widehat{\mathbf{B}}_n := \frac{1}{n}\sum_{t=1}^{n}\begin{bmatrix} \frac{1}{\widehat{\sigma}_n^2}\widehat{U}_t^2\mathbf{X}_t\mathbf{X}_t' & \frac{1}{2\widehat{\sigma}_n^6}(\widehat{U}_t - \widehat{\sigma}_n^2)\widehat{U}_t\mathbf{X}_t \\ \frac{1}{2\widehat{\sigma}_n^6}(\widehat{U}_t - \widehat{\sigma}_n^2)\widehat{U}_t\mathbf{X}_t' & \frac{1}{4\widehat{\sigma}_n^8}(\widehat{U}_t^2 - \widehat{\sigma}_n^2)^2 \end{bmatrix},$$

where $\widehat{U}_t := Y_t - \mathbf{X}_t'\widehat{\boldsymbol{\beta}}_n$, and $(\widehat{\boldsymbol{\beta}}_n', \widehat{\sigma}_n^2)'$ is the ML estimator. These matrix estimators are used to construct the test statistics. Next, the parametric bootstrap is applied to our tests. As Horowitz (1994) points out, the parametric bootstrap helps to prevent the level distortion problem that is particularly enormous for testing the information matrix equality (*e.g.*, Taylor, 1987; Orme, 1990; Chesher and Spady, 1991; White, 1982). CW provide a detailed procedure for applying the parametric bootstrap.[2]

[2]The following URL provides the GAUSS program codes for testing the information matrix equality of the linear normal, linear exponential, linear Weibull, linear probit, linear logit, linear gompit, linear scobit, and linear tobit models: http://web.yonsei.ac.kr/jinseocho/research.htm.

In addition to the tests of the current study, we also make comparisons with other tests that are popularly used for testing the information matrix equality. Jarque and Bera's (1987) test is used for testing the normality assumption. We denote this as $\widehat{\mathfrak{J}}_n$ and use it to test the normality distribution assumption of the error. Chesher's (1983) and Lancaster's (1984) information matrix test is also compared with our tests. It is denoted $\widehat{\mathfrak{I}}_n$. We also apply the parametric bootstrap to $\widehat{\mathfrak{J}}_n$ and $\widehat{\mathfrak{I}}_n$.

The null simulation results are contained in Table 1. The bootstrap repetition is 500, and we report the empirical rejection rates in Table 1 by repeating independent experiments 5,000 times. The level of significance is 5%. All test statistics including $\widehat{\mathfrak{J}}_n$ and $\widehat{\mathfrak{I}}_n$ have empirical rejection rates very close to the nominal level. This result is not limited to the large sample size case alone. Even when the sample size is as small as 50, the empirical rejection rates are close to the nominal level. This aspect implies that the researcher can control the type I error without considering limitations on the sample size.

The global power simulation results are contained in Tables 2 to 5. The bootstrap repetition is 500, and we repeat independent experiments 2,000 times. The empirical rejection rates are contained in the tables using a significance of 5%.

We summarize the global power simulation results as follows. First, all tests are consistent: as n tends to infinity, the empirical rejection rates approach unity. Second, if the error distribution has heavy tails, $\widehat{\mathfrak{J}}_n$ or $\widehat{\mathfrak{I}}_n$ are more powerful than the tests of this study. Specifically, for ALT2 (resp. ALT3), $\widehat{\mathfrak{I}}_n$ (resp. $\widehat{\mathfrak{J}}_n$) is more powerful than any other test. In particular, Jarque and Bera's (1987) test is obtained by applying the Lagrange multiplier testing principle to the Pearson family distributions and the t-distribution belongs to this family. Accordingly, $\widehat{\mathfrak{J}}_n$ turns out to be the most powerful test for every case in ALT3. On the other hand, if the conditional variance or mean is misspecified, our tests are more powerful than $\widehat{\mathfrak{J}}_n$ or $\widehat{\mathfrak{I}}_n$. For example, for ALT1 (resp. ALT4), $\widehat{\mathfrak{S}}_n^{(3)}$ (resp. $\widehat{\mathfrak{B}}_n^{(2)}$ or $\widehat{\mathfrak{D}}_n^{(2)}$) is more powerful than other tests for most sample sizes. This aspect implies that it is generally hard to say in advance, and of course without prior information about the alternative, which test is more powerful than the others. Third, when comparing only our tests, we notice a tendency that the tests indexed by (2) or (3) become more powerful than that indexed by (1), as the sample size increases. For illustration, we focus on $\widehat{\mathfrak{B}}_n^{(1)}$, $\widehat{\mathfrak{B}}_n^{(2)}$, and $\widehat{\mathfrak{B}}_n^{(3)}$ of Table 3. When the sample size is small, the most powerful test among the three tests is $\widehat{\mathfrak{B}}_n^{(1)}$. As the sample size increases, however, $\widehat{\mathfrak{B}}_n^{(3)}$ becomes most powerful, and $\widehat{\mathfrak{B}}_n^{(2)}$ becomes more powerful than $\widehat{\mathfrak{B}}_n^{(1)}$. For $n \geq 1,200$, this power relationship is maintained. This tendency is not only for the $\widehat{\mathfrak{B}}_n$-indexed tests. The other tests indexed by $\widehat{\mathfrak{D}}_n$ and $\widehat{\mathfrak{S}}_n$ exhibit similar tendencies, and all the tests in Tables 2, 4, and 5 show the same tendency. This finding affirms Theorem 2(*i*). The leading terms of $\widehat{\mathfrak{B}}_n^{(1)}$, $\widehat{\mathfrak{D}}_n^{(1)}$, and $\widehat{\mathfrak{S}}_n^{(1)}$ cannot be greater than those of the other tests. Accordingly, they are unlikely to be the most powerful test when the sample size is moderately

large. Although the most powerful test is also determined by factors besides T_*, D_*, H_*, S_*, C_*, and G_*, Tables 2 to 5 show the general tendency for the tests indexed by (2) and (3) to be globally more powerful than that index by (1).

The local power simulation results are contained in Tables 6 to 9. The bootstrap repetition number is 500, and 3,000 independent replications were conducted. As before, the level of significance is 5%.

We summarize the local power simulation results as follows. First, for every local DGP, all tests converge to stable empirical rejection rates, as the sample size increases. The limits of the empirical rejection rates are between 5% and 100%. This aspect corroborates the convergence rate $n^{-1/2}$ as the determining rate for stable local distributions of the tests. Second, the local power of the Jarque-Bera (1987) test is higher than the others for LOC2 and LOC3. As discussed above, the Jarque-Bera (1987) test is designed to test for Pearson family distributions such as the t-distribution, and mixtures of normal distributions are better approximated by Pearson family distributions than those with (normal) misspecification. This property explains why the local powers of $\widehat{\mathfrak{J}}_n$ are higher. For the other local DGPs, the locally most powerful test among our tests is locally more powerful than $\widehat{\mathfrak{J}}_n$ and $\widehat{\mathfrak{I}}_n$. From this finding, we may conclude that our tests have comparable powers to other popularly used test statistics. Third, when comparing our tests only, we notice similar local power patterns among the tests. For every DGP, the tests indexed by (1) (resp. (2) and (3)) have similar local powers among themselves. For illustration, we focus on Table 6. Note that the empirical rejection rates of $\widehat{\mathfrak{B}}_n^{(1)}$, $\widehat{\mathfrak{D}}_n^{(1)}$, and $\widehat{\mathfrak{S}}_n^{(1)}$ converge to a certain number around 5%, whereas those of $\widehat{\mathfrak{B}}_n^{(2)}$, $\widehat{\mathfrak{B}}_n^{(3)}$, $\widehat{\mathfrak{D}}_n^{(2)}$, $\widehat{\mathfrak{D}}_n^{(3)}$, $\widehat{\mathfrak{S}}_n^{(2)}$, and $\widehat{\mathfrak{S}}_n^{(3)}$ converge to around 12%. This observation is not limited to Table 6 but applies also to Tables 7 – 9. Theorem 3 is corroborated by this finding. Note that the local approximations of $\widehat{\mathfrak{B}}_n^{(1)}$, $\widehat{\mathfrak{D}}_n^{(1)}$, and $\widehat{\mathfrak{S}}_n^{(1)}$ are equivalent, and so are those of $\widehat{\mathfrak{B}}_n^{(2)}$, $\widehat{\mathfrak{B}}_n^{(3)}$, $\widehat{\mathfrak{D}}_n^{(2)}$, $\widehat{\mathfrak{D}}_n^{(3)}$, $\widehat{\mathfrak{S}}_n^{(2)}$, and $\widehat{\mathfrak{S}}_n^{(3)}$. Finally, the local power patterns can be different from the global power patterns. The tests indexed by (1) can be locally more powerful than those indexed by (2) and (3). For instance, observe Table 9 and note that the empirical rejection rates of $\widehat{\mathfrak{B}}_n^{(1)}$, $\widehat{\mathfrak{D}}_n^{(1)}$, and $\widehat{\mathfrak{S}}_n^{(1)}$ are generally higher than those of $\widehat{\mathfrak{B}}_n^{(2)}$, $\widehat{\mathfrak{B}}_n^{(3)}$, $\widehat{\mathfrak{D}}_n^{(2)}$, $\widehat{\mathfrak{D}}_n^{(3)}$, $\widehat{\mathfrak{S}}_n^{(2)}$, and $\widehat{\mathfrak{S}}_n^{(3)}$.

4.2 Probit

For the next experiment we use a probit specification. The probit model specifies the conditional mean of a limited dependent variable Y_t as $\mathbb{E}[Y_t|\mathbf{X}_t] = \Phi(\mathbf{X}_t'\boldsymbol{\beta})$, where Φ is the standard normal CDF and $\mathbf{X}_t := (1, X_t)'$.

We examine the following DGPs for our experiments. First,

- $\mathbb{E}[Y_t|X_t] = \text{Probit}(1 + X_t)$ and $X_t \sim \text{IID } N(0, 1)$,

where 'Probit(x)' means $\Phi(x)$. The model is correctly specified for this DGP, and we use this DGP to examine the asymptotic null behavior of the tests. Second, we examine the following DGPs for the global power properties of the tests:

- ALT1: $\mathbb{E}[Y_t|X_t] = \text{Probit}(1 + X_t + X_t^4)$ and $X_t \sim \text{IID } N(0,1)$;
- ALT2: $\mathbb{E}[Y_t|X_t] = \text{Logit}[-(1 + X_t)]$ and $X_t \sim \text{IID } N(0,1)$,

where 'Logit(x)' denotes $\{1 + \exp(-x)\}^{-1}$. ALT1 contains a nonlinear component X_t^4, so that the linear probit model is misspecified. ALT2 is a linear logit process, and although no nonlinear component is involved, the linear probit model is functionally misspecified. Third, the local power properties of the tests are examined by means of the following DGPs:

- LOC1: $\mathbb{E}[Y_t|X_t] = \text{Probit}(1 + X_t + n^{-1/2}X_t^4)$ and $X_t \sim \text{IID } N(0,1)$;
- LOC2: $\mathbb{E}[Y_t|X_t] = (1 - n^{-1/2})\text{Probit}(1 + X_t) + n^{-1/2}\text{Logit}[-(1 + X_t)]$ and $X_t \sim \text{IID } N(0,1)$.

LOC1 and LOC2 are considered as local processes of ALT1 and ALT2, respectively. If the sample size is finite and small, LOC1 and LOC2 are better approximated by the null DGP than ALT1 and ALT, respectively. LOC2 is generated as a mixture of the probit and logit distributions. We examine these two DGPs to investigate the local powers of the tests. We also apply the parametric bootstrap as before and compare Chesher's (1983) and Lancaster's (1984) information matrix test to our tests.

The null simulation results are contained in Table 10. The bootstrap repetitions are 500, and we report the empirical rejection rates in Table 10 using 5,000 replications, as before. The nominal level of significance is 5%. From these simulation results, the empirical nominal levels of all tests are very accurate, just as for the normal linear model case.

The global power simulation results are given in Tables 11 and 12. The bootstrap repetitions were 500 and we used 2,000 replications. The same nominal 5% significance level was used.

We summarize the global power simulation results as follows. First, all tests including the information matrix test are consistent, and the overall performance of our tests are better than the information matrix test. Second, the rejection rates of the tests are DGP-dependent. For ALT1, the empirical rejection rates of all the tests approach unity very quickly as the sample size increases. On the other hand, the convergence rates are very slow for ALT2. This is due to the fact that the logit and probit probability functions are very similar to each other. Unless the sample size is very large, it is hard to distinguish them. Third, it is hard to corroborate Theorem 2(*i*) using ALT1 as the tests approach unity very quickly. On the other hand, ALT2

shows that the rejection rates of the tests indexed by (2) and (3) are higher than those indexed by (1). These results affirm Theorem 2(i) as for the linear normal model case.

The local power simulation results are contained in Tables 13 and 14. The bootstrap repetitions are 500, and 3,000 replications were used. Again, the nominal significane level was 5%.

We summarize the local power simulation results as follows. First, there is a different power relationship among the tests. For LOC1, the locally most powerful tests are our tests indexed by (2) and (3), and the next is the information matrix test. Our tests indexed by (1) have the lowest local power. For LOC2, the locally most powerful tests are our tests indexed by (2) and (3), overall. Next are the tests indexed by (1), and the information matrix test has the lowest local power. Second, the simulations affirm Theorem 3. For both LOC1 and LOC2, our tests indexed by (2) and (3) show more or less similar empirical rejection rates, and those indexed by (1) also show similar rejection rates. This is because the tests indexed by (2) and (3) are equivalent tests, and so are those indexed by (1). This outcome is predicted by Theorem 3.

5 Empirical Applications

In the political economy and political science literature, a longstanding research question is to explain voting turnout. For example, in the seminal study of Wolfinger and Rosenstone (1980) on voting behavior, the authors estimated a voting turnout model using 1972 presidential election data. In later work in economics, Feddersen and Pesendorfer (1996) provided an economic model for voting turnout based on the asymmetric information of voters. In addition to these contributions, many research papers have attempted to explain voting turnout using empirical analysis and economic theory (*e.g.*, Nagler, 1991, 1994; Bénabou 2000; Besley and Case, 2003; Berry, DeMeritt, and Esarey, 2010, among others).

Wolfinger and Rosenstone's (1980) empirical model has been particularly influential in this literature. Using the level of education, (*Education*), the squared level of education (*Education*^{2}), age (*Age*), squared age (*Age*^{2}), a dummy for the South (*South*), a dummy for the presence of a gubernatorial election in the state (*Gubernatorial Election*), and the number of days before the election that registration closes (*Closing Date*), they estimate a linear probit model and find that it is the registration requirement in voting law that most severely affects the least educated group. Here, *Closing date* is used as a measure for the voting law requirement. Specifically, their model estimation shows that if *Closing Date* were hypothetically set to zero, the average voting turnout shows the greatest increase for the least educated group, whereas the increase is least for the most educated group turnout. Wolfinger and Rosenstone (1980) explain this in terms of the cost of voting: more educated people pay a lower cost for understanding the implications of complex and

abstract political issues. This finding is now regarded as a stylized fact in the political economic and science literature.

Nagler (1991) points out that Wolfinger and Rosenstone's (1980) central empirical result is an artifact of the probit model methodology. The probit probability is most highly affected if the explanatory variable is around zero. In other words, the first-order derivative is greater at this level than at any other level of the explanatory variable. Therefore, if *Education* is near zero, the predicted probability level increase is greater than any group with higher education. So, the empirical finding of Wolfinger and Rosenstone (1980) result is simply an artifact of how the probit model manages the impact of data rather than a meaningful empirical finding. The same claim holds for a logit specification. To remedy this problem, Nagler (1991) estimates another probit model with two additional explanatory variables: *Closing Date* $\times$ *Education* and *Closing Date* $\times$ *Education*^{2}. He includes them to capture the interactive effects of *Closing Date* and *Education* to the turnout. From estimation of this model using 1972 and 1984 presidential election data, Nagler (1991) rejects Wolfinger and Rosenstone's (1980) empirical result.

Nagler (1994) attempts to improve these findings. Instead of specifying a probit model, he specifies a scobit model that assumes

$$\mathbb{P}(Y_t = 1|\mathbf{X}_t) = 1 - \frac{1}{(1 + \exp(\mathbf{X}_t'\boldsymbol{\beta}_*))^{\alpha_*}},$$

where Y_t is a dummy for voting, and $\mathbf{X}_t$ is a vector of explanatory variables. This specification follows a Burr type 10 distribution, for which the logit distribution is a special case obtained by setting $\alpha_* = 1$. If $\alpha_* \neq 1$, the distribution is skewed and so the model is also called the skewed-logit model. Using estimates of this model, Nagler (1991) modifies the earlier claim and finds that, under the skewed probability model, the interactive terms are not significant for the 1984 presidential election data and confirms Wolfinger and Rosenstone's (1980) earlier finding that the least educated group **is** more severely affected by the voting law requirement. He also argues that the scobit model is particularly useful in allowing for misspecification in logit or probit models.

Nagler's (1994) empirical conclusion may still be misleading and for the same reason. As Nagler (1991) points out, the probit model advocated by Wolfinger and Rosenstone (1980) can yield biased empirical findings because it is misspecified. In the same way, if the scobit model is misspecified, a similar critique applies. Although use of the skewness measure (α_*) generalizes the logit model, the model can easily be misspecified by other factors. As Berry, DeMeritt, and Esarey (2010) discuss, among the various features of probit and logit models for modeling voting turnout, use of the correct model assumptions and specification for the data is critical in reaching the conclusion of Nagler (1994).

Against this background, our tests of model specification provide a relevant new methodology to assess whether these empirical models of voting turnout are correctly specified or not. Using 1984 US presidential election data from Altman and McDonald (2003), we estimate the same model considered by Wolfinger and Rosenstone (1980) and Nagler (1991, 1994). The results are given in Table 15. Probit models without and with interactive terms are estimated by following Wolfinger and Rosenstone (1980) and Nagler (1991), respectively. The same model is also considered by Berry, DeMeritt, and Esarey (2010). Logit and scobit models without and with interactive terms are also estimated by following Nagler (1994). All the estimated parameters are similar to those in the literature. The only difference is that the p-values of the t-test statistics are computed by robust standard errors using the method in White (1980). These are provided in parentheses. The same table also provides the test results and their p-values. All the specification test statistics test the validity of the information matrix equality, and these are computed using the methodology of Section 4.2.

The findings in Table 15 can be summarized as follows. First, all empirical models for voting turnout appear to be misspecified, even though they have significantly dominated the empirical literature for some time. All the tests $\widehat{\mathfrak{B}}_n^{(1)}$, $\widehat{\mathfrak{B}}_n^{(2)}$, $\widehat{\mathfrak{B}}_n^{(3)}$, $\widehat{\mathfrak{D}}_n^{(1)}$, $\widehat{\mathfrak{D}}_n^{(2)}$, $\widehat{\mathfrak{D}}_n^{(3)}$, $\widehat{\mathfrak{S}}_n^{(1)}$, $\widehat{\mathfrak{S}}_n^{(2)}$, and $\widehat{\mathfrak{S}}_n^{(3)}$ reject the information matrix equality. None of the test p-values differ from zero, which implies that the conditional distributions of voting turnout that are assumed in these models are all misspecified. The scobit models appear to do best and have greater log-likelihood values than the probit and logit models. But the allowance for a skewed distribution is not enough to eliminate model misspecification. Second, the interactive terms in the scobit model are not statistically significant. The p-values of the interactive terms are 0.2283 and 0.6378, and this finding corresponds with Nagler (1994), although correct specification was assumed in reaching that conclusion in Nagler's work. Nevertheless, the outcome might be different if the correct model specification was used for voting turnout. In other words, the empirical findings obtained using the scobit specification may well be as misleading as for those from the probit and logit models.

Inferences drawn from these models in empirical work are inevitably approximate and the quality of the approximation depends on the scope of the model, its capability in testing the validity of a theory, and on the relevance of the model to the data. If the empirical model has a sufficiently flexible form that enables adequate estimation of the core part of the relevant theory, we may be able to exploit the model scope to test the theory within the framework of quasi-maximum likelihood estimation, as pointed out by Berry, DeMeritt, and Esarey (2010). The tests given here help to point to weaknesses in specification that may be repaired by the use of more flexible models with greater scope for empirical relevance.

6 Conclusion

The information matrix equality is a fundamental feature of correct specification in likelihood based econometric work. We provide a new methodology for testing such equality in empirical applications. Our approach is embedded in the general framework of testing the equality of two positive-definite matrices. The new approach improves earlier analytic attempts to control size in information matrix equality testing and delivers a class of test procedures that are easily implemented in practical work. The test mechanism relies on a simple characterization of equality between two k dimensional positive-definite matrices $\mathbf{A}$ and $\mathbf{B}$ involving only the traces of the two matrices $\mathbf{AB}^{-1}$ and $\mathbf{BA}^{-1}$, which greatly facilitates practical use and leads to a group of omnibus test statistics for equality of covariance matrices.

Asymptotic theory for these tests under null, local, and global alternatives are obtained under mild regularity conditions that support wide use of these procedures in empirical work. Simulation evidence affirms that good size control is obtained and test power in specification testing against various alternatives is generally strong, but power can be dominated in some cases by specific testing procedures such as those based on direct tests for Gaussianity. The methods of specification testing based on the information matrix equality are well illustrated in the commonly occurring cases of logit and probit models. Empirical application of these methods to voting turnout models show that classic models used in this literature all seem to suffer from specification failure, putting some of the empirical conclusions in the literature about voting turnout behavior at risk.

7 Technical Appendix and Proofs

7.1 Preliminary Lemmas

Before proving the main claims in the paper, we provide the following preliminary lemmas.

Lemma A1. *Given Assumption A, if for some $d_* > 0$, $\mathbf{B}_* = d_*\mathbf{A}_*$,*

$$\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*] + d_*^2\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*^{-1}] = 2d_*[\mathrm{tr}[\mathbf{R}_{j,*}\mathbf{R}_{i,*}]]. \qquad \square$$

Lemma A2. *Given Assumption B and $\mathcal{H}_\ell$,*

(i) $\mathbf{A}_{*,n}^{-1} = \mathbf{A}_*^{-1} - n^{-1/2}\mathbf{C}_*\mathbf{A}_*^{-1} + n^{-1}\mathbf{C}_*^2\mathbf{A}_*^{-1} + O(n^{-3/2})$*;*

(ii) $\mathbf{B}_{*,n}^{-1} = \mathbf{B}_*^{-1} - n^{-1/2}\mathbf{F}_*\mathbf{B}_*^{-1} + n^{-1}\mathbf{F}_*^2\mathbf{B}_*^{-1} + O(n^{-3/2})$*;*

(iii) $\mathbf{U}_{a,n} = \mathbf{U}_{o,n} - n^{-1/2}\mathbf{C}_*\mathbf{U}_{o,n} + O_{\mathbb{P}}(n^{-3/2})$*;*

(iv) $\mathbf{W}_{a,n} = \mathbf{W}_{o,n} - n^{-1/2}\mathbf{F}_*\mathbf{W}_{o,n} + O_{\mathbb{P}}(n^{-3/2})$;

(v) $\mathbf{A}_{*,n}^{-1}\mathbf{B}_{*,n} = \mathbf{I} + n^{-1/2}\mathbf{V}_* - n^{-1}\mathbf{C}_*\mathbf{V}_* + O(n^{-3/2})$;

(vi) $\mathbf{B}_{*,n}^{-1}\mathbf{A}_{*,n} = \mathbf{I} - n^{-1/2}\mathbf{V}_* + n^{-1}\mathbf{F}_*\mathbf{V}_* + O(n^{-3/2})$;

(vii) $\mathbf{P}_{a,n} = \mathbf{P}_{o,n} - n^{-1/2}(\mathbf{F}_*\mathbf{W}_{o,n} - \mathbf{C}_*\mathbf{U}_{o,n}) + O_{\mathbb{P}}(n^{-3/2})$;

(viii) $\mathbf{B}_{*,n}^{-1}\partial_j\mathbf{B}_{*,n} = \mathbf{B}_*^{-1}\partial_j\mathbf{B}_* + n^{-1/2}(\mathbf{B}_*^{-1}\partial_j\bar{\mathbf{B}}_* - \mathbf{F}_*\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*) + O(n^{-1})$;

(ix) $\mathbf{A}_{*,n}^{-1}\partial_j\mathbf{A}_{*,n} = \mathbf{A}_*^{-1}\partial_j\mathbf{A}_* + n^{-1/2}(\mathbf{A}_*^{-1}\partial_j\bar{\mathbf{A}}_* - \mathbf{C}_*\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*) + O(n^{-1})$;

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

(xi) $\mathbf{L}_{a,n} = \mathbf{L}_{o,n} - n^{-1/2}\{(\mathbf{F}_*\mathbf{W}_{o,n} - \mathbf{C}_*\mathbf{U}_{o,n}) - \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})(\mathbf{Q}_{j,*} - (\mathbf{F}_*\mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{C}_*\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*))\} + O_{\mathbb{P}}(n^{-3/2})$; *and*

(xii) $\mathbf{L}_{a,n}\mathbf{A}_{*,n}^{-1}\mathbf{B}_{*,n} = \mathbf{L}_{o,n} - n^{-1/2}(\mathbf{F}_*\mathbf{W}_{o,n} - \mathbf{C}_*\mathbf{U}_{o,n}) + n^{-1/2}\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})(\mathbf{Q}_{j,*} - (\mathbf{F}_*\mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{C}_*\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)) + n^{-1/2}\mathbf{L}_{o,n}(\mathbf{F}_* - \mathbf{C}_*) - n^{-1/2}(\mathbf{F}_*\mathbf{W}_{o,n} - \mathbf{C}_*\mathbf{U}_{o,n}) + O_{\mathbb{P}}(n^{-3/2})$. □

Lemma A3. *Given Assumption B and* $\mathcal{H}_\ell$,

(i) $k^{-1}\mathrm{tr}[\mathbf{B}_{*,n}^{-1}\mathbf{A}_{*,n}] = 1 - n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{V}_*] + n^{-1}k^{-1}\mathrm{tr}[\mathbf{F}_*\mathbf{V}_*] + O(n^{-3/2})$;

(ii) $(k^{-1}\mathrm{tr}[\mathbf{B}_{*,n}^{-1}\mathbf{A}_{*,n}])^2 = 1 - 2n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{V}_*] + 2(nk)^{-1}\mathrm{tr}[\mathbf{F}_*\mathbf{V}_*] + n^{-1}k^{-2}\mathrm{tr}[\mathbf{V}_*]^2 + O(n^{-3/2})$;

(iii) $(k^{-1}\mathrm{tr}[\mathbf{B}_{*,n}^{-1}\mathbf{A}_{*,n}])^{-1} = 1 + n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{V}_*] - n^{-1}k^{-1}\mathrm{tr}[\mathbf{F}_*\mathbf{V}_*] + n^{-1}k^{-2}\mathrm{tr}[\mathbf{V}_*]^2 + O(n^{-3/2})$; *and*

(iv) $(k^{-1}\mathrm{tr}[\mathbf{B}_{*,n}^{-1}\mathbf{A}_{*,n}])^{-2} = 1 + 2n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{V}_*] - 2(nk)^{-1}\mathrm{tr}[\mathbf{F}_*\mathbf{V}_*] + 3n^{-1}k^{-2}\mathrm{tr}[\mathbf{V}_*]^2 + O(n^{-3/2})$. □

Lemma A4. *Given Assumption B and* $\mathcal{H}_\ell$,

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

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

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

(iv) $\det[\mathbf{D}_{*,n}] = 1 + n^{-1/2}\mathrm{tr}[\mathbf{V}_*] + \frac{1}{2n}(\mathrm{tr}[\mathbf{V}_*]^2 + \mathrm{tr}[\mathbf{C}_*^2] - \mathrm{tr}[\mathbf{F}_*^2]) + O(n^{-3/2})$;

(v) $\det[\mathbf{D}_{*,n}]^{1/k} = 1 + \frac{1}{\sqrt{nk}}\mathrm{tr}[\mathbf{V}_*] + \frac{1}{2nk}(\mathrm{tr}[\mathbf{C}_*^2] - \mathrm{tr}[\mathbf{F}_*^2]) + \frac{1}{2nk^2}\mathrm{tr}[\mathbf{V}_*]^2 + O(n^{-3/2})$; *and*

(vi) $\det[\mathbf{D}_{*,n}]^{1/k}\mathrm{tr}[\mathbf{L}_{a,n}] = \frac{1}{k}\mathrm{tr}[\mathbf{K}_{o,n}] + \frac{1}{\sqrt{nk^2}}\mathrm{tr}[\mathbf{V}_*]\mathrm{tr}[\mathbf{K}_{o,n}] - \frac{1}{\sqrt{nk}}\mathrm{tr}[\mathbf{F}_*\mathbf{W}_{o,n} - \mathbf{C}_*\mathbf{U}_{o,n}] + \frac{1}{\sqrt{nk}}[\mathrm{tr}[\mathbf{Q}_{j,*} - (\mathbf{F}_*\mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{C}_*\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)]]'(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*) + O(n^{-3/2})$. □

Lemma A5. *Given Assumption B and* $\mathcal{H}_\ell$,

(i) $\widehat{T}_{o,n} = k^{-1}\mathrm{tr}[\mathbf{K}_{o,n}] + O_{\mathbb{P}}(n^{-1})$;

(ii) $\widehat{D}_{o,n} = k^{-1}\mathrm{tr}[\mathbf{K}_{o,n}] + O_{\mathbb{P}}(n^{-1})$;

(iii) $\widehat{H}_{o,n} = k^{-1}\mathrm{tr}[\mathbf{K}_{o,n}] + O_{\mathbb{P}}(n^{-1})$;

(iv) $\widehat{C}_{o,n} = k^{-1}\mathrm{tr}[\mathbf{K}_{o,n}^2] - (k^{-1}\mathrm{tr}[\mathbf{K}_{o,n}])^2 + o_{\mathbb{P}}(n^{-1})$; *and*

(v) $\widehat{G}_{o,n} = (2k)^{-1}\mathrm{tr}[\mathbf{K}_{o,n}^2] - (2k^2)^{-1}\mathrm{tr}[\mathbf{K}_{o,n}]^2 + o_{\mathbb{P}}(n^{-1})$. □

Before proving the preliminary lemmas, we note that $\mathbf{P}_{o,n} = \mathbf{M}_{o,n}$ and $\mathbf{R}_{j,*} = \mathbf{S}_{j,*}$ under $\mathcal{H}_\ell$, so that $\mathbf{L}_{o,n} = \mathbf{K}_{o,n}$.

Proof of Lemma A1: By lemma A5(*i*) of CW and the fact that $\mathbf{A}_*^{-1}\mathbf{B}_* = d_*\mathbf{I}$,

$$\begin{aligned}\partial_{ji}^2 \mathrm{tr}[\mathbf{D}_*] &= \mathrm{tr}[\mathbf{A}_*^{-1}\mathbf{B}_*\{(\mathbf{B}_*^{-1}\partial_{ji}^2\mathbf{B}_* - \mathbf{A}_*^{-1}\partial_{ji}^2\mathbf{A}_*) - (\mathbf{R}_{j,*}\mathbf{A}_*^{-1}\partial_i\mathbf{A}_* + \mathbf{R}_{i,*}\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)\}] \\ &= d_*\mathrm{tr}[(\mathbf{B}_*^{-1}\partial_{ji}^2\mathbf{B}_* - \mathbf{A}_*^{-1}\partial_{ji}^2\mathbf{A}_*) - (\mathbf{R}_{j,*}\mathbf{A}_*^{-1}\partial_i\mathbf{A}_* + \mathbf{R}_{i,*}\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)].\end{aligned}$$

The asymptotic expansion of $\partial_{ji}^2\mathrm{tr}[\mathbf{D}_*^{-1}]$ is also obtained by simply interchanging the roles of $\mathbf{A}_*$ and $\mathbf{B}_*$:

$$\begin{aligned}\partial_{ji}^2 \mathrm{tr}[\mathbf{D}_*^{-1}] &= \mathrm{tr}[\mathbf{B}_*^{-1}\mathbf{A}_*\{(\mathbf{A}_*^{-1}\partial_{ji}^2\mathbf{A}_* - \mathbf{B}_*^{-1}\partial_{ji}^2\mathbf{B}_*) + (\mathbf{R}_{j,*}\mathbf{B}_*^{-1}\partial_i\mathbf{B}_* + \mathbf{R}_{i,*}\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*)\}] \\ &= d_*^{-1}\mathrm{tr}[(\mathbf{A}_*^{-1}\partial_{ji}^2\mathbf{A}_* - \mathbf{B}_*^{-1}\partial_{ji}^2\mathbf{B}_*) + (\mathbf{R}_{j,*}\mathbf{B}_*^{-1}\partial_i\mathbf{B}_* + \mathbf{R}_{i,*}\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*)].\end{aligned}$$

Therefore, $\partial_{ji}^2\mathrm{tr}[\mathbf{D}_*] + d_*^2\partial_{ji}^2\mathrm{tr}[\mathbf{D}_*^{-1}] = 2d_*\mathrm{tr}[\mathbf{R}_{j,*}\mathbf{R}_{i,*}]$ by noting that $\mathbf{R}_{i,*} := \mathbf{B}_*^{-1}\partial_i\mathbf{B}_* - \mathbf{A}_*^{-1}\partial_i\mathbf{A}_*$ and $\mathbf{R}_{j,*} := \mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{A}_*^{-1}\partial_j\mathbf{A}_*$. ■

Proof of Lemma A2: (*i*) Note that $\mathbf{A}_{*,n}^{-1} = [\mathbf{I} - n^{-1/2}\mathbf{A}_*^{-1}(-\bar{\mathbf{A}}_*)]^{-1}\mathbf{A}_*^{-1}$. For large enough n, $[\mathbf{I} - n^{-1/2}\mathbf{A}_*^{-1}(-\bar{\mathbf{A}}_*)]^{-1} = \mathbf{I} - n^{-1/2}\mathbf{A}_*^{-1}\bar{\mathbf{A}}_* + n^{-1}\mathbf{A}_*^{-1}\bar{\mathbf{A}}_*\mathbf{A}_*^{-1}\bar{\mathbf{A}}_* + ...$, which implies that

$$\begin{aligned}\mathbf{A}_{*,n}^{-1} &= [\mathbf{I} - n^{-1/2}\mathbf{A}_*^{-1}(-\bar{\mathbf{A}}_*)]^{-1}\mathbf{A}_*^{-1} \\ &= \mathbf{A}_*^{-1} - n^{-1/2}\mathbf{A}_*^{-1}\bar{\mathbf{A}}_*\mathbf{A}_*^{-1} + n^{-1}\mathbf{A}_*^{-1}\bar{\mathbf{A}}_*\mathbf{A}_*^{-1}\bar{\mathbf{A}}_*\mathbf{A}_*^{-1} + O(n^{-3/2}) \\ &= \mathbf{A}_*^{-1} - n^{-1/2}\mathbf{C}_*\mathbf{A}_*^{-1} + n^{-1}\mathbf{C}_*^2\mathbf{A}_*^{-1} + O(n^{-3/2}).\end{aligned}$$

(*ii*) This follows from Lemma A2(*i*) and the symmetric structure between $\mathbf{A}_{*,n}$ and $\mathbf{B}_{*,n}$.

(*iii*) Note that $\mathbf{U}_{a,n} = \mathbf{A}_{*,n}^{-1}(\mathbf{A}_n - \mathbf{A}_{*,n}) = \mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_{*,n}) - n^{-1/2}\mathbf{C}_*\mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_{*,n}) + O_{\mathbb{P}}(n^{-3/2})$ by Lemma A2(*i*). Here, the right side is $\mathbf{U}_{o,n} - n^{-1/2}\mathbf{C}_*\mathbf{U}_{o,n} + O_{\mathbb{P}}(n^{-3/2})$ by the definition of $\mathbf{U}_{o,n}$.

(*iv*) This follows from Lemma A2(*iii*) and the symmetric structure between $\mathbf{A}_{*,n}$ and $\mathbf{B}_{*,n}$.

(*v*) Note that

$$\begin{aligned}\mathbf{A}_{*,n}^{-1}\mathbf{B}_{*,n} &= (\mathbf{A}_*^{-1} - n^{-1/2}\mathbf{C}_*\mathbf{A}_*^{-1} + n^{-1}\mathbf{C}_*^2\mathbf{A}_*^{-1} + O(n^{-3/2}))(\mathbf{B}_* + n^{-1/2}\bar{\mathbf{B}}_*) \\ &= \mathbf{I} + n^{-1/2}(\mathbf{A}_*^{-1}\bar{\mathbf{B}}_* - \mathbf{A}_*^{-1}\bar{\mathbf{A}}_*) + n^{-1}(\mathbf{C}_*^2\mathbf{A}_*^{-1}\mathbf{B}_* - \mathbf{C}_*\mathbf{A}_*^{-1}\bar{\mathbf{B}}_*) + O(n^{-3/2}) \\ &= \mathbf{I} + n^{-1/2}(\mathbf{B}_*^{-1}\bar{\mathbf{B}}_* - \mathbf{A}_*^{-1}\bar{\mathbf{A}}_*) - n^{-1}\mathbf{C}_*(\mathbf{B}_*^{-1}\bar{\mathbf{B}}_* - \mathbf{C}_*) + O(n^{-3/2})\end{aligned}$$

by Lemma A2(*i*) and the definition of $\mathbf{B}_{*,n}$. We now note that $\mathbf{V}_* := \mathbf{B}_*^{-1}\bar{\mathbf{B}}_* - \mathbf{A}_*^{-1}\bar{\mathbf{A}}_* = \mathbf{B}_*^{-1}\bar{\mathbf{B}}_* - \mathbf{C}_*$.

(*vi*) This follows from Lemma A2(*v*) and the symmetric structure between $\mathbf{A}_{*,n}$ and $\mathbf{B}_{*,n}$.

(*vii*) This follows from the definition of $\mathbf{P}_{a,n} := \mathbf{W}_{a,n} - \mathbf{U}_{a,n}$ and Lemmas A2(*iii* and *iv*).

(*viii*) By Lemma A2(*ii*),

$$\begin{aligned}\mathbf{B}_{*,n}^{-1}\partial_j\mathbf{B}_{*,n} &= (\mathbf{B}_*^{-1} - n^{-1/2}\mathbf{F}_*\mathbf{B}_*^{-1} + n^{-1}\mathbf{F}_*^2\mathbf{B}_*^{-1} + O(n^{-3/2}))\partial_j(\mathbf{B}_* + n^{-1/2}\bar{\mathbf{B}}_*) \\ &= \mathbf{B}_*^{-1}\partial_j\mathbf{B}_* + n^{-1/2}(\mathbf{B}_*^{-1}\partial_j\bar{\mathbf{B}}_* - \mathbf{F}_*\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*) + O(n^{-1}).\end{aligned}$$

(*ix*) We can apply the proof of Lemma A2(*viii*).

(*x*) Apply Lemmas A2(*viii* and *ix*) to obtain the desired result.

(*xi*) By Lemmas A2(*vii* and *ix*),

$$\begin{aligned}\mathbf{L}_{a,n} = \mathbf{P}_{o,n} &+ \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})(\mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{A}_*^{-1}\partial_j\mathbf{A}_*) - n^{-1/2}(\mathbf{F}_*\mathbf{W}_{o,n} - \mathbf{C}_*\mathbf{U}_{o,n}) \\ &+ n^{-1/2}\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})(\mathbf{Q}_{j,*} - (\mathbf{F}_*\mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{C}_*\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)) + O_{\mathbb{P}}(n^{-3/2}).\end{aligned}$$

The desired result follows from the definition of $\mathbf{L}_{o,n}$.

(*xii*) We combine Lemmas A2(*v* and *xi*) and collect the terms according to their convergence rates. This completes the proof. ■

Proof of Lemma A3: (*i*) This immediately follows from Lemma A2(*vi*).

(*ii*) This immediately follows from Lemma A2(*vi*).

(*iii*) Taylor expansion of $1/x$ at $x = 1$ gives $1/x = 1 - (x-1) + (x-1)^2 - (x-1)^3 + ...$ We now let x be $k^{-1}\mathrm{tr}[\mathbf{D}_{*,n}^{-1}]$ and use Lemma A3(*i*). If the terms are rearranged according to their convergence rates, the desired result follows.

(*iv*) This immediately follows from Lemma A3(*iii*). ■

Proof of Lemma A4: (*i*) By the proof of lemma A2 (*i*) of CW,

$$\begin{aligned}\det[\mathbf{A}_n] - \det[\mathbf{A}_*] &= \det[\mathbf{A}_*]\mathrm{tr}[\mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_*)] \\ &+ \frac{1}{2}\det[\mathbf{A}_*]\{\mathrm{tr}[\mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_*)]^2 - \mathrm{tr}[\mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_*)\mathbf{A}_*^{-1}(\mathbf{A}_n - \mathbf{A}_*)]\} + O_{\mathbb{P}}(n^{-3/2}).\end{aligned}$$

We now simply let $\mathbf{A}_n$ be $\mathbf{A}_{*,n}$ and note that $\mathbf{C}_* = \mathbf{A}_*^{-1}(\mathbf{A}_{*,n} - \mathbf{A}_*) = \mathbf{A}_*^{-1}\bar{\mathbf{A}}_*$ under $\mathcal{H}_\ell$. This yields the

desired result.

(*ii*) This immediately follows from Lemma A4(*i*) and the symmetric structure between $\mathbf{A}_{*,n}$ and $\mathbf{B}_{*,n}$.

(*iii*) Lemma A2(*iii*) of CW shows that $\det[\mathbf{A}_n]^{-1} - \det[\mathbf{A}_*]^{-1} = -\det[\mathbf{A}_*]^{-1}(\mathrm{tr}[\mathbf{U}_n] + \frac{1}{2}\mathrm{tr}[\mathbf{U}_n]^2 - \frac{1}{2}\mathrm{tr}[\mathbf{U}_n^2]) + O_{\mathbb{P}}(n^{-1})$. Under $\mathcal{H}_\ell$, $\mathbf{U}_n = \mathbf{C}_*$. If we further let their $\mathbf{A}_n$ be $\mathbf{A}_{*,n}$, then

$$\det[\mathbf{A}_{*,n}]^{-1} - \det[\mathbf{A}_*]^{-1} = -\det[\mathbf{A}_*]^{-1}\{\mathrm{tr}[\mathbf{A}_*^{-1}(\mathbf{A}_{*,n} - \mathbf{A}_*)] + \frac{1}{2}\mathrm{tr}[\mathbf{A}_*^{-1}(\mathbf{A}_{*,n} - \mathbf{A}_*)]^2 - \frac{1}{2}\mathrm{tr}[\mathbf{A}_*^{-1}(\mathbf{A}_{*,n} - \mathbf{A}_*)^2]\} + O_{\mathbb{P}}(n^{-3/2}).$$

The desired result follows by noting that $\mathbf{C}_* = \mathbf{A}_*^{-1}(\mathbf{A}_{*,n} - \mathbf{A}_*) = \mathbf{A}_*^{-1}\bar{\mathbf{A}}_*$.

(*iv*) Note that

$$\begin{aligned}&\det[\mathbf{D}_{*,n}] = \det[\mathbf{A}_{*,n}]^{-1}\det[\mathbf{B}_{*,n}] \\ &= \left\{1 + \frac{1}{\sqrt{n}}\mathrm{tr}[\mathbf{F}_*] + \frac{1}{2n}(\mathrm{tr}[\mathbf{F}_*]^2 - \mathrm{tr}[\mathbf{F}_*^2])\right\}\left\{1 - \frac{1}{\sqrt{n}}\mathrm{tr}[\mathbf{C}_*] + \frac{1}{2n}(\mathrm{tr}[\mathbf{C}_*]^2 + \mathrm{tr}[\mathbf{C}_*^2])\right\} + O(n^{-3/2}),\end{aligned}$$

where the second equality follows from Lemmas A4(*ii* and *iii*) and the fact that $\det[\mathbf{D}_*] = 1$ under $\mathcal{H}_\ell$. Thus,

$$\det[\mathbf{D}_{*,n}] = 1 + \frac{1}{\sqrt{n}}\mathrm{tr}[\mathbf{F}_* - \mathbf{C}_*] + \frac{1}{2n}(\mathrm{tr}[\mathbf{F}_* - \mathbf{C}_*]^2 + \mathrm{tr}[\mathbf{C}_*^2] - \mathrm{tr}[\mathbf{F}_*^2]) + O(n^{-3/2}).$$

We further note that $\mathbf{V}_* := \mathbf{F}_* - \mathbf{C}_*$ to yield the result.

(*v*) Taylor expansion applied to $\det[\mathbf{D}_{*,n}]^{1/k}$ gives

$$\det[\mathbf{D}_{*,n}]^{1/k} = \det[\mathbf{D}_*]^{1/k} + \frac{1}{k}\det[\mathbf{D}_{*,n}]^{1/k-1}\{\det[\mathbf{D}_{*,n} - \det[\mathbf{D}_*]\} + \frac{1}{2k}\left(\frac{1}{k} - 1\right)\{\det[\mathbf{D}_{*,n} - \det[\mathbf{D}_*]\}^2 + \ldots. \quad (4)$$

Lemma A4(*iv*) implies that $\det[\mathbf{D}_{*,n}] - \det[\mathbf{D}_*] = \frac{1}{\sqrt{n}}\mathrm{tr}[\mathbf{V}_*] + \frac{1}{2n}(\mathrm{tr}[\mathbf{V}_*]^2 + \mathrm{tr}[\mathbf{C}_*^2] - \mathrm{tr}[\mathbf{F}_*^2]) + O(n^{-3/2})$ by noting that $\det[\mathbf{D}_*] = 1$ under $\mathcal{H}_\ell$. We now substitute this into (4) and arrange the terms according to their convergence rates. This yields the desired result.

(*vi*) To show this, we combine Lemmas A2(*xi*) and A3(*v*) and rearrange the terms according to their convergence rates. This completes the proof. ■

As Lemma A5 is immediately obtained by applying Corollary 1 and Lemma 4(*ii*), we omit the proof.

1.2 Proofs of the Main Results

Proof of Lemma 1: (*i*) If $\mathbf{A} = \mathbf{B}$, then clearly $\mathrm{tr}[\mathbf{D}] = \mathrm{tr}[\mathbf{A}^{-1}\mathbf{B}] = \mathrm{tr}[\mathbf{I}] = k$ and $\mathrm{tr}[\mathbf{D}^{-1}] = \mathrm{tr}[\mathbf{B}^{-1}\mathbf{A}] = \mathrm{tr}[\mathbf{I}] = k$. For the converse, note that $k^{-1}\sum_{j=1}^{k}\lambda_j = 1$, where λ_j is the j-th largest eigenvalue of $\mathbf{D}$ and so $\mathrm{tr}[\mathbf{D}] = k$. In addition, $k^{-1}\mathrm{tr}[\mathbf{D}^{-1}] = 1$ implies that $k^{-1}\sum_{j=1}^{k}\lambda_j^{-1} = 1$, so that the harmonic mean of the eigenvalues of $\mathbf{D}$ is 1. That is, the arithmetic mean of the eigenvalues is identical to the harmonic mean. Therefore, for some λ, $\lambda = \lambda_1 = \ldots = \lambda_k$. The given condition also implies that $\lambda = 1$. If we now let $\mathbf{C}$ be the orthonormal matrix of the eigenvectors of $\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}$, $\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2} = \mathbf{C}\mathbf{I}\mathbf{C}' = \mathbf{I}$. Therefore, $\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2} = \mathbf{I}$ implies $\mathbf{A}^{1/2}\mathbf{A}^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}\mathbf{A}^{1/2} = \mathbf{A}^{1/2}\mathbf{A}^{1/2}$, which simplifies to $\mathbf{B} = \mathbf{A}$.

(*ii*) We can combine Lemma 1(*i*) with lemma 1 of CW. ■

Proofs of Lemma 2 follow from lemma 4 of CW and Lemma 3 in our study. We thus omit its proof. Furthermore, Corollary 1 follows from Lemma 2. We now prove Lemma 3.

Proof of Lemma 3: (*i*) Lemma 4(*i*) of CW gives the expansion of $\mathrm{tr}[\widehat{\mathbf{B}}_n\widehat{\mathbf{A}}_n^{-1}]$. We apply this expansion to expand $k^{-1}\mathrm{tr}[\widehat{\mathbf{D}}_n^{-1}]$ by simply interchanging the roles of $\mathbf{A}_n$ and $\mathbf{B}_n$. That is,

$$\begin{aligned}\frac{1}{k}\mathrm{tr}[\widehat{\mathbf{D}}_n^{-1}] - \frac{1}{k}\mathrm{tr}[\mathbf{D}_*^{-1}] = -\frac{1}{k}\mathrm{tr}[\mathbf{L}_n\mathbf{B}_*^{-1}\mathbf{A}_*] + \frac{1}{k}\mathrm{tr}[\mathbf{L}_n\mathbf{W}_n\mathbf{B}_*^{-1}\mathbf{A}_*] \\ + \frac{1}{k}[\mathrm{tr}[(-\mathbf{J}_{j,n} + \mathbf{P}_n\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*)\mathbf{B}_*^{-1}\mathbf{A}_*]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \\ + \frac{1}{2k}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla_{\boldsymbol{\theta}}^2\mathrm{tr}[\mathbf{D}_*^{-1}](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1}). \quad (5)\end{aligned}$$

We also note that by Taylor expansion of $\frac{1}{x}$ yields that $\frac{1}{x} - \frac{1}{x_0} = -\frac{1}{x_0^2}(x - x_0) + \frac{1}{x_0^3}(x - x_0)^2 + \mathcal{R}$, where $\mathcal{R}$ is the remainder. We now let x and x_0 be $\frac{1}{k}\mathrm{tr}[\widehat{\mathbf{D}}_n^{-1}]$ and $\frac{1}{k}\mathrm{tr}[\mathbf{D}_*^{-1}]$, respectively and also note that $\widehat{H}_n = k/\mathrm{tr}[\widehat{\mathbf{D}}_n^{-1}] - 1$ and $H_* = k/\mathrm{tr}[\mathbf{D}_*^{-1}] - 1$. We finally arrange the terms according to their convergence rates to obtain the desired result.

(*ii*) If $\mathcal{H}_0$ further holds, $H_* = 0$, $\mathbf{B}_*^{-1}\mathbf{A}_* = \mathbf{I}$, $\mathbf{L}_n = \mathbf{K}_n$, $k^{-1}\mathrm{tr}[\mathbf{D}_*^{-1}] = 1$, and $\mathbf{P}_n = \mathbf{M}_n$. If all these equalities are applied to (5), the asymptotic expansion of $\widehat{H}_n$ reduces to the desired expansion. ■

Proof of Lemma 4: (*i*) Lemmas 4(*i.a* and *i.b*) immediately follow from Lemma 2(*i*, *ii*, and *iii*).

(*ii*) (*ii.a*) From the fact that $\mathbf{B}_* = d_*\mathbf{A}_*$, it follows that $\mathrm{tr}[\mathbf{D}_*^{-1}] = k/d_*$, $\mathbf{D}_* = d_*\mathbf{I}$, and $\mathbf{D}_*^{-1} = d_*^{-1}\mathbf{I}$.

We now substitute these into $\widehat{H}_n$ in Lemma 3 and obtain

$$\widehat{H}_n = d_* - 1 + d_* k^{-1}\mathrm{tr}[\mathbf{L}_n] + d_*(k^{-1}\mathrm{tr}[\mathbf{L}_n])^2 - d_* k^{-1}[\mathrm{tr}[(-\mathbf{J}_{j,n} + \mathbf{P}_n\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*)]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \\ - \frac{d_*}{k}\mathrm{tr}[\mathbf{L}_n\mathbf{W}_n] - \frac{d_*^2}{2k}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*^{-1}](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1}). \quad (6)$$

In the same way, we substitute $\mathrm{tr}[\mathbf{D}_*^{-1}] = k/d_*$, $\mathbf{D}_* = d_*\mathbf{I}$, and $\mathbf{D}_*^{-1} = d_*^{-1}\mathbf{I}$ into (2) and obtain

$$\widehat{T}_n = d_* - 1 + d_* k^{-1}\mathrm{tr}[\mathbf{L}_n] - d_* k^{-1}\mathrm{tr}[\mathbf{L}_n\mathbf{U}_n] \\ + \frac{d_*}{k}[\mathrm{tr}[\mathbf{J}_{j,n} - \mathbf{P}_n\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*]]'(\widehat{\boldsymbol{\theta}}_n\boldsymbol{\theta}_*) + \frac{1}{2k}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1}). \quad (7)$$

Therefore, the asymptotic expansion of $\widehat{C}_n$ is obtained as

$$\widehat{C}_n := \widehat{T}_n - \widehat{H}_n = d_* k^{-1}\mathrm{tr}[\mathbf{L}_n\mathbf{P}_n] + d_* k^{-1}[\mathrm{tr}[\mathbf{P}_n\mathbf{R}_{j,*}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) - d_* k^{-2}\mathrm{tr}[\mathbf{L}_n]^2 \\ + \frac{1}{2k}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\left\{\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*] + d_*^2\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*^{-1}]\right\}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1}). \quad (8)$$

Here, the definition of $\mathbf{P}_n := \mathbf{W}_n - \mathbf{U}_n$ is used to simplify the expression. Given this, note that Lemma A1 implies that $\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*] + d_*^2\nabla^2_{\boldsymbol{\theta}}\mathrm{tr}[\mathbf{D}_*^{-1} = 2d_*\mathrm{tr}[\mathbf{R}_{j,*}\mathbf{R}_{i,*}]$. Therefore,

$$\widehat{C}_n = d_* k^{-1}\mathrm{tr}[\mathbf{L}_n\mathbf{P}_n] + d_* k^{-1}[\mathrm{tr}[\mathbf{P}_n\mathbf{R}_{j,*}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) - d_* k^{-2}\mathrm{tr}[\mathbf{L}_n]^2 \\ + d_* k^{-1}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'[\mathrm{tr}[\mathbf{R}_{j,*}\mathbf{R}_{i,*}]](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1}).$$

We recall the definition of $\mathbf{L}_n := \mathbf{P}_n + \sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})\mathbf{R}_{j,*}$, and note that this implies

$$\widehat{C}_n = d_* k^{-1}\mathrm{tr}[\mathbf{P}_n^2] + 2d_* k^{-1}[\mathrm{tr}[\mathbf{P}_n\mathbf{R}_{j,*}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) - d_* k^{-2}\mathrm{tr}[\mathbf{L}_n]^2 \\ + d_* k^{-1}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'[\mathrm{tr}[\mathbf{R}_{j,*}\mathbf{R}_{i,*}]](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1}), \quad (9)$$

that is also equal to $d_* k^{-1}\mathrm{tr}[\mathbf{L}_n^2] - d_* k^{-2}\mathrm{tr}[\mathbf{L}_n]^2 + o_{\mathbb{P}}(n^{-1})$.

(*ii.b*) Note that corollary 5(*ii*) of CW shows that $\widehat{S}_n = -\frac{d_*}{2k^2}\mathrm{tr}[\mathbf{L}_n]^2 + \frac{d_*}{2k}\mathrm{tr}[\mathbf{L}_n^2] + o_{\mathbb{P}}(n^{-1})$. Also, $\widehat{G}_n := \widehat{C}_n - \widehat{S}_n$. Thus, $\widehat{G}_n = 2^{-1}d_*\{k^{-1}\mathrm{tr}[\mathbf{L}_n^2] - (k^{-1}\mathrm{tr}[\mathbf{L}_n])^2\} + o_{\mathbb{P}}(n^{-1})$ using (*ii.a*).

(*iii*) Given Lemma 4(*ii*), we let $d_* = 1$ and $\mathbf{L}_n = \mathbf{K}_n$ to complete the proof. ■

Theorem 1(*i*) follows as a corollary of theorem 1 of CW, and Theorems 1(*ii* and *iii*) also follow as corollaries of Corollary 1 and Lemma 4(*iii*). Corollary 2 is implied by Lemmas 3, 4, and lemma 5 of CW.

We now prove Theorem 2.

Proof of Theorem 2: The claim structures given for the statistics in Corollary 2(*i–vii*) are symmetric and similar. We therefore prove only the claim on $\widehat{\mathfrak{B}}_n^{(1)}$ in (*i*) and (*ii*) to save the space. The others are proved in a similar fashion.

(*i*) For $\widehat{\mathfrak{B}}_n^{(1)}$ to have the greatest leading term, it has to be greater than those $\widehat{\mathfrak{B}}_n^{(2)}$ and $\widehat{\mathfrak{B}}_n^{(3)}$. This implies that $\frac{1}{2}D_*^2 \geq \frac{1}{2}T_*^2 + 2(T_* - G_*)$ and $\frac{1}{2}T_*^2 \geq \frac{1}{2}D_*^2 + 2(T_* - G_*)$. These two inequalities hold only when $T_* = G_*$, so that all eigenvalues of $\mathbf{D}_*$ are identical. This implies that $\mathbf{B}_*$ is proportional to $\mathbf{A}_*$ and contradicts the assumption of Theorem 2.

(*ii*) For $\widehat{\mathfrak{B}}_n^{(2)}$ to have the greatest leading term, it has to be greater than those of the other tests. From this condition, we have the following 8 inequalities:

$$\frac{1}{2}T_*^2 + 2S_* \geq \frac{1}{2}D_*^2, \tag{10}$$

$$T_*^2 \geq D_*^2, \tag{11}$$

$$\frac{1}{2}T_*^2 + 2S_* \geq \frac{1}{2}H_*^2, \tag{12}$$

$$S_* \geq G_*, \tag{13}$$

$$T_*^2 + S_* \geq H_*^2 + G_*, \tag{14}$$

$$T_*^2 + 2S_* \geq \frac{1}{2}D_*^2 + \frac{1}{2}H_*^2, \tag{15}$$

$$T_*^2 + 2S_* \geq D_*^2 + 2G_*, \tag{16}$$

$$T_*^2 + 2S_* \geq H_*^2 + 2G_*. \tag{17}$$

Each inequality is obtained by letting the leading term of Corollary 2(*i.a*) be greater than the leading terms of Corollaries 2(*i.b–i.i*) and the fact that $C_* \equiv S_* + G_*$. These 8 inequalities are necessary for the desired condition.

Given this, note that (11), (13), and (14) are the conditions for $\widehat{\mathfrak{B}}_n^{(2)}$ to have the greatest leading term that are given by Theorem 2(*ii*). This proves sufficiency. For necessity, note that (11) implies (10); (13) and (14) imply (12); (11) and (12) imply (15); (11) and (13) imply (16); and (13) and (14) imply (17). Therefore, (11), (13), and (14) imply the other inequalities: (10), (12), (15), (16), and (17). From this, if $\widehat{\mathfrak{B}}_n^{(2)}$ has the greatest leading term, (11), (13), and (14) hold. This completes the proof. ■

Proof of Lemma 5: (*i*) We apply lemma 4(*i*) of CW and obtain the following expansion for $\widehat{T}_n$:

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

where $T_{*,n} := k^{-1}\mathrm{tr}[\mathbf{B}_{*,n}\mathbf{A}_{*,n}^{-1}] - 1$. We now use Lemma A2(*ii*, *v*, *vii*, and *xii*) for the first three terms and obtain

$$\begin{aligned}\widehat{T}_n = {} & \frac{1}{n^{1/2}k}\mathrm{tr}[\mathbf{V}_*] - \frac{1}{nk}\mathrm{tr}[\mathbf{C}_*\mathbf{V}_*] - \frac{1}{n^{1/2}k}\mathrm{tr}[(\mathbf{F}_*\mathbf{W}_{o,n} - \mathbf{C}_*\mathbf{U}_{o,n})] \\ & + \frac{1}{n^{1/2}k}\mathrm{tr}[\mathbf{L}_{o,n}\mathbf{V}_*] + \frac{1}{n^{1/2}k}[\mathrm{tr}[(\mathbf{Q}_{j,*} - (\mathbf{F}_*\mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{C}_*\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*))]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \\ & + \frac{1}{k}\mathrm{tr}[\mathbf{L}_{o,n}] + \frac{1}{k}[\mathrm{tr}[(\mathbf{J}_{j,o,n} - \mathbf{P}_{o,n}\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)\mathbf{A}_*^{-1}\mathbf{B}_*]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \\ & - \frac{1}{k}\mathrm{tr}[\mathbf{L}_{o,n}\mathbf{U}_{o,n}] + \frac{1}{2k}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla_{\boldsymbol{\theta}}^2\mathrm{tr}[\mathbf{D}_*](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1}). \quad (18)\end{aligned}$$

Note that $\mathbf{P}_{o,n} = \mathbf{M}_{o,n}$, $\mathbf{L}_{o,n} = \mathbf{K}_{o,n}$ under $\mathcal{H}_\ell$ and also that

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

This represents the last second to the last fourth terms in (18). Substituting $\widehat{T}_{o,n}$ into these terms completes the proof.

(*ii*) We apply Lemma 4(*ii*) of CW and obtain the following expansion for $\widehat{D}_n$:

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

where $D_{*,n} := \det[\mathbf{B}_{*,n}\mathbf{A}_{*,n}^{-1}]^{1/k} - 1$. We note that Lemma A3(*v*) implies that $D_{*,n} = \frac{1}{\sqrt{nk}}\mathrm{tr}[\mathbf{V}_*] + \frac{1}{2nk}(\mathrm{tr}[\mathbf{C}_*^2] - \mathrm{tr}[\mathbf{F}_*^2]) + \frac{1}{2nk^2}\mathrm{tr}[\mathbf{V}_*]^2 + O(n^{-3/2})$, and the asymptotic expansion of $\det[\mathbf{D}_{*,n}]^{1/k}\mathrm{tr}[\mathbf{L}_{a,n}]$ is

given by Lemma A3(*vi*). If we collect all these terms,

$$\begin{aligned}\widehat{D}_n = \frac{1}{\sqrt{n}k}\mathrm{tr}[\mathbf{V}_*] + \frac{1}{2nk}(\mathrm{tr}[\mathbf{C}_*^2] - \mathrm{tr}[\mathbf{F}_*^2]) + \frac{1}{2nk^2}\mathrm{tr}[\mathbf{V}_*]^2 + \frac{1}{k}\mathrm{tr}[\mathbf{K}_{o,n}] + \frac{1}{\sqrt{n}k^2}\mathrm{tr}[\mathbf{V}_*]\mathrm{tr}[\mathbf{K}_{o,n}] \\ - \frac{1}{\sqrt{n}k}\mathrm{tr}[\mathbf{F}_*\mathbf{W}_{o,n}] + \frac{1}{\sqrt{n}k}\mathrm{tr}[\mathbf{C}_*\mathbf{U}_{o,n}] + \frac{1}{2k}(\mathrm{tr}[\mathbf{M}_{o,n}]^2 + \mathrm{tr}[\mathbf{U}_{o,n}^2]) + \frac{1}{2k}\left(\frac{1}{k} - 1\right)\mathrm{tr}[\mathbf{K}_{o,n}]^2 \\ - \frac{1}{2k}\mathrm{tr}[\mathbf{W}_{o,n}^2] + \frac{1}{\sqrt{n}k}[\mathrm{tr}[\mathbf{Q}_{j,*} - (\mathbf{F}_*\mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{C}_*\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \\ + \frac{1}{k}[\mathrm{tr}[\mathbf{U}_{o,n}\mathbf{A}_*^{-1}\partial_j\mathbf{A}_* - \mathbf{W}_{o,n}\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + \frac{1}{k}\mathrm{tr}[\mathbf{M}_{o,n}][\mathrm{tr}[\mathbf{S}_{j,*}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \\ + \frac{1}{k}[\mathrm{tr}[\mathbf{J}_{j,o,n}]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + \frac{1}{2k}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla_{\boldsymbol{\theta}}^2\det[\mathbf{D}_*](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1}). \quad (19)\end{aligned}$$

This equation is derived by using the fact that $\mathbf{L}_{o,n} = \mathbf{K}_{o,n}$, $\mathbf{R}_{j,*} = \mathbf{S}_{j,*}$, and $\mathbf{P}_{o,n} = \mathbf{M}_{o,n}$ under $\mathcal{H}_\ell$. We now note the definition of $\widehat{D}_{o,n}$:

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

If the right-side terms of (19) that correspond to the definition of $\widehat{\mathbf{D}}_{o,n}$ are collected into $\widehat{\mathbf{D}}_{o,n}$, the desired result follows.

(*iii*) Note that Lemma 3(*i*) is simplified into

$$\begin{aligned}\widehat{H}_n = H_{*,n} + k^{-1}\mathrm{tr}[\mathbf{L}_{a,n}\mathbf{B}_{*,n}^{-1}\mathbf{A}_{*,n}]/(k^{-1}\mathrm{tr}[\mathbf{D}_{*,n}^{-1}])^2 \\ + (k^{-1}\mathrm{tr}[\mathbf{L}_{o,n}])^2 - k^{-1}\mathrm{tr}[\mathbf{L}_{o,n}\mathbf{W}_{o,n}] - k^{-1}[\mathrm{tr}[(-\mathbf{J}_{j,n} + \mathbf{P}_{o,n}\mathbf{B}_*^{-1}\partial_j\mathbf{B}_*)\mathbf{B}_*^{-1}\mathbf{A}_*]]'(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \\ - (2k)^{-1}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'\nabla_{\boldsymbol{\theta}}^2\mathrm{tr}[\mathbf{D}_*^{-1}](\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_{\mathbb{P}}(n^{-1}) \quad (20)\end{aligned}$$

under $\mathcal{H}_\ell$, where $H_{*,n} := (k^{-1}\mathrm{tr}[\mathbf{B}_{*,n}^{-1}\mathbf{A}_{*,n}])^{-1} - 1$. Given this, we further note that

$$\begin{aligned}k^{-1}\mathrm{tr}[\mathbf{L}_{a,n}\mathbf{B}_{*,n}^{-1}\mathbf{A}_{*,n}]/(k^{-1}\mathrm{tr}[\mathbf{D}_{*,n}])^2 = k^{-1}\mathrm{tr}[\mathbf{L}_{o,n}] - n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{F}_*\mathbf{W}_{o,n} - \mathbf{C}_*\mathbf{U}_{o,n}] \\ + n^{-1/2}k^{-1}\sum_{j=1}^{\ell}(\widehat{\theta}_{j,n} - \theta_{j,*})[\mathbf{Q}_{j,*} - (\mathbf{F}_*\mathbf{B}_*^{-1}\partial_j\mathbf{B}_* - \mathbf{C}_*\mathbf{A}_*^{-1}\partial_j\mathbf{A}_*)] \\ - n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{L}_{o,n}\mathbf{V}_*] + 2n^{-1/2}k^{-2}\mathrm{tr}[\mathbf{V}_*]\mathrm{tr}[\mathbf{L}_{o,n}] + o_{\mathbb{P}}(n^{-1}) \quad (21)\end{aligned}$$

using Lemmas A1(*vi*, *xi*) and A2(*iv*). If we substitute (21) into (20) and use Lemma A1(*vi*), the the desired result is obtained.

(*iv*) We now use Lemmas 5(*i* and *ii*) and compute $\widehat{S}_n$ by its definition. That is,

$$\widehat{S}_n := \widehat{T}_n - \widehat{D}_n = \widehat{S}_{o,n} + \frac{1}{2k}\left\{\frac{1}{n}\mathrm{tr}[\mathbf{V}_*^2] + \frac{2}{\sqrt{n}}\mathrm{tr}[\mathbf{K}_{o,n}\mathbf{V}_*] + \mathrm{tr}[\mathbf{K}_{o,n}^2]\right\} - \frac{1}{2k}\mathrm{tr}[\mathbf{K}_{o,n}^2]$$
$$- \frac{1}{2k^2}\left\{\frac{1}{n}\mathrm{tr}[\mathbf{V}_*]^2 + \frac{2}{\sqrt{n}}\mathrm{tr}[\mathbf{V}_*]\mathrm{tr}[\mathbf{K}_{o,n}] + \mathrm{tr}[\mathbf{K}_{o,n}]^2\right\} + \frac{1}{2k^2}\mathrm{tr}[\mathbf{K}_{o,n}]^2 + o_{\mathbb{P}}(n^{-1}).$$

Note further that $\frac{1}{n}\mathrm{tr}[\mathbf{V}_*^2] + \frac{2}{\sqrt{n}}\mathrm{tr}[\mathbf{K}_{o,n}\mathbf{V}_*] + \mathrm{tr}[\mathbf{K}_{o,n}^2] = \mathrm{tr}[(\mathbf{K}_{o,n} + n^{-1/2}\mathbf{V}_*)^2]$ and $\frac{1}{n}\mathrm{tr}[\mathbf{V}_*]^2 + \frac{2}{\sqrt{n}}\mathrm{tr}[\mathbf{V}_*]$ $\mathrm{tr}[\mathbf{K}_{o,n}] + \mathrm{tr}[\mathbf{K}_{o,n}]^2 = \mathrm{tr}[\mathbf{K}_{o,n} + n^{-1/2}\mathbf{V}_*]^2$. Using these facts, we obtain that

$$S_n = \widehat{S}_{o,n} + \frac{1}{2k}\{\mathrm{tr}[(\mathbf{K}_{o,n} + n^{-1/2}\mathbf{V}_*)^2] - \mathrm{tr}[\mathbf{K}_{o,n}^2]\} - \frac{1}{2k^2}\{\mathrm{tr}[\mathbf{K}_{o,n} + n^{-1/2}\mathbf{V}_*]^2 - \mathrm{tr}[\mathbf{K}_{o,n}]^2\} + o_{\mathbb{P}}(n^{-1}).$$

This is the desired result.

(*v*) Note that $\widehat{C}_n \equiv \widehat{T}_n - \widehat{H}_n$ and that the asymptotic approximations of $\widehat{T}_n$ and $\widehat{H}_n$ are provided in Lemmas 5(*i* and *ii*).

$$\widehat{C}_n = \widehat{T}_{o,n} - \widehat{H}_{o,n} + (nk)^{-1}\mathrm{tr}[\mathbf{C}_*^2 - 2\mathbf{C}_*\mathbf{F}_* + \mathbf{F}_*^2] + 2n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{K}_{o,n}\mathbf{V}_*]$$
$$- n^{-1}k^{-2}\mathrm{tr}[\mathbf{V}_*]^2 - 2n^{-1/2}k^{-2}\mathrm{tr}[\mathbf{V}_*]\mathrm{tr}[\mathbf{K}_{o,n}] + o_{\mathbb{P}}(n^{-1}). \quad (22)$$

Note that $\mathrm{tr}[\mathbf{C}_*^2 - 2\mathbf{C}_*\mathbf{F}_* + \mathbf{F}_*^2] = \mathrm{tr}[(\mathbf{F}_* - \mathbf{C}_*)^2] = \mathrm{tr}[\mathbf{V}_*^2]$. The desired result follows from this.

(*vi*) Note that $\widehat{G}_n \equiv \widehat{C}_n - \widehat{S}_n$. Furthermore, the asymptotic approximations of $\widehat{C}_n$ and $\widehat{S}_n$ are provided in Lemmas 5(*i* and *iii*). From these, it follows that

$$\widehat{G}_n = \widehat{C}_{o,n} - \widehat{S}_{o,n} + (2nk)^{-1}\mathrm{tr}[\mathbf{V}_*^2] - n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{V}_*]\mathrm{tr}[\mathbf{K}_{o,n}]$$
$$- (2nk^2)^{-1}\mathrm{tr}[\mathbf{V}_*]^2 + n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{K}_{o,n}\mathbf{V}_*] + o_{\mathbb{P}}(n^{-1}).$$

We finally note that $\widehat{C}_{o,n} - \widehat{S}_{o,n} = \widehat{G}_{o,n}$ to complete the proof. ■

Proof of Theorem 3: (*i*) Note that $\widehat{\mathfrak{B}}_n^{(1)} := \frac{nk}{4}(\widehat{T}_n^2 + \widehat{D}_n^2)$. Furthermore, the asymptotic approximations of

$\widehat{T}_n$ and $\widehat{D}_n$ are provided in Lemmas 5(*i* and *ii*). Therefore,

$$\begin{aligned}\widehat{\mathfrak{B}}_n^{(1)} &= \frac{nk}{4}\left\{\widehat{T}_{o,n}^2 + \widehat{D}_{o,n}^2 + \frac{2}{\sqrt{n}k}(\widehat{T}_{o,n}+\widehat{D}_{o,n})\mathrm{tr}[\mathbf{V}_*] + \frac{2}{nk^2}\mathrm{tr}[\mathbf{V}_*]^2\right\} + o_{\mathbb{P}}(1)\\ &= \frac{nk}{4}\left\{\frac{2}{k^2}\mathrm{tr}[\mathbf{K}_{o,n}]^2 + \frac{4}{\sqrt{n}k^2}\mathrm{tr}[\mathbf{K}_{o,n}]\mathrm{tr}[\mathbf{V}_*] + \frac{2}{nk^2}\mathrm{tr}[\mathbf{V}_*]^2\right\} + o_{\mathbb{P}}(1)\\ &= \frac{1}{2k}\left\{n\mathrm{tr}[\mathbf{K}_{o,n}]^2 + 2\sqrt{n}\mathrm{tr}[\mathbf{K}_{o,n}]\mathrm{tr}[\mathbf{V}_*] + \mathrm{tr}[\mathbf{V}_*]^2\right\} + o_{\mathbb{P}}(1)\\ &= \frac{1}{2k}\mathrm{tr}[\mathbf{V}_* + \sqrt{n}\mathbf{K}_{o,n}]^2 + o_{\mathbb{P}}(1),\end{aligned}$$

where the third to last equality holds by the definitions of $\widehat{T}_{o,n}$ and $\widehat{D}_{o,n}$. This shows the asymptotic approximation of $\widehat{\mathfrak{B}}_n^{(1)}$. We next note that $\widehat{\mathfrak{B}}_n^{(2)} := \frac{nk}{2}(\widehat{T}_n^2 + 2\widehat{S}_n)$. Therefore,

$$\begin{aligned}\widehat{\mathfrak{B}}_n^{(2)} &= \frac{nk}{2}\widehat{T}_{o,n}^2 + nk\widehat{S}_{o,n} + \sqrt{n}\mathrm{tr}[\mathbf{V}_*]\widehat{T}_{o,n} + \frac{1}{2k}\mathrm{tr}[\mathbf{V}_*]^2 + \frac{1}{2}\mathrm{tr}[(\mathbf{V}_* + \sqrt{n}\mathbf{K}_{o,n})^2]\\ &\quad - \frac{1}{2k}\mathrm{tr}[\mathbf{V}_* + \sqrt{n}\mathbf{K}_{o,n}]^2 + \frac{n}{2k}\mathrm{tr}[\mathbf{K}_{o,n}]^2 - \frac{n}{2}\mathrm{tr}[\mathbf{K}_{o,n}^2] + o_{\mathbb{P}}(1) = \frac{1}{2}\mathrm{tr}[(\mathbf{V}_* + \sqrt{n}\mathbf{K}_{o,n})^2] + o_{\mathbb{P}}(1),\end{aligned}$$

where the last equality holds by virtue of the definitions of $\widehat{T}_{o,n}$ and $\widehat{S}_{o,n}$.

Finally, the structure of $\widehat{\mathfrak{B}}_n^{(3)}$ is symmetric to that of $\widehat{\mathfrak{B}}_n^{(2)}$. In the same way, it follows that $\widehat{\mathfrak{B}}_n^{(3)} = \frac{1}{2}\mathrm{tr}[(\mathbf{V}_* + \sqrt{n}\mathbf{K}_{o,n})^2] + o_{\mathbb{P}}(1)$.

(*ii*) From Lemmas 5(*i* and *iii*), it follows that

$$\begin{aligned}\widehat{T}_n^2 &= (\widehat{T}_{o,n} + n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{V}_*])^2 + o_{\mathbb{P}}(n^{-1}) = (k^{-1}\mathrm{tr}[\mathbf{K}_{o,n}] + n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{V}_*] + O_{\mathbb{P}}(n^{-1}))^2 + o_{\mathbb{P}}(n^{-1})\\ &= (k^{-1}\mathrm{tr}[\mathbf{K}_{o,n} + n^{-1/2}\mathbf{V}_*]))^2 + o_{\mathbb{P}}(n^{-1}),\end{aligned} \tag{23}$$

and

$$\begin{aligned}\widehat{H}_n^2 &= (\widehat{H}_{o,n} + n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{V}_*])^2 + o_{\mathbb{P}}(n^{-1}) = (k^{-1}\mathrm{tr}[\mathbf{K}_{o,n}] + n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{V}_*] + O_{\mathbb{P}}(n^{-1}))^2 + o_{\mathbb{P}}(n^{-1})\\ &= (k^{-1}\mathrm{tr}[\mathbf{K}_{o,n} + n^{-1/2}\mathbf{V}_*]))^2 + o_{\mathbb{P}}(n^{-1}),\end{aligned} \tag{24}$$

where the second equality holds by Lemmas A5(*i* and *iii*). Given that $\widehat{\mathfrak{D}}_n^{(1)} := \frac{nk}{4}(\widehat{T}_n^2 + \widehat{H}_n^2)$, (23) and (24) imply that $\widehat{\mathfrak{D}}_n^{(1)} = \frac{1}{2k}\mathrm{tr}[\mathbf{V}_* + \sqrt{n}\mathbf{K}_{o,n}]^2 + o_{\mathbb{P}}(n^{-1})$, as desired.

From the definition $\widehat{\mathfrak{D}}_n^{(2)} := \frac{nk}{2}(\widehat{T}_n^2 + \widehat{C}_n)$, if we combine this with Lemma A5(*iv*) and (23), it follows

that

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

This is the desired result for $\widehat{\mathfrak{D}}_n^{(2)}$.

Finally, from the fact that (23) has the same asymptotic approximation as that of (24), the asymptotic approximation of $\widehat{\mathfrak{D}}_n^{(3)}$ is identical to that of $\widehat{\mathfrak{D}}_n^{(2)}$.

(*iii*) From Lemma 5(*ii*),

$$\widehat{D}_n^2 = (\widehat{D}_{o,n} + n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{V}_*])^2 + O_{\mathbb{P}}(n^{-3/2}) = (k^{-1}\mathrm{tr}[\mathbf{K}_{o,n}] + n^{-1/2}k^{-1}\mathrm{tr}[\mathbf{V}_*] + O_{\mathbb{P}}(n^{-1}))^2 + O_{\mathbb{P}}(n^{-3/2}) \\ = (k^{-1}\mathrm{tr}[\mathbf{K}_{o,n} + n^{-1/2}\mathbf{V}_*]))^2 + O_{\mathbb{P}}(n^{-3/2}), \tag{25}$$

where the second equality holds by Lemma A5(*ii*). Given that $\widehat{\mathfrak{S}}_n^{(1)} := \frac{nk}{4}(\widehat{D}_n^2 + \widehat{H}_n^2)$, (23) and (25) imply that $\widehat{\mathfrak{S}}_n^{(1)} = \mathrm{tr}[\mathbf{V}_* + \sqrt{n}\mathbf{K}_{o,n}]^2 + o_{\mathbb{P}}(n^{-1})$, as desired.

From the definition of $\widehat{\mathfrak{S}}_n^{(2)} := \frac{nk}{2}(\widehat{D}_n^2 + 2\widehat{G}_n)$, it follows that

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

by using Lemma A5(*v*) and (24). This is the desired approximation for $\widehat{\mathfrak{S}}_n^{(2)}$. Finally, (25) has the same asymptotic approximation as that of (24), and this implies that the asymptotic expansion of $\widehat{\mathfrak{S}}_n^{(3)}$ is identical to that of $\widehat{\mathfrak{S}}_n^{(2)}$. ■

References

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

ALTMAN, M. AND MCDONALD, M. (2003): "Replication with Attnetion to Accuracy," *Political Analysis*, 11, 302–307.

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

BÉNABOU, R. (2000): "Unequal Societies: Income Distribution and the Social Contract," *American Economic Review*, 90, 96–129.

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.

BERNSTEIN, D. (2005): *Matrix Mathematics*. Princeton, NJ: Princeton University Press.

BERRY, W., DEMERITT, J., AND ESARLEY, J. (2010): "Testing for Interaction in Binary Logit and Probit Models: Is a Product Terms Essential?," *American Journal of Political Science*, 54, 248–266.

BESLEY, T. AND CASE, ANNE (2003): "Political Institutions and Policy Choices: Evidence from the United States," *Journal of Economics Lieterature*, 41, 7–73.

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. (2014): "Testing the Equality of Two Positive-Definite Matrices with Application to Information Matrix Testing," *Advances in Econometrics: Essays in Honor of Peter C. B. Phillips*. Vol. 33. Eds. Yoosoon Chang, Thomas B. Fomby, and Joon Y. Park. West Yorkshire, UK: Emerald Group Publishing Limited, forthcoming.

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.

FEDDERSEN, T. AND PESENDORFER, W. (1996): "The Swing Voter's Curse," *American Economic Review*, 86, 408–424.

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.

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.

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 Liklihood 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. AND PILLAI, K. (1973): "The Distribution of the Sphericity Test Criterion," *Journl of Multivariate Analysis*, 3, 226–235.

NAGLER, J. (1994): "Scobit: An Alternative Estimator to Logit and Probit," *American Journal of Political Science*, 38, 230–255.

NAGLER, J. (1991): "The Effect of Registration Laws and Education on U.S. Vouter Turnout," *The American Political Science Review*, 85, 1393–1405.

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.

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

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

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

WHITE, H. (1980): "A Heteroskedasticity-Cosistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

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

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.

WOLFINGER, R. AND ROSENSTONE, S. (1980): *Who Votes?* New Haven: Yale University Press.

Statistics $\setminus n$	50	100	200	300	400	500
$\widehat{\mathfrak{B}}_n^{(1)}$	5.20	5.30	4.58	4.88	4.64	5.24
$\widehat{\mathfrak{B}}_n^{(2)}$	5.28	5.16	4.64	5.08	4.74	5.14
$\widehat{\mathfrak{B}}_n^{(3)}$	5.28	4.94	4.62	4.94	4.74	5.16
$\widehat{\mathfrak{D}}_n^{(1)}$	4.74	5.14	5.02	4.66	5.00	5.22
$\widehat{\mathfrak{D}}_n^{(2)}$	5.28	5.22	4.74	5.04	4.74	5.34
$\widehat{\mathfrak{D}}_n^{(3)}$	5.62	4.82	4.84	4.78	4.68	5.14
$\widehat{\mathfrak{S}}_n^{(1)}$	4.58	5.00	5.02	4.34	4.96	5.28
$\widehat{\mathfrak{S}}_n^{(2)}$	5.56	5.08	4.72	5.00	4.66	5.16
$\widehat{\mathfrak{S}}_n^{(3)}$	5.44	5.06	5.02	4.96	4.72	5.14
$\mathfrak{J}_n$	5.28	5.06	4.30	4.62	4.14	5.02
$\mathfrak{I}_n$	5.20	4.92	4.70	4.56	4.94	4.92

Table 1: EMPIRICAL LEVELS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). Repetitions: 5,000. Bootstrap Repetitions: 500. Model: $Y_t = \mathbf{X}_t'\boldsymbol{\beta} + U_t$, $\mathbf{X}_t = (1, X_t)'$ and $U_t \sim N(0, \sigma^2)$. DGP: $Y_t = X_t + U_t$, $U_t|X_t \sim N(0,1)$, and $X_t \sim N(0,1)$.

Statistics $\setminus n$	50	100	150	200	250	300
$\widehat{\mathfrak{B}}_n^{(1)}$	52.10	79.60	92.80	97.05	98.70	99.60
$\widehat{\mathfrak{B}}_n^{(2)}$	86.10	99.65	100.00	100.00	100.0	100.0
$\widehat{\mathfrak{B}}_n^{(3)}$	87.45	99.85	100.00	100.00	100.0	100.0
$\widehat{\mathfrak{D}}_n^{(1)}$	69.05	93.10	98.75	99.75	99.90	100.0
$\widehat{\mathfrak{D}}_n^{(2)}$	86.70	99.80	100.00	100.0	100.0	100.0
$\widehat{\mathfrak{D}}_n^{(3)}$	89.65	99.90	100.00	100.0	100.0	100.0
$\widehat{\mathfrak{S}}_n^{(1)}$	33.70	63.85	83.75	92.30	97.25	99.30
$\widehat{\mathfrak{S}}_n^{(2)}$	89.35	99.90	100.00	100.0	100.0	100.0
$\widehat{\mathfrak{S}}_n^{(3)}$	90.30	99.95	100.00	100.0	100.0	100.0
$\mathfrak{J}_n$	49.45	77.15	92.30	96.60	98.50	99.25
$\mathfrak{I}_n$	20.40	88.00	99.90	100.0	100.0	100.0

Table 2: EMPIRICAL GLOBAL POWERS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). Repetitions: 2,000. Bootstrap Repetitions: 500. Model: $Y_t = \mathbf{X}_t'\boldsymbol{\beta} + U_t$, $\mathbf{X}_t = (1, X_t)'$, and $U_t \sim N(0, \sigma^2)$. DGP: $Y_t = X_t + U_t$, $U_t|X_t \sim N(0, \exp(X_t))$, and $X_t \sim N(1,1)$.

Statistics $\setminus n$	100	500	800	1,200	1,600	2,000
$\widehat{\mathfrak{B}}_n^{(1)}$	14.05	55.45	74.00	89.60	97.50	98.90
$\widehat{\mathfrak{B}}_n^{(2)}$	1.20	36.05	67.60	90.60	98.70	99.65
$\widehat{\mathfrak{B}}_n^{(3)}$	3.20	44.15	73.20	92.30	98.95	99.70
$\widehat{\mathfrak{D}}_n^{(1)}$	18.35	61.95	79.80	92.70	98.20	99.30
$\widehat{\mathfrak{D}}_n^{(2)}$	1.50	38.65	69.30	91.20	98.70	99.65
$\widehat{\mathfrak{D}}_n^{(3)}$	6.45	52.25	78.80	93.55	99.05	99.75
$\widehat{\mathfrak{S}}_n^{(1)}$	21.35	66.60	83.00	94.25	98.85	99.50
$\widehat{\mathfrak{S}}_n^{(2)}$	5.05	48.25	75.60	92.70	99.00	99.70
$\widehat{\mathfrak{S}}_n^{(3)}$	7.70	53.95	79.45	93.85	99.15	99.75
$\mathfrak{J}_n$	1.05	51.15	82.15	95.80	99.50	99.85
$\mathfrak{I}_n$	25.25	74.70	89.45	96.75	99.35	99.80

Table 3: EMPIRICAL GLOBAL POWERS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). Repetitions: 2,000. Bootstrap Repetitions: 500. Model: $Y_t = \mathbf{X}_t'\boldsymbol{\beta} + U_t$, $\mathbf{X}_t = (1, X_t)'$, and $U_t \sim N(0, \sigma^2)$. DGP: $Y_t = X_t + U_t$, $U_t|X_t \sim 0.5 \cdot N(-1, 1) + 0.5 \cdot N(1, 1)$, and $X_t \sim N(0, 1)$.

Statistics $\setminus n$	200	500	800	1,200	1,600	2,000
$\widehat{\mathfrak{B}}_n^{(1)}$	27.60	49.50	64.70	82.10	89.85	94.85
$\widehat{\mathfrak{B}}_n^{(2)}$	35.35	56.30	72.55	85.80	93.85	96.60
$\widehat{\mathfrak{B}}_n^{(3)}$	33.75	54.85	71.00	84.95	93.60	96.40
$\widehat{\mathfrak{D}}_n^{(1)}$	23.95	44.90	61.30	79.85	87.90	93.85
$\widehat{\mathfrak{D}}_n^{(2)}$	35.25	56.00	72.10	85.55	93.80	96.60
$\widehat{\mathfrak{D}}_n^{(3)}$	30.75	52.60	69.05	83.70	92.70	95.95
$\widehat{\mathfrak{S}}_n^{(1)}$	14.60	37.40	55.30	75.20	84.45	91.50
$\widehat{\mathfrak{S}}_n^{(2)}$	32.60	54.20	70.45	84.60	93.30	96.25
$\widehat{\mathfrak{S}}_n^{(3)}$	30.20	52.00	68.85	83.55	92.50	95.85
$\mathfrak{J}_n$	38.75	62.50	78.40	90.50	96.45	97.85
$\mathfrak{I}_n$	1.40	2.50	6.70	19.90	37.60	55.30

Table 4: EMPIRICAL GLOBAL POWERS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). Repetitions: 2,000. Bootstrap Repetitions: 500. Model: $Y_t = \mathbf{X}_t'\boldsymbol{\beta} + U_t$, $\mathbf{X}_t = (1, X_t)'$, and $U_t \sim N(0, \sigma^2)$. DGP: $Y_t = X_t + U_t$, $U_t|X_t \sim t_3$, and $X_t \sim N(0, 1)$.

Statistics $\setminus n$	50	100	150	200	250	300
$\widehat{\mathfrak{B}}_n^{(1)}$	39.75	69.85	85.70	93.50	96.90	98.70
$\widehat{\mathfrak{B}}_n^{(2)}$	43.70	71.85	88.05	94.45	97.45	99.20
$\widehat{\mathfrak{B}}_n^{(3)}$	40.10	68.40	86.25	93.85	97.00	99.10
$\widehat{\mathfrak{D}}_n^{(1)}$	33.55	64.55	82.00	91.60	95.85	97.90
$\widehat{\mathfrak{D}}_n^{(2)}$	44.55	72.60	88.00	94.50	97.45	99.20
$\widehat{\mathfrak{D}}_n^{(3)}$	34.10	64.50	83.50	91.85	96.35	98.85
$\widehat{\mathfrak{S}}_n^{(1)}$	13.40	47.10	70.55	84.75	91.85	95.95
$\widehat{\mathfrak{S}}_n^{(2)}$	40.90	69.20	86.45	93.60	96.80	99.05
$\widehat{\mathfrak{S}}_n^{(3)}$	33.10	64.00	83.15	91.70	96.10	98.90
$\mathfrak{J}_n$	17.75	38.50	51.05	62.35	72.50	79.30
$\mathfrak{I}_n$	20.50	32.60	44.50	59.75	71.20	83.95

Table 5: EMPIRICAL GLOBAL POWERS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). Repetitions: 2,000. Bootstrap Repetitions: 500. Model: $Y_t = \mathbf{X}_t'\boldsymbol{\beta}+U_t$, $\mathbf{X}_t = (1, X_t)'$, and $U_t \sim N(0, \sigma^2)$. DGP: $Y_t = X_t + \frac{1}{2}X_t^2 + U_t$, $U_t|X_t \sim N(0,1)$, and $X_t \sim N(0,1)$.

Statistics $\setminus n$	50	100	200	300	400	500
$\widehat{\mathfrak{B}}_n^{(1)}$	8.10	7.53	6.80	5.67	6.53	5.93
$\widehat{\mathfrak{B}}_n^{(2)}$	14.45	13.23	13.10	12.63	12.90	12.07
$\widehat{\mathfrak{B}}_n^{(3)}$	13.95	13.40	13.23	12.87	12.90	12.50
$\widehat{\mathfrak{D}}_n^{(1)}$	7.50	7.33	5.93	5.17	6.20	5.70
$\widehat{\mathfrak{D}}_n^{(2)}$	14.45	13.57	13.37	12.63	12.93	12.37
$\widehat{\mathfrak{D}}_n^{(3)}$	13.65	13.57	13.73	13.60	12.97	12.53
$\widehat{\mathfrak{S}}_n^{(1)}$	5.70	6.00	5.47	4.60	5.73	5.27
$\widehat{\mathfrak{S}}_n^{(2)}$	14.50	13.53	13.77	12.87	12.87	12.53
$\widehat{\mathfrak{S}}_n^{(3)}$	13.85	13.63	13.67	13.60	13.13	12.60
$\mathfrak{J}_n$	7.60	7.33	7.00	6.07	5.80	5.27
$\mathfrak{I}_n$	5.50	6.37	7.13	8.93	9.67	10.80

Table 6: EMPIRICAL LOCAL POWERS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). Repetitions: 3,000. Bootstrap Repetitions: 500. Model: $Y_t = \mathbf{X}_t'\boldsymbol{\beta} + U_t$, $\mathbf{X}_t = (1, X_t)'$, and $U_t \sim N(0, \sigma^2)$. DGP: $Y_t = X_t + U_t$, $U_t|X_t \sim N(0, \exp(2n^{-1/2}X_t))$, and $X_t \sim N(1,1)$.

Statistics $\setminus n$	1,000	2,000	4,000	5,000	10,000	20,000
$\widehat{\mathfrak{B}}_n^{(1)}$	4.57	4.43	5.80	5.47	10.40	11.37
$\widehat{\mathfrak{B}}_n^{(2)}$	7.53	16.47	30.27	36.60	60.07	68.40
$\widehat{\mathfrak{B}}_n^{(3)}$	7.93	16.27	30.53	36.37	59.98	68.07
$\widehat{\mathfrak{D}}_n^{(1)}$	5.07	4.53	5.70	5.23	9.52	10.70
$\widehat{\mathfrak{D}}_n^{(2)}$	7.73	16.40	30.37	36.63	60.07	68.33
$\widehat{\mathfrak{D}}_n^{(3)}$	8.27	16.57	30.47	36.67	59.68	68.03
$\widehat{\mathfrak{S}}_n^{(1)}$	5.67	4.70	5.57	5.17	8.36	10.07
$\widehat{\mathfrak{S}}_n^{(2)}$	8.13	16.53	30.33	36.33	59.77	68.00
$\widehat{\mathfrak{S}}_n^{(3)}$	8.27	16.70	30.47	36.30	59.41	67.80
$\mathfrak{J}_n$	8.57	20.63	39.27	45.57	69.40	77.97
$\mathfrak{I}_n$	9.03	14.87	25.00	29.87	52.96	62.17

Table 7: EMPIRICAL LOCAL POWERS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). Repetitions: 3,000. Bootstrap Repetitions: 500. Model: $Y_t = \mathbf{X}_t'\boldsymbol{\beta} + U_t$, $\mathbf{X}_t = (1, X_t)'$, and $U_t \sim N(0, \sigma^2)$. DGP: $Y_t = X_t + U_t$, $U_t|X_t \sim (10n^{-1/2}) \cdot N(-(1 - 10n^{-1/2}), 1) + (1 - 10n^{-1/2}) \cdot N(10n^{-1/2}, 1)$, and $X_t \sim N(0, 1)$.

Statistics $\setminus n$	1,000	2,000	3,000	4,000	5,000	10,000
$\widehat{\mathfrak{B}}_n^{(1)}$	42.10	51.00	55.20	60.63	60.23	61.97
$\widehat{\mathfrak{B}}_n^{(2)}$	49.43	58.87	60.93	65.27	64.87	66.77
$\widehat{\mathfrak{B}}_n^{(3)}$	48.03	57.80	59.23	64.43	64.30	65.80
$\widehat{\mathfrak{D}}_n^{(1)}$	39.30	48.33	52.90	58.00	58.27	60.73
$\widehat{\mathfrak{D}}_n^{(2)}$	49.33	58.83	60.73	65.20	64.80	66.50
$\widehat{\mathfrak{D}}_n^{(3)}$	45.80	55.90	57.47	62.73	63.37	64.87
$\widehat{\mathfrak{S}}_n^{(1)}$	34.87	45.27	49.63	55.10	56.17	59.07
$\widehat{\mathfrak{S}}_n^{(2)}$	47.40	57.13	58.97	63.90	64.17	65.43
$\widehat{\mathfrak{S}}_n^{(3)}$	45.43	55.60	57.33	62.37	63.23	64.67
$\mathfrak{J}_n$	56.13	65.40	68.93	72.57	73.33	75.13
$\mathfrak{I}_n$	4.90	11.07	18.00	24.40	29.83	39.97

Table 8: EMPIRICAL LOCAL POWERS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). Repetitions: 3,000. Bootstrap Repetitions: 500. Model: $Y_t = \mathbf{X}_t'\boldsymbol{\beta} + U_t$, $\mathbf{X}_t = (1, X_t)'$, and $U_t \sim N(0, \sigma^2)$. DGP: $Y_t = X_t + U_t$, $U_t|X_t \sim N(0, 1)/\{1 + n^{-1/2}(\mathcal{X}_3^2 - 1)\}^{1/2}$, and $X_t \sim N(0, 1)$.

Statistics $\setminus n$	3,000	4,000	5,000	10,000	20,000	30,000
$\widehat{\mathfrak{B}}_n^{(1)}$	20.90	17.57	16.00	9.33	6.97	6.67
$\widehat{\mathfrak{B}}_n^{(2)}$	20.43	15.87	14.77	8.73	6.47	6.33
$\widehat{\mathfrak{B}}_n^{(3)}$	19.57	15.27	14.33	8.67	6.50	6.27
$\widehat{\mathfrak{D}}_n^{(1)}$	19.47	16.50	15.40	8.97	6.90	6.50
$\widehat{\mathfrak{D}}_n^{(2)}$	20.53	15.80	14.80	8.83	6.50	6.33
$\widehat{\mathfrak{D}}_n^{(3)}$	18.77	14.87	13.80	8.33	6.60	6.10
$\widehat{\mathfrak{S}}_n^{(1)}$	18.70	15.43	14.93	8.60	6.47	6.23
$\widehat{\mathfrak{S}}_n^{(2)}$	19.57	15.13	14.30	8.70	6.50	6.23
$\widehat{\mathfrak{S}}_n^{(3)}$	18.67	14.70	13.90	8.40	6.57	6.13
$\mathfrak{J}_n$	5.37	5.20	5.77	5.33	4.63	5.00
$\mathfrak{I}_n$	10.37	8.87	9.10	7.10	6.00	6.00

Table 9: EMPIRICAL LOCAL POWERS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). Repetitions: 3,000. Bootstrap Repetitions: 500. Model: $Y_t = \mathbf{X}_t'\boldsymbol{\beta} + U_t$ and $U_t \sim N(0, \sigma^2)$. DGP: $Y_t = X_t + 5n^{-1/2}X_t^2 + U_t$, $U_t|X_t \sim N(0,1)$, and $X_t \sim N(0,1)$.

Statistics $\setminus n$	50	100	200	300	400	500
$\widehat{\mathfrak{B}}_n^{(1)}$	3.11	4.28	5.06	5.03	4.82	4.72
$\widehat{\mathfrak{B}}_n^{(2)}$	3.37	4.39	4.86	4.86	4.88	4.74
$\widehat{\mathfrak{B}}_n^{(3)}$	2.91	3.99	4.78	4.71	4.74	4.64
$\widehat{\mathfrak{D}}_n^{(1)}$	2.72	4.09	4.84	4.90	4.69	4.80
$\widehat{\mathfrak{D}}_n^{(2)}$	3.43	4.43	4.84	4.87	4.89	4.75
$\widehat{\mathfrak{D}}_n^{(3)}$	2.44	3.76	4.53	4.59	4.70	4.50
$\widehat{\mathfrak{S}}_n^{(1)}$	2.58	3.99	4.69	4.73	4.78	4.75
$\widehat{\mathfrak{S}}_n^{(2)}$	2.92	3.98	4.79	4.69	4.75	4.62
$\widehat{\mathfrak{S}}_n^{(3)}$	2.39	3.76	4.51	4.61	4.71	4.56
$\mathfrak{I}_n$	3.36	4.18	4.77	5.07	4.82	4.75

Table 10: EMPIRICAL LEVELS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). Repetitions: 5,000. Bootstrap Repetitions: 500. Model for $\mathbb{E}[Y_t|X_t]$: Probit$(\mathbf{X}_t'\boldsymbol{\beta})$ and $\mathbf{X}_t = (1, X_t)'$. DGP: $\mathbb{E}[Y_t|X_t] =$ Probit$(1 + X_t)$ and $X_t \sim N(0,1)$.

Statistics $\setminus n$	50	100	150	200	250	300
$\widehat{\mathfrak{B}}_n^{(1)}$	11.85	77.30	96.80	99.65	100.0	100.0
$\widehat{\mathfrak{B}}_n^{(2)}$	10.05	77.25	96.80	99.75	100.0	100.0
$\widehat{\mathfrak{B}}_n^{(3)}$	14.45	81.80	97.65	99.85	100.0	100.0
$\widehat{\mathfrak{D}}_n^{(1)}$	15.55	82.30	97.60	99.85	100.0	100.0
$\widehat{\mathfrak{D}}_n^{(2)}$	9.80	76.20	96.75	99.75	100.0	100.0
$\widehat{\mathfrak{D}}_n^{(3)}$	17.65	84.20	98.10	99.85	100.0	100.0
$\widehat{\mathfrak{S}}_n^{(1)}$	18.00	84.80	98.05	99.85	100.0	100.0
$\widehat{\mathfrak{S}}_n^{(2)}$	14.35	81.80	97.55	99.85	100.0	100.0
$\widehat{\mathfrak{S}}_n^{(3)}$	17.65	84.40	98.10	99.85	100.0	100.0
$\mathfrak{I}_n$	11.85	74.75	95.95	99.60	100.0	100.0

Table 11: EMPIRICAL GLOBAL POWERS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). Repetitions: 2,000. Bootstrap Repetitions: 500. Model for $\mathbb{E}[Y_t|X_t]$: Probit($\mathbf{X}_t'\boldsymbol{\beta}$) and $\mathbf{X}_t = (1, X_t)'$. DGP: $\mathbb{E}[Y_t|X_t] =$ Probit$(1 + X_t + X_t^4)$ and $X_t \sim N(0, 1)$.

Statistics $\setminus n$	1,000	2,000	3,000	4,000	5,000	10,000
$\widehat{\mathfrak{B}}_n^{(1)}$	8.85	12.15	14.30	18.55	18.60	31.75
$\widehat{\mathfrak{B}}_n^{(2)}$	9.90	13.80	16.30	20.80	23.10	39.80
$\widehat{\mathfrak{B}}_n^{(3)}$	9.55	13.30	15.80	20.25	22.60	39.40
$\widehat{\mathfrak{D}}_n^{(1)}$	8.20	11.65	14.05	17.85	18.30	31.00
$\widehat{\mathfrak{D}}_n^{(2)}$	9.90	13.80	16.25	20.85	23.10	39.85
$\widehat{\mathfrak{D}}_n^{(3)}$	8.70	12.80	15.20	19.90	22.10	39.00
$\widehat{\mathfrak{S}}_n^{(1)}$	7.55	11.10	13.65	17.15	17.90	30.55
$\widehat{\mathfrak{S}}_n^{(2)}$	9.50	13.30	15.80	20.25	22.60	39.40
$\widehat{\mathfrak{S}}_n^{(3)}$	8.70	12.75	15.20	19.90	22.10	39.00
$\mathfrak{I}_n$	3.55	3.75	5.70	7.70	9.30	20.25

Table 12: EMPIRICAL GLOBAL POWERS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). Repetitions: 2,000. Bootstrap Repetitions: 500. Model for $\mathbb{E}[Y_t|X_t]$: Probit($\mathbf{X}_t'\boldsymbol{\beta}$) and $\mathbf{X}_t = (1, X_t)'$. DGP: $\mathbb{E}[Y_t|X_t] =$ Logit$[-(1 + X_t)]$ and $X_t \sim N(0, 1)$.

Statistics $\setminus n$	1,000	2,000	3,000	4,000	5,000	10,000
$\widehat{\mathfrak{B}}_n^{(1)}$	26.73	39.27	43.50	40.78	41.12	30.73
$\widehat{\mathfrak{B}}_n^{(2)}$	79.62	82.27	82.13	80.26	80.12	71.80
$\widehat{\mathfrak{B}}_n^{(3)}$	79.07	81.87	81.53	79.89	79.52	71.53
$\widehat{\mathfrak{D}}_n^{(1)}$	22.44	35.67	40.00	37.88	38.73	29.23
$\widehat{\mathfrak{D}}_n^{(2)}$	79.56	82.27	82.13	80.23	80.12	71.77
$\widehat{\mathfrak{D}}_n^{(3)}$	78.47	81.47	81.13	79.76	78.99	71.13
$\widehat{\mathfrak{S}}_n^{(1)}$	16.11	29.83	35.37	34.72	36.24	28.00
$\widehat{\mathfrak{S}}_n^{(2)}$	79.04	81.83	81.50	79.89	79.52	71.53
$\widehat{\mathfrak{S}}_n^{(3)}$	78.42	81.43	81.07	79.73	78.96	71.13
$\mathfrak{I}_n$	47.24	51.50	55.67	57.22	57.48	59.47

Table 13: EMPIRICAL LOCAL POWERS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). Repetitions: 3,000. Bootstrap Repetitions: 500. Model for $\mathbb{E}[Y_t|X_t]$: Probit$(\mathbf{X}_t'\boldsymbol{\beta})$ and $\mathbf{X}_t = (1, X_t)'$. DGP: $\mathbb{E}[Y_t|X_t] =$ Probit$(1 + X_t + n^{-1/2}X_t^4)$ and $X_t \sim N(0, 1)$.

Statistics $\setminus n$	300	400	500	1,000	2,000	3,000
$\widehat{\mathfrak{B}}_n^{(1)}$	39.00	39.00	47.00	61.60	69.77	75.40
$\widehat{\mathfrak{B}}_n^{(2)}$	40.50	42.50	49.25	65.27	73.23	77.77
$\widehat{\mathfrak{B}}_n^{(3)}$	39.75	40.75	48.25	64.20	72.67	77.40
$\widehat{\mathfrak{D}}_n^{(1)}$	35.75	37.75	44.50	60.00	68.33	74.70
$\widehat{\mathfrak{D}}_n^{(2)}$	40.50	42.50	49.25	65.27	73.23	77.77
$\widehat{\mathfrak{D}}_n^{(3)}$	37.75	39.50	46.50	62.93	71.47	76.93
$\widehat{\mathfrak{S}}_n^{(1)}$	32.00	35.25	40.75	57.40	66.57	73.13
$\widehat{\mathfrak{S}}_n^{(2)}$	39.75	40.75	48.25	64.20	72.67	77.40
$\widehat{\mathfrak{S}}_n^{(3)}$	37.75	39.50	46.50	62.93	71.47	76.87
$\mathfrak{I}_n$	3.00	3.25	4.00	8.07	13.70	20.83

Table 14: EMPIRICAL LOCAL POWERS OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). Repetitions: 3,000. Bootstrap Repetitions: 500. Model for $\mathbb{E}[Y_t|X_t]$: Probit$(\mathbf{X}_t'\boldsymbol{\beta})$ and $\mathbf{X}_t = (1, X_t)'$. DGP: $\mathbb{E}[Y_t|X_t] = (1 - n^{-1/2})$Probit$(1 + X_t) + n^{-1/2}$Logit$[-(1 + X_t)]$ and $X_t \sim N(0, 1)$.

Statistics \ Models	Probit Models		Logit Models		Scobit Models	
	w/ Products	w/o Products	w/ Products	w/o Products	w/ Products	w/o Products
Constant	-2.7431	-2.5229	-4.4129	-4.0727	-5.3465	-4.4062
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Closing Date	0.0006	-0.0078	-0.0001	-0.0132	-0.0024	-0.0217
	(0.8685)	(0.0000)	(0.9832)	(0.0000)	(0.7997)	(0.0000)
Education	0.2645	0.1818	0.3585	0.2426	0.3494	0.2094
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0041)	(0.0000)
Education^2	0.0050	0.0123	0.0192	0.0282	0.0663	0.0711
	(0.2433)	(0.0000)	(0.0133)	(0.0000)	(0.0000)	(0.0000)
Age	0.06965	0.0697	0.1141	0.1142	0.1837	0.1813
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Age^2	-0.0005	-0.0005	-0.0008	-0.0008	-0.0012	-0.0012
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
South	-0.1154	-0.1159	-0.1897	-0.1904	-0.2975	-0.2956
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Gubernatorial Election	0.0034	0.0034	0.0048	0.0052	-0.0014	-0.0000
	(0.7670)	(0.7666)	(0.8012)	(0.7853)	(0.9637)	(0.9998)
Closing Date $\times$ *Education*	-0.0031		-0.0044		-0.0052	
	(0.0399)		(0.0956)		(0.2283)	
Closing Date $\times$ *Education*^{2}	0.0002		0.0003		0.0002	
	(0.0075)		(0.2219)		(0.6378)	
α_*					0.4105	0.4194
					(0.0000)	(0.0000)
Sample Size	99,676	99,676	99,676	99,676	99,676	99,676
Log-Likelihood	-55,815.28	-55,818.03	-55,774.55	-55,777.67	-55,725.09	-55,730.63
$\widehat{\mathfrak{B}}_n^{(1)}$	589.90	272.79	509.42	224.26	812.99	651.59
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\widehat{\mathfrak{B}}_n^{(2)}$	3,007.43	1,875.64	2,849.08	1,803.54	5,672.81	5,337.46
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\widehat{\mathfrak{B}}_n^{(3)}$	2,925.71	1,834.28	2,775.50	1,766.55	5,490.45	5,162.59
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\widehat{\mathfrak{D}}_n^{(1)}$	553.60	254.71	476.66	208.09	737.84	582.65
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\widehat{\mathfrak{D}}_n^{(2)}$	2,951.45	1,831.22	2,801.19	1,765.28	5,482.02	5,118.08
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\widehat{\mathfrak{D}}_n^{(3)}$	2,797.14	1,753.71	2,662.08	1,695.95	5,184.35	4,805.33
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\widehat{\mathfrak{S}}_n^{(1)}$	512.75	234.03	439.87	189.60	646.16	495.21
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\widehat{\mathfrak{S}}_n^{(2)}$	2,813.76	1,745.44	2,679.71	1,690.03	5,106.86	4,723.83
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$\widehat{\mathfrak{S}}_n^{(3)}$	2,741.16	1,709.28	2,614.19	1,657.70	4,956.56	4,585.95
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Table 15: EMPIRICAL MODEL ESTIMATIONS AND INFERENCES OF THE TEST STATISTICS (LEVEL OF SIGNIFICANCE: 5%). The figures in parentheses stand for the p-values. The p-values of the parameter estimates are computed by White's (1980) heteroskedasticity consistent standard errors, and the p-values of the test statistics are obtained by implementing the parametric bootstrap.