Int. J. Anal. Appl. (2026), 24:382

International Journal of Analysis and Applications

Performance and Robustness of Covariance Matrix Equality Tests under High-Dimensional Covariance Structures

Sirithip Wasinrat[1], **Boonyarit Choopradit**[2,*]

[1]*Faculty of Science, Chandrakasem Rajabhat University, Bangkok, Thailand*
[2]*Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathumthani, Thailand*

[*]*Corresponding author:* boonyarit@mathstat.sci.tu.ac.th

Abstract. When the dimension is close to or above the number of observations, classical tests for equal covariance matrices cannot be used, and the behaviour of their high-dimensional alternatives may change with the covariance structure. We examine how well three such procedures, the Srivastava–Yanagihara (SY), Ahmad (AHM), and Qayed–Han (QH) tests, hold the nominal level and how often they detect unequal covariance matrices. Monte Carlo simulations with 5,000 replications were conducted for three multivariate normal populations with sample sizes from 10 to 160 and dimensions equal to or larger than the sample size, under first-order autoregressive (AR(1)), banded Toeplitz (BTOEP), and simple diagonal (SIM) covariance structures. QH kept its empirical Type I error rate between 0.041 and 0.059 in all settings, and SY stayed inside Bradley's interval in all but one setting, whereas AHM was liberal for sample sizes of 20 or less, with rates up to 0.104. SY was the most powerful test in most AR(1) and SIM settings, while QH was the most powerful under BTOEP, with power of at least 0.91 even for samples of size 10. Because QH depends on the data only through the sum of the variables and is equivalent to Bartlett's test applied to that sum, it has high power when the variance of this sum differs between populations but only trivial power, equal to its size, against alternatives that leave it unchanged. SY is therefore recommended as the primary test when the covariance structure is unknown, with QH as a complementary test.

1. Introduction

Many multivariate procedures assume that several populations share a common covariance matrix [1,2]. However, if the dimension exceeds the sample sizes, every sample covariance matrix is singular, and methods that rely on its inverse or determinant, for instance Hotelling's T^2 test,

Received: Aug. 15, 2026.

2020 *Mathematics Subject Classification.* 62H15; 62F03; 65C05.

Key words and phrases. high-dimensional data; homogeneity of covariance matrices; Type I error rate; power; Monte Carlo simulation.

https://doi.org/10.28924/2291-8639-24-2026-382
ISSN: 2291-8639

© 2026 the author(s).

linear discriminant analysis, and Box's M test for equal covariance matrices, break down [1,3]. For k populations, the hypotheses of interest are

$$H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \mathbf{\Sigma}_k = \mathbf{\Sigma} \tag{1}$$

versus

$$H_1 : \text{ Not all } \mathbf{\Sigma}_i \text{ are equal,}$$

where $\mathbf{\Sigma}_i$ denotes the covariance matrix of population i ($i = 1, \ldots, k$). Throughout, the dimension p is at least as large as each sample size N_i ($p \geq N_i$). High-dimensional tests of (1) have been proposed in [1], [3], [4], [5], and [6].

However, the Srivastava–Yanagihara (SY) [5], Ahmad (AHM) [3], and Qayed–Han (QH) [1] tests, which are based on different principles, have not yet been evaluated side by side. Through simulation, this paper examines how closely each test keeps its nominal level and how much power it has under three covariance structures.

The three tests are reviewed in Section 2. The simulation design is given in Section 3, the results are presented and discussed in Section 4, and conclusions are drawn in Section 5.

2. Three Tests for Homogeneity of Covariance Matrices

For each population i ($i = 1, \ldots, k$), the observations $\mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{iN_i}$ are independent and identically distributed (i.i.d.), each following a p-variate normal distribution with mean $\boldsymbol{\mu}_i$ and positive definite covariance matrix $\mathbf{\Sigma}_i$; here N_i is the size of the sample drawn from population i.

In what follows, $\mathbf{M}'$ and $\operatorname{tr}(\mathbf{M})$ are, respectively, the transpose and the sum of the diagonal elements of a matrix $\mathbf{M}$, $\operatorname{diag}(\cdot)$ is the diagonal matrix with the listed entries, $E(\cdot)$ is the expectation operator, log is the natural logarithm, and α is the nominal significance level. Let $n_i = N_i - 1$, $N^* = \sum_{i=1}^{k} n_i$, $N_{\min} = \min\{N_1, \ldots, N_k\}$ and $N_{\max} = \max\{N_1, \ldots, N_k\}$. For the i-th sample, we use the mean vector, the centred cross-product matrix, and the sample covariance matrix, namely $\overline{\mathbf{X}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{X}_{ij}$, $\mathbf{V}_i = \sum_{j=1}^{N_i} \left(\mathbf{X}_{ij} - \overline{\mathbf{X}}_i\right)\left(\mathbf{X}_{ij} - \overline{\mathbf{X}}_i\right)'$ and $\mathbf{S}_i = \mathbf{V}_i / n_i$.

2.1. Srivastava–Yanagihara's Test.

The first test is due to Srivastava and Yanagihara (Eq. (1.17) of [5]); it compares trace-based estimates computed from the individual samples. Its statistic is

$$SY = \sum_{i=1}^{k} \frac{\left(\widehat{a}_{2i} - \bar{\widehat{a}}_2\right)^2}{\widehat{\eta}_i^2} \tag{2}$$

where

$$\bar{\widehat{a}}_2 = \frac{\sum_{i=1}^{k} \frac{\widehat{a}_{2i}}{\widehat{\eta}_i^2}}{\sum_{i=1}^{k} \frac{1}{\widehat{\eta}_i^2}}, \quad \widehat{\eta}_i^2 = \frac{4}{n_i^2}\widehat{a}_2^2\left(1 + \frac{2n_i\widehat{a}_4}{p\widehat{a}_2^2}\right), \quad \widehat{a}_{2i} = \frac{1}{p\,(n_i - 1)\,(n_i + 2)}\left\{\operatorname{tr}\mathbf{V}_i^2 - \frac{1}{n_i}\left(\operatorname{tr}\mathbf{V}_i\right)^2\right\},$$

$$\widehat{a}_4 = \frac{1}{c_0}\left(\frac{1}{p}\operatorname{tr}\mathbf{V}^4 - pc_1\widehat{a}_1\widehat{a}_3 - p^2c_2\widehat{a}_1^2\widehat{a}_2 - pc_3\widehat{a}_2^2 - N^*p^3\widehat{a}_1^4\right), \quad \mathbf{V} = \sum_{i=1}^{k}\mathbf{V}_i,$$

$$\widehat{a_1} = \frac{1}{N^* p}\,\mathrm{tr}\,\mathbf{V}, \quad \widehat{a_2} = \frac{1}{(N^*-1)(N^*+2)\,p}\left\{\mathrm{tr}\,\mathbf{V}^2 - \frac{1}{N^*}\left(\mathrm{tr}\,\mathbf{V}\right)^2\right\},$$

$$\widehat{a_3} = \frac{1}{N^*\left(N^{*2}+3N^*+4\right)}\left\{\frac{1}{p}\,\mathrm{tr}\,\mathbf{V}^3 - 3N^*\left(N^*+1\right)p\widehat{a_1}\widehat{a_2} - N^* p^2 \widehat{a_1}^3\right\},$$

$$c_0 = N^*\left(N^{*3}+6N^{*2}+21N^*+18\right), \quad c_1 = 2N^*\left(2N^{*2}+6N^*+9\right), \quad c_2 = 2N^*\left(3N^*+2\right),$$

$$c_3 = N^*\left(2N^{*2}+5N^*+7\right).$$

Here $\widehat{a_l}$ is a consistent estimator of $a_l = \mathrm{tr}\left(\mathbf{\Sigma}^l\right)/p$, $l = 1, \ldots, 4$, based on the pooled matrix $\mathbf{V}$, and $\widehat{a_{2i}}$ estimates $\mathrm{tr}\left({\mathbf{\Sigma}_i}^2\right)/p$ from the i-th sample. The factor $\widehat{a_3}$ in the second term of $\widehat{a_4}$ follows from the expression for $E\left(\mathrm{tr}\,\mathbf{V}^4\right)$ in Corollary 6.1(c) of [5]; it is missing from the statement of Theorem 2.2 in [5]. Omitting it makes SY change under a rescaling of the data and distorts its Type I error rate.

If conditions (A1)–(A4) of [5] are satisfied and $(N^*, p) \to \infty$, SY has, under H_0, a limiting chi-squared distribution whose degrees of freedom equal $k-1$. Writing $\chi^2_{1-\alpha,k-1}$ for the $100(1-\alpha)$th percentile of this limiting distribution, we reject H_0 whenever $SY > \chi^2_{1-\alpha,k-1}$.

2.2. **Ahmad's Test.** Ahmad [3] proposed the statistic:

$$AHM = \frac{\widehat{F_k}}{\widehat{\sigma_0}} \tag{3}$$

where

$$\widehat{F_k} = \frac{1}{p^2}\left[(k-1)\sum_{i=1}^{k} B_i - 2\sum_{i<i'} B_{ii'}\right], \quad B_i = \frac{1}{N_i\left(N_i-1\right)}\sum_{j\neq j'}\left(\mathbf{X}'_{ij}\mathbf{X}_{ij'}\right)^2,$$

$$B_{ii'} = \frac{1}{N_i N_{i'}}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_{i'}}\left(\mathbf{X}'_{ij}\mathbf{X}_{i'j'}\right)^2,$$

where $1 \le i < i' \le k$, the sum in B_i extends over all ordered pairs (j, j') of distinct indices from $\{1, \ldots, N_i\}$, and σ_0^2 denotes the leading-order term in the variance of $\widehat{F_k}$ under H_0 (Eq. (39) of [3]), given by $\sigma_0^2 = 4\left[(k-1)^2\sum_{i=1}^{k}\frac{1}{N_i^2} + \sum_{i<i'}\frac{2}{N_i N_{i'}}\right]\frac{\left[\mathrm{tr}\left(\mathbf{\Sigma}^2\right)\right]^2}{p^4}$.

The statistic $\widehat{F_k}$ has expectation $F_k = \frac{1}{p^2}\sum_{i<i'}\mathrm{tr}\left\{\left(\mathbf{\Sigma}_i - \mathbf{\Sigma}_{i'}\right)^2\right\}$, a scaled sum of squared Frobenius distances between pairs of covariance matrices that vanishes under H_0. The estimator $\widehat{\sigma_0}$ is obtained by replacing $\mathrm{tr}\left(\mathbf{\Sigma}^2\right)$ in σ_0 with the pooled unbiased estimator $\widehat{\mathrm{tr}\left(\mathbf{\Sigma}^2\right)} = \frac{\sum_{i<i'} N_i N_{i'} B_{ii'}}{\sum_{i<i'} N_i N_{i'}}$. Under normality, a direct calculation shows that the exact null variance of $\widehat{F_k}$ is obtained from σ_0^2 by replacing $\frac{1}{N_i^2}$ with $\frac{1}{\{N_i(N_i-1)\}}$ and $\left[\mathrm{tr}\left(\mathbf{\Sigma}^2\right)\right]^2$ with $\left[\mathrm{tr}\left(\mathbf{\Sigma}^2\right)\right]^2 + \mathrm{tr}\left(\mathbf{\Sigma}^4\right)$; hence σ_0^2 underestimates the null variance when N_i is small or $\frac{\mathrm{tr}\left(\mathbf{\Sigma}^4\right)}{\left[\mathrm{tr}\left(\mathbf{\Sigma}^2\right)\right]^2}$ is not negligible. As in [3], the statistic is written for zero mean vectors, which is the case in the simulation design of Section 3.1; for data with unknown means, the bilinear forms must be computed from suitably centred observations.

Under H_0, AHM is asymptotically standard normal under Assumptions 1–6 of [3] as $N_i, p \to \infty$ (Theorem 11 of [3]). The null hypothesis is rejected when $AHM > Z_{1-\alpha}$, where $Z_{1-\alpha}$ is the $100(1-\alpha)$th percentile of N(0, 1).

2.3. **Qayed–Han's Test.** Qayed and Han [1] proposed the statistic:

$$QH = \rho\left(N^* \log \widehat{S} - \sum_{i=1}^{k} n_i \log \widehat{S}_i\right) \tag{4}$$

where ρ is the scale factor of Bartlett [12]

$$\rho = \left\{1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k} \frac{1}{n_i} - \frac{1}{N^*}\right)\right\}^{-1}$$

and $\widehat{S}_i = \mathbf{1}'\mathbf{S}_i\mathbf{1}$, $\widehat{S} = \frac{1}{N^*}\sum_{i=1}^{k} n_i\widehat{S}_i$, where $\mathbf{1} = (1, \ldots, 1)'$ is the p-vector with every entry equal to one. Under H_0, as $N_{\min} \to \infty$ and p grows with $N_{\max}$ as required in Theorem 3.2 of [1] (which requires the p components of each observation to form a stationary process), the null distribution of QH tends to the chi-squared law with $k-1$ degrees of freedom, and H_0 is rejected if $QH > \chi^2_{1-\alpha,k-1}$.

Remark. Since $n_i\widehat{S}_i = \sum_{j=1}^{N_i}\left(\mathbf{1}'\mathbf{X}_{ij} - \mathbf{1}'\overline{\mathbf{X}}_i\right)^2$, QH is Bartlett's test of homogeneity of variances [12] applied to the univariate sums $\mathbf{1}'\mathbf{X}_{ij}$. Under normality, $n_i\widehat{S}_i / \left(\mathbf{1}'\mathbf{\Sigma}_i\mathbf{1}\right)$ has an exact chi-squared distribution with n_i degrees of freedom; therefore, under H_0, the distribution of QH is free of p and of the covariance structure. However, QH responds only to differences among the scalars $\mathbf{1}'\mathbf{\Sigma}_i\mathbf{1}$, i = 1, . . . , k; its power equals its size against any alternative with $\mathbf{1}'\mathbf{\Sigma}_1\mathbf{1} = \cdots = \mathbf{1}'\mathbf{\Sigma}_k\mathbf{1}$, even when the covariance matrices differ.

3. Methodology

The Type I error rates and powers of SY, AHM, and QH (Equations (2)–(4)) were estimated by Monte Carlo simulation, following [2], [7], and [8], under the first-order autoregressive (AR(1)), banded Toeplitz (BTOEP), and simple diagonal (SIM) covariance structures described in Section 3.1. Each setting was replicated 5,000 times, so the Monte Carlo standard error of an empirical rate near 0.05 is about 0.003. All computations used Python 3.13 with NumPy 2.5 and SciPy 1.18 (random generator seed 20260815); normal vectors were generated through the Cholesky factor of the covariance matrix, and the asymptotic critical values of Section 2 were used at the nominal significance level $\alpha = 0.05$.

3.1. **Data Simulation.** In each replication, independent p-variate normal samples with zero mean vectors were drawn from three populations ($k = 3$). The sample sizes N_i were equal and took five values: 10, 20, 40, 80, and 160, ranging from very small to very large samples, and the dimension took the values $p \in \{10, 20, 40, 80, 160\}$, with $p \geq N_i$ (i = 1, 2, 3). To evaluate the Type I error rate and power, the covariance matrices were specified under the following three structures.

Case 1: First-order autoregressive (AR(1)) structure:

$$H_0^{(1)} : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}_3 = \mathbf{A}_0, \qquad H_1^{(1)} : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{A}_0,\ \mathbf{\Sigma}_3 = \mathbf{A}_1,$$

where $\mathbf{A}_0 = \left(0.7^{|s-t|}\right)_{p\times p}$, and $\mathbf{A}_1 = \left(0.9^{|s-t|}\right)_{p\times p}$, $s, t = 1, \ldots, p$.

Case 2: Banded Toeplitz (BTOEP) structure:

$$H_0^{(2)} : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}_3 = \mathbf{\Gamma}_0, \qquad H_1^{(2)} : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Gamma}_0,\ \mathbf{\Sigma}_3 = \mathbf{\Gamma}_1,$$

where $\mathbf{\Gamma}_0$ is the $p \times p$ tridiagonal Toeplitz matrix with ones on the main diagonal, -0.5 on the first sub- and super-diagonals, and zeros elsewhere, and $\mathbf{\Gamma}_1$ is defined in the same way with -0.05 in place of -0.5.

Case 3: Simple diagonal (SIM) structure:

$$H_0^{(3)} : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}_3 = \mathbf{D}_0, \qquad H_1^{(3)} : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{D}_0,\ \mathbf{\Sigma}_3 = \mathbf{D}_1,$$

where $\mathbf{D}_0 = \operatorname{diag}\left(d_1, \ldots, d_p\right)$, $\mathbf{D}_1 = \operatorname{diag}\left(u_1, \ldots, u_p\right)$ in which d_s are i.i.d. Uniform (0,1), and u_s are i.i.d. Uniform (0.5,1.5), for $s = 1, \ldots, p$. These values are generated anew in each replication.

3.2. **Empirical Type I Error Rate and Empirical Power.** The empirical Type I error rate is the proportion of the 5,000 replications in which H_0 is rejected when it is true, and the empirical power is the corresponding proportion under H_1. Following [9], we say that a test keeps its nominal level when its empirical rate falls within Bradley's interval [0.025, 0.075]; following [10], it is called robust when the rate falls within the narrower Cochran's interval [0.04, 0.06]. In Tables 1–3, rates outside Bradley's interval are marked with *, rates inside Bradley's but outside Cochran's interval with †, and power is shown only when the corresponding Type I error rate is within Bradley's interval.

4. Results and Discussion

The simulated rejection rates for the AR(1), BTOEP, and SIM structures are listed in Tables 1, 2, and 3, respectively; every SY result is based on the corrected estimator $\widehat{a_4}$ of Section 2.1. Differences of less than about 0.006 (two Monte Carlo standard errors) from the nominal level should not be over-interpreted.

4.1. **Results under the AR(1) Structure.** Under the AR(1) structure (Table 1), QH keeps the nominal level for every combination of p and N_i; its empirical rates (0.0448–0.0576) lie inside Cochran's interval throughout. SY also keeps the nominal level in all combinations, its rates ranging from 0.0450 to 0.0674; the three rates outside Cochran's interval ($N_i = 10$ with $p \le 20$, and $N_i = 20$ with $p = 80$) exceed 0.06 by at most 0.0074. AHM is liberal for small samples: its rates exceed Bradley's upper bound for $N_i = 10$ (except $p = 80$) and for $N_i = 20$ with $p \le 40$, reaching 0.1040 for $N_i = p = 10$, and they fall inside Cochran's interval only for $(N_i, p) = (80, 160)$ and $(160, 160)$. Part of this liberality is due to σ_0^2, which underestimates the null variance of $\widehat{F_k}$ (Section 2.2): under AR(1), the exact null variance is $1.66\sigma_0^2$ for $N_i = p = 10$ but only $1.05\sigma_0^2$ for $N_i = p = 160$.

TABLE 1. Simulated Type I error rates and powers of the SY, AHM, and QH tests, AR(1) structure

N_i	p	Type I Error Rate			Power		
		SY	AHM	QH	SY	AHM	QH
10	10	0.0674†	0.1040*	0.0516	0.1428	-	0.1204
	20	0.0606†	0.0876*	0.0570	0.2384	-	0.2178
	40	0.0574	0.0910*	0.0520	0.3828	-	0.3190
	80	0.0470	0.0744†	0.0492	0.5676	0.4878	0.3894
	160	0.0450	0.0768*	0.0478	0.7548	-	0.4084
20	20	0.0576	0.0910*	0.0448	0.3984	-	0.4426
	40	0.0538	0.0846*	0.0464	0.6752	-	0.6404
	80	0.0610†	0.0710†	0.0576	0.8974	0.8258	0.7242
	160	0.0480	0.0690†	0.0460	0.9898	0.8950	0.7510
40	40	0.0504	0.0742†	0.0464	0.9208	0.9602	0.9118
	80	0.0506	0.0646†	0.0560	0.9964	0.9928	0.9524
	160	0.0492	0.0632†	0.0490	0.9998	0.9994	0.9702
80	80	0.0480	0.0630†	0.0458	1.0000	1.0000	0.9998
	160	0.0554	0.0580	0.0478	1.0000	1.0000	0.9996
160	160	0.0516	0.0554	0.0500	1.0000	1.0000	1.0000

Note: * rate below 0.025 or above 0.075; † rate between 0.025 and 0.075 but outside [0.04, 0.06]; - power omitted because the test does not keep its nominal level.

The power of all three tests increases with p and N_i. Of the tests that keep the nominal level, SY is the most powerful in most combinations, and its margin is largest when p is large relative to N_i (e.g., 0.7548 for SY versus 0.4084 for QH when $N_i = 10$ and $p = 160$). QH is more powerful than SY only for $N_i = p = 20$ (0.4426 versus 0.3984). For fixed N_i, the power of QH grows slowly with p because the ratio $\mathbf{1}'\mathbf{A}_1\mathbf{1}/\mathbf{1}'\mathbf{A}_0\mathbf{1}$, which alone determines its power (Remark in Section 2.3), increases only from about 1.75 at $p = 10$ to 3.21 at $p = 160$. Where AHM attains an acceptable Type I error rate, its power usually lies between those of QH and SY; it exceeds that of SY only for $N_i = p = 40$ (0.9602 versus 0.9208), although its Type I error rate there (0.0742) exceeds Cochran's upper bound. For $N_i \geq 80$, all three tests have power close to one.

4.2. **Results under the BTOEP Structure.** Under the BTOEP structure (Table 2), QH again keeps the nominal level in every combination, its rates ranging from 0.0408 to 0.0590. SY stays inside Bradley's interval throughout (0.0292–0.0600) but becomes somewhat conservative when p is much larger than N_i (0.0362 and 0.0292 for $N_i = 10$ with $p = 80$ and 160, and 0.0392 for $N_i = 20$ with $p = 160$). AHM exceeds Bradley's upper bound for $N_i = 10$ with $p \leq 40$ and for $N_i = p = 20$; its rates lie inside Cochran's interval for $N_i \geq 40$ and for $N_i = 20$ with $p = 40$. These rates are generally lower than under AR(1). This agrees with Section 2.2, because the ratio $\frac{\mathrm{tr}(\boldsymbol{\Sigma}^4)}{[\mathrm{tr}(\boldsymbol{\Sigma}^2)]^2}$, and hence the underestimation of the null variance by σ_0^2, is smaller for $\boldsymbol{\Gamma}_0$ (about $1.9/p$) than for $\mathbf{A}_0$ (about $6/p$).

TABLE 2. Simulated Type I error rates and powers of the SY, AHM, and QH tests, BTOEP structure

N_i	p	Type I Error Rate			Power		
		SY	AHM	QH	SY	AHM	QH
10	10	0.0600	0.0820*	0.0452	0.1094	-	0.9138
	20	0.0534	0.0828*	0.0590	0.1244	-	0.9916
	40	0.0500	0.0818*	0.0478	0.1534	-	0.9996
	80	0.0362†	0.0730†	0.0446	0.1624	0.2404	1.0000
	160	0.0292†	0.0738†	0.0532	0.1484	0.2434	1.0000
20	20	0.0518	0.0774*	0.0478	0.2688	-	1.0000
	40	0.0490	0.0586	0.0408	0.4098	0.4692	1.0000
	80	0.0484	0.0712†	0.0534	0.5662	0.4628	1.0000
	160	0.0392†	0.0618†	0.0544	0.6674	0.4828	1.0000
40	40	0.0492	0.0564	0.0526	0.8248	0.8834	1.0000
	80	0.0574	0.0570	0.0488	0.9668	0.9006	1.0000
	160	0.0462	0.0564	0.0516	0.9956	0.9038	1.0000
80	80	0.0534	0.0548	0.0492	1.0000	1.0000	1.0000
	160	0.0492	0.0558	0.0470	1.0000	1.0000	1.0000
160	160	0.0556	0.0546	0.0522	1.0000	1.0000	1.0000

Note: * rate below 0.025 or above 0.075; † rate between 0.025 and 0.075 but outside [0.04, 0.06]; - power omitted because the test does not keep its nominal level.

The power comparison is dominated by QH, whose power is at least 0.9138 for $N_i = 10$ and equal to 1.0000 for $N_i \geq 20$, whereas SY has power of only 0.1094–0.1624 for $N_i = 10$ and AHM about 0.24 where its Type I error rate is acceptable. This advantage of QH is a consequence of the alternative: $\mathbf{1}'\boldsymbol{\Gamma}_0\mathbf{1} = 1$ whereas $\mathbf{1}'\boldsymbol{\Gamma}_1\mathbf{1} = 0.9p + 0.1$, so the variance of the sum of the variables differs greatly between the populations, while the trace-based quantities used by SY and AHM change much less ($\mathrm{tr}\left(\boldsymbol{\Gamma}_0{}^2\right)/p \approx 1.5$ versus $\mathrm{tr}\left(\boldsymbol{\Gamma}_1{}^2\right)/p \approx 1.0$). Between the two trace-based tests, SY is more powerful than AHM for $N_i = 20$ and 40 with $p \geq 80$ (e.g., 0.6674 versus 0.4828 for $N_i = 20$ and $p = 160$), whereas AHM is more powerful in the other four settings in which both powers are reported: $N_i = 10$ with $p = 80$ and 160 (0.2404 and 0.2434 versus 0.1624 and 0.1484), where its Type I error rates are inflated (0.0730 and 0.0738), $N_i = 20$ with $p = 40$ (0.4692 versus 0.4098), and $N_i = p = 40$ (0.8834 versus 0.8248). For $N_i \geq 80$, all three tests have power equal to one.

4.3. **Results under the SIM Structure.** Under the SIM structure (Table 3), QH keeps the nominal level in all combinations (0.0448–0.0542). SY has rates between 0.0430 and 0.0546 in all but two combinations: it is slightly conservative for $N_i = 20$ with $p = 160$ (0.0398) and falls just below Bradley's lower bound for $N_i = 10$ with $p = 160$ (0.0246). AHM is again liberal for small samples, exceeding Bradley's upper bound for $N_i = 10$ with $p \leq 80$ (0.0764–0.0888) and lying between

Cochran's and Bradley's upper bounds for $N_i = 10$ with $p = 160$ and for $N_i = 20$ (0.0626–0.0730); for $N_i \geq 40$ its rates are between 0.0518 and 0.0602.

TABLE 3. Simulated Type I error rates and powers of the SY, AHM, and QH tests, SIM structure

N_i	p	Type I Error Rate			Power		
		SY	AHM	QH	SY	AHM	QH
10	10	0.0546	0.0870*	0.0516	0.6542	-	0.1848
	20	0.0540	0.0888*	0.0488	0.7884	-	0.1802
	40	0.0450	0.0802*	0.0538	0.8536	-	0.1656
	80	0.0430	0.0764*	0.0488	0.8508	-	0.1612
	160	0.0246*	0.0730†	0.0532	-	0.6318	0.1720
20	20	0.0538	0.0712†	0.0512	0.9806	0.8988	0.3342
	40	0.0494	0.0638†	0.0458	0.9988	0.9400	0.3348
	80	0.0472	0.0694†	0.0512	1.0000	0.9602	0.3336
	160	0.0398†	0.0626†	0.0460	1.0000	0.9652	0.3274
40	40	0.0514	0.0588	0.0448	1.0000	0.9998	0.5990
	80	0.0470	0.0560	0.0468	1.0000	0.9998	0.6144
	160	0.0470	0.0544	0.0486	1.0000	0.9998	0.6026
80	80	0.0508	0.0602†	0.0542	1.0000	1.0000	0.8836
	160	0.0528	0.0530	0.0490	1.0000	1.0000	0.8894
160	160	0.0528	0.0518	0.0508	1.0000	1.0000	0.9926

Note: * rate below 0.025 or above 0.075; † rate between 0.025 and 0.075 but outside [0.04, 0.06]; - power omitted because the test does not keep its nominal level.

SY is the most powerful test under the SIM structure. Its power is 0.6542–0.8536 for $N_i = 10$ ($p \leq 80$), at least 0.9806 for $N_i = 20$, and equal to one for $N_i \geq 40$. AHM is somewhat less powerful (0.8988–0.9652 for $N_i = 20$) and has power of at least 0.9998 for $N_i \geq 40$. QH has much lower power: about 0.17 for $N_i = 10$, 0.33 for $N_i = 20$, 0.60 for $N_i = 40$, 0.89 for $N_i = 80$, and 0.99 for $N_i = 160$, with almost no dependence on p. These values are close to the power of Bartlett's test for a variance ratio of 2 (0.17, 0.33, 0.61, 0.90, and 1.00 for the same sample sizes), which agrees with the Remark in Section 2.3, because the ratio $\mathbf{1}'\mathbf{D}_1\mathbf{1}/\mathbf{1}'\mathbf{D}_0\mathbf{1}$ fluctuates around 2 for every p. Thus, under heterogeneous diagonal covariance matrices, the trace-based tests SY and AHM detect differences much more efficiently than QH.

5. Conclusion

Using 5,000 Monte Carlo replications, we compared the Srivastava–Yanagihara (SY), Ahmad (AHM), and Qayed–Han (QH) tests of equality of three covariance matrices in high dimensions, for the AR(1), BTOEP, and SIM structures with $N_i = 10$–160 and $p \geq N_i$. SY was computed with the corrected estimator $\widehat{a_4}$ of Section 2.1. Three conclusions emerge. First, QH and SY keep the nominal level well. All 45 rates of QH are within Cochran's interval (0.0408–0.0590), and those of SY lie

inside Bradley's interval in 44 of the 45 settings, the exception being a slightly conservative rate of 0.0246 for $N_i = 10$ and $p = 160$ under SIM. AHM, in contrast, exceeds Bradley's upper bound in 14 settings, all with $N_i \leq 20$, with rates up to 0.1040, partly because σ_0^2 underestimates the null variance of $\widehat{F}_k$ when N_i and p are small, and stays inside Bradley's interval for $N_i \geq 40$. Second, which test is most powerful depends on the way in which the covariance matrices differ. Among tests with an acceptable Type I error rate, SY attains the largest power in 13 of the 15 AR(1) settings and in all SIM settings in which its Type I error rate is acceptable, whereas QH is the most powerful test under BTOEP, with power of at least 0.91 even for $N_i = 10$. Third, these differences are explained by the quantities on which the tests are based: SY and AHM use traces of the sample covariance matrices, whereas QH depends on the data only through the sums $\mathbf{1}'\mathbf{X}_{ij}$ and is equivalent to Bartlett's test applied to these sums. QH is therefore very powerful when the alternative changes $\mathbf{1}'\boldsymbol{\Sigma}_i\mathbf{1}$, as in the BTOEP design, but its power equals its size when $\mathbf{1}'\boldsymbol{\Sigma}_1\mathbf{1} = \cdots = \mathbf{1}'\boldsymbol{\Sigma}_k\mathbf{1}$, even if the covariance matrices differ.

In practice, when the covariance structure is unknown, SY is recommended as the primary test of covariance homogeneity for high-dimensional normal data, because it combines reliable Type I error control with the highest power under the AR(1) and SIM structures. QH is a useful complement: its null distribution is exact under normality, and it is very sensitive to differences in the variance of the sum of the variables, but a non-significant QH result should not be taken as evidence of homogeneity. AHM should be used only when $N_i \geq 40$, unless its critical values are calibrated, for example by resampling.

Because AHM and QH were developed for non-normal data, whereas SY assumes normality, future studies should compare the three tests under non-normal distributions, such as skewed distributions (e.g., [11]) and data contaminated by outliers. Their power should also be examined against alternatives in which the covariance matrices differ but $\mathbf{1}'\boldsymbol{\Sigma}_i\mathbf{1}$ is the same for all populations, because QH cannot detect such differences. Unequal sample sizes, more than three populations, and applications to real data would further clarify the practical limitations of the tests.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] A. Qayed, D. Han, Homogeneity Test of Several High-Dimensional Covariance Matrices for Stationary Processes under Non-Normality, Commun. Stat. Theory Methods 52 (2023), 2783–2798. https://doi.org/10.1080/03610926.2021.1960375.

[2] S. Chaipitak, B. Choopradit, A New Test for Equality of Two Covariance Matrices in High-Dimensional Data, Math. Stat. 12 (2024), 455–464. https://doi.org/10.13189/ms.2024.120507.

[3] M.R. Ahmad, Testing Homogeneity of Several Covariance Matrices and Multi-Sample Sphericity for High-Dimensional Data under Non-Normality, Commun. Stat. Theory Methods 46 (2017), 3738–3753. https://doi.org/10.1080/03610926.2015.1073310.

[4] J.R. Schott, A Test for the Equality of Covariance Matrices When the Dimension Is Large Relative to the Sample Sizes, Comput. Stat. Data Anal. 51 (2007), 6535–6542. https://doi.org/10.1016/j.csda.2007.03.004.

[5] M.S. Srivastava, H. Yanagihara, Testing the Equality of Several Covariance Matrices with Fewer Observations than the Dimension, J. Multivar. Anal. 101 (2010), 1319–1329. https://doi.org/10.1016/j.jmva.2009.12.010.

[6] P. Zhong, R. Li, S. Santo, Homogeneity Tests of Covariance Matrices with High-Dimensional Longitudinal Data, Biometrika 106 (2019), 619–634. https://doi.org/10.1093/biomet/asz011.

[7] S. Chaipitak, J. Jaipetch, K. Wichayacheewin, W. Suwanrach, B. Choopradit, Efficiency Evaluation of Statistical Tests for Homogeneity of Variances under Normal, Beta, and Weibull Distributional Frameworks, Int. J. Anal. Appl. 23 (2025), 240. https://doi.org/10.28924/2291-8639-23-2025-240.

[8] S. Chaipitak, A Novel Test for the Homogeneity of Several Covariance Matrices in High-Dimensional Data: Application to Gene Expression Analysis, WSEAS Trans. Math. 25 (2026), 36–50. https://doi.org/10.37394/23206.2026.25.5.

[9] J.V. Bradley, Robustness?, Br. J. Math. Stat. Psychol. 31 (1978), 144–152. https://doi.org/10.1111/j.2044-8317.1978.tb00581.x.

[10] W.G. Cochran, The χ^2 Test of Goodness of Fit, Ann. Math. Stat. 23 (1952), 315–345. https://doi.org/10.1214/aoms/1177729380.

[11] S. Wasinrat, S. Chaipitak, P. Thammachote, B. Choopradit, The G-Bell Family of Distributions, Int. J. Math. Comput. Sci. 20 (2025), 1033–1042. https://doi.org/10.69793/ijmcs/04.2025/wctc.

[12] M.S. Bartlett, Properties of Sufficiency and Statistical Tests, Proc. R. Soc. Lond. A 160 (1937), 268–282. https://doi.org/10.1098/rspa.1937.0109.