

Diagonal and unscaled Wald-type tests in general factorial designs

Lukasz Smaga

Faculty of Mathematics and Computer Science

Adam Mickiewicz University

Umultowska 87, 61–614 Poznań, Poland

e-mail: ls@amu.edu.pl

Abstract: In this paper, the asymptotic and permutation testing procedures are developed in general factorial designs without assuming homoscedasticity or a particular error distribution. The one-way layout, crossed and hierarchically nested designs are contained in our general framework. New test statistics are modifications of Wald-type statistic, where a weight matrix is a certain diagonal matrix. Asymptotic properties of the new solutions are also investigated. In particular, the consistency of the tests under fixed alternatives or asymptotic validity of the permutation procedures are proved in many cases. Simulation studies show that, in the case of small sample sizes, some of the proposed methods perform comparably to or even better in certain situations than the Wald-type permutation test of Pauly et al. (2015). Illustrative real data examples of the use of the tests in practice are also given.

MSC 2010 subject classifications: Primary 62G09, 62G10; secondary 62K15.

Keywords and phrases: Diagonal Wald-type test, general factorial design, hypothesis testing, Moore-Penrose inverse, permutation test, unscaled Wald-type test.

Received May 2016.

Contents

1	Introduction	2614
2	Diagonal Wald-type test	2616
2.1	Diagonal Wald-type permutation test	2618
2.2	Diagonal Wald-type tests based on standardized test statistic	2619
3	Unscaled Wald-type test	2620
4	Simulation experiments	2623
4.1	Simulation design	2623
4.2	Simulation results	2624
4.2.1	Size control	2624
4.2.2	Power	2625
4.2.3	Summary	2628
5	Real data illustrative examples	2629
5.1	One-way layout	2629
5.2	Two-way cross-classification design	2629

5.3	Nested hierarchical design	2630
6	Concluding remarks	2630
A	Proofs	2631
B	Asymptotic standardized diagonal and unscaled Wald-type tests . .	2632
C	Additional simulation studies	2634
D	Simulation results tables	2634
	Acknowledgements	2644
	References	2644

1. Introduction

Let us consider the following general factorial design introduced by Pauly et al. [20]. We consider independent observations

$$X_{ij} = \mu_i + \varepsilon_{ij}, \quad (1.1)$$

$i = 1, \dots, d$, $j = 1, \dots, n_i$, where ε_{ij} , $j = 1, \dots, n_i$ are i.i.d. random errors satisfying the following conditions:

$$E(\varepsilon_{i1}) = 0, \quad E(\varepsilon_{i1}^2) = \sigma_i^2 > 0, \quad E(\varepsilon_{i1}^4) < \infty, \quad i = 1, \dots, d. \quad (1.2)$$

The total sample size is denoted by $N = \sum_{i=1}^d n_i$. In order to derive asymptotic results, we will make the following assumption:

$$\lim_{N \rightarrow \infty} \frac{n_i}{N} = \kappa_i > 0, \quad i = 1, \dots, d. \quad (1.3)$$

It is worth noting that different variances, samples sizes as well as distributions of error terms are allowed as long as assumptions in (1.2) hold.

Let \mathbf{I}_n be the $n \times n$ identity matrix and let $\mathbf{1}_n$ be the $n \times 1$ vector of ones. In matrix notation, (1.1) can be written as $\mathbf{X} = \text{diag}(\mathbf{1}_{n_1}, \dots, \mathbf{1}_{n_d})\boldsymbol{\mu} + \boldsymbol{\varepsilon}$, where $\mathbf{X} = (X_{11}, \dots, X_{dn_d})'$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)'$ and $\boldsymbol{\varepsilon} = (\varepsilon_{11}, \dots, \varepsilon_{dn_d})'$ with $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\text{Cov}(\boldsymbol{\varepsilon}) = \text{diag}(\sigma_1^2 \mathbf{I}_{n_1}, \dots, \sigma_d^2 \mathbf{I}_{n_d}) > \mathbf{0}$. In our general design, a factorial structure within the components of the vector $\boldsymbol{\mu}$ by splitting up the indices is allowed (the index i is appropriately split in subindices i_1, i_2, \dots). In this way, we can consider for example the one-way layout and crossed and hierarchically nested designs (see [20] and Section 5, for more detail and some real data examples).

To formulate a general hypothesis testing problem, we need a contrast matrix \mathbf{H} , i.e., $\mathbf{H}\mathbf{1} = \mathbf{0}$, where $\mathbf{1}$ is the column vector of ones of the appropriate size. We are interested in testing the null hypothesis $H_0 : \mathbf{H}\boldsymbol{\mu} = \mathbf{0}$. This hypothesis is equivalent to $H_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$, where $\mathbf{T} = \mathbf{H}'(\mathbf{H}\mathbf{H}')^{-1}\mathbf{H}$ is the unique projection matrix (\mathbf{M}^- denotes a generalized inverse of \mathbf{M}), which is symmetric and idempotent. For instance, in the one-way layout where factor A has a levels, the centring matrix $\mathbf{T} = \mathbf{P}_a = \mathbf{I}_a - (1/a)\mathbf{1}_a\mathbf{1}_a'$ is used for testing the hypothesis of no treatment effect $H_0^A : \mu_1 = \dots = \mu_a$. Other examples of contrast matrix \mathbf{H} and projection matrix \mathbf{T} are given in [20] and in Section 5.

For testing H_0 , there exist many inference methods for normal or homoscedastic models (see [20] and the references therein). Nevertheless, such assumptions are often not met or it is difficult to check them in practice. To avoid these limitations, Pauly et al. [20] proposed the testing procedures based on the Wald-type statistic (WTS)

$$Q_N(\mathbf{T}) = N\bar{\mathbf{X}}'\mathbf{T}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T})^+\mathbf{T}\bar{\mathbf{X}},$$

where $\bar{\mathbf{X}} = (\bar{X}_1, \dots, \bar{X}_d)'$ is the vector of the means $\bar{X}_i = 1/n_i \sum_{j=1}^{n_i} X_{ij}$, $\hat{\mathbf{V}}_N = N \cdot \text{diag}(\hat{\sigma}_1^2/n_1, \dots, \hat{\sigma}_d^2/n_d)$, $\hat{\sigma}_i^2 = 1/(n_i - 1) \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$ is the empirical variance of $\mathbf{X}_i = (X_{i1}, \dots, X_{in_i})'$, $i = 1, \dots, d$, and \mathbf{M}^+ denotes the Moore-Penrose inverse of \mathbf{M} [11]. The asymptotic WTS test is given by $\varphi_N = \{Q_N(\mathbf{T}) > \chi_{\text{rank}(\mathbf{T}), 1-\alpha}^2\}$, where $\chi_{p,\alpha}^2$ is the α -quantile from the χ_p^2 -distribution. This test is asymptotically exact, but it requires large sample sizes to keep the nominal type I error level.

Pauly et al. [20] proposed a permutation test based on WTS to overcome this problem. They applied a modified permutation principle [12, 18]. This principle can be used in situations where exchangeability of the data does not hold. The test retains the finite exactness property under exchangeability. Moreover, it is even asymptotically exact and consistent when the data are not exchangeable. From simulations, the Wald-type permutation statistic (WTPS) test tends to result in accurate test decision for small sample sizes in many cases, but it is also more or less liberal for extremely skewed distributions (like log-normal one) in the case of unequal variances. Small sample sizes usually means that there are a few or over a dozen observations in each sample, which of course depends on the number of groups.

In the Wald-type statistic proposed by Pauly et al. [20], the Moore-Penrose inverse of $\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}$ is used as a so-called weight matrix. However, it can be chosen in other way resulting in a modified Wald-type statistic. The choice of a weight matrix may be significant for performance of a test (see, for example, Duchesne and Francq [7]). Smaga [23] considered the asymptotic and permutation tests based on modified WTS, where the Moore-Penrose inverse is replaced by a $\{2\}$ -inverse, i.e., a matrix satisfying the second relation defining the Moore-Penrose inverse [7, 9]. Under some assumptions, these testing procedures are also asymptotically valid but, however, consistent for a smaller class of fixed alternatives than the tests based on WTS and WTPS. For extremely skewed distributions, heteroscedastic designs and small sample sizes, the methods based on $\{2\}$ -inverses seem to be a more conservative replacement for the WTPS. However, they may perform worse under symmetric distributions, i.e., they may be more conservative or more liberal than the WTPS test.

The testing procedures of Pauly et al. [20] and Smaga [23] are constructed without assumption of equal sample sizes, equal variances and a particular distribution of the errors. However, no one of them is better than the other and in general they do not perform satisfactory well for extremely skewed distributions, heteroscedastic designs and small sample sizes. In this paper, we propose new inference methods based on the ideas from tests for problems in high-dimensional data analysis. More precisely, we consider modified Wald-type statistics, where

a weight matrix is a certain diagonal matrix, which may be related to $\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}$, or not. In such a way, the singularity problem of the matrix $\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}$ is circumvented. We consider the asymptotic and permutation methods to approximate null distribution of a test statistic. Simulation studies show that some of our new solutions perform comparable to or even better in certain scenarios than existing competitors.

The remainder of the paper is organized as follows. New testing procedures are introduced in Sections 2 and 3. Their properties are also given there. Section 4 contains a Monte Carlo simulation study providing an idea of the size control and power of the tests. Illustrative real data examples are presented in Section 5. Some conclusions are given in Section 6. Proofs are outlined in the Appendix.

2. Diagonal Wald-type test

For high-dimensional low sample size data, the Hotelling's T^2 test suffers from a singularity problem in the covariance matrix estimation and therefore is not valid in that setting. To overcome this problem, some remedies are proposed in the literature. One of them is the assumption of diagonal covariance matrix. This idea was first considered by Wu et al. [29] and further investigated by Dong et al. [6], Park and Nag Ayyala [19], Srivastava [25], Srivastava and Du [26], Srivastava et al. [27]. Here, we use this idea to handle with the problem of singularity of the matrix $\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}$. Specifically, we propose the following test statistic

$$Q_N^D(\mathbf{T}) = N\bar{\mathbf{X}}'\mathbf{T}\{\text{diag}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T})\}^{-1}\mathbf{T}\bar{\mathbf{X}},$$

where $\text{diag}(\mathbf{M})$ denotes the diagonal matrix with diagonal entries of the quadratic matrix \mathbf{M} , and the vector $\bar{\mathbf{X}}$ and the matrix $\hat{\mathbf{V}}_N$ were defined in Section 1. To methods based on $Q_N^D(\mathbf{T})$, we refer as the diagonal Wald-type tests. The null hypothesis is rejected for large values of $Q_N^D(\mathbf{T})$.

For small sample sizes, the sample variances in $\hat{\mathbf{V}}_N$ are perhaps not reliable estimators. This may have a negative effect on testing procedures which use this matrix. Since the tests based on $Q_N^D(\mathbf{T})$ only use the diagonal elements of $\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}$, a negative effect of unreliable estimation seems to be smaller for these tests than for the WTS and WTPS ones, where whole matrix $\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}$ is used. On the other hand, the diagonal Wald-type tests do not use the information from the off-diagonal elements of the matrix $\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}$, in contrast to the WTS and WTPS testing procedures. However, our results indicate that negative effect of the off-diagonal elements of $\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}$ is stronger than the information from them (see Section 4), which is favourable for the diagonal Wald-type tests.

We should check that the statistic $Q_N^D(\mathbf{T})$ is well defined. By definition of the estimator $\hat{\mathbf{V}}_N$, the i th diagonal element of $\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}$ is of the form $\mathbf{T}_i\hat{\mathbf{V}}_N\mathbf{T}'_i = N\sum_{j=1}^d T_{ij}^2\hat{\sigma}_j^2/n_j$, where $\mathbf{T}_i = (T_{i1}, \dots, T_{id})$ is the i th row of the projection matrix \mathbf{T} . Since $\hat{\sigma}_j^2 > 0, j = 1, \dots, d$ with probability one, the i th diagonal element $\mathbf{T}_i\hat{\mathbf{V}}_N\mathbf{T}'_i$ of the matrix $\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}$ is equal to zero if and only if $\mathbf{T}_i = \mathbf{0}'_d$. However,

this is impossible, because \mathbf{T} is a projection matrix. Hence, $\{\text{diag}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T})\}^{-1}$ exists with probability one and the diagonal Wald-type statistic is well defined.

First, we construct the asymptotic test based on the diagonal Wald-type statistic. The asymptotic null distribution of $Q_N^D(\mathbf{T})$ is given in the following theorem. Throughout the paper, \xrightarrow{d} and \xrightarrow{P} denote convergence in distribution and probability, respectively.

Theorem 2.1. *Under assumptions (1.2)–(1.3) and the null hypothesis $H_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$, if $r = \text{rank}(\mathbf{T})$, then $Q_N^D(\mathbf{T}) \xrightarrow{d} \sum_{i=1}^r \lambda_{D,i} Z_{D,i}^2$ as $N \rightarrow \infty$, where $\lambda_{D,1}, \dots, \lambda_{D,r}$ are the nonzero eigenvalues of the matrix $\{\text{diag}(\mathbf{T}\mathbf{V}\mathbf{T})\}^{-1}\mathbf{T}\mathbf{V}\mathbf{T}$, $\mathbf{V} = \text{diag}(\sigma_1^2/\kappa_1, \dots, \sigma_d^2/\kappa_d)$, and $Z_{D,1}, \dots, Z_{D,r}$ are the independent $N(0, 1)$ random variables.*

By Theorem 2.1, the asymptotic null distribution of the diagonal Wald-type statistic is a central χ^2 -type mixture distribution (see [30]). This distribution can be approximated by a scaled χ^2 -distribution, i.e., by the distribution of $g_D \chi_{f_D}^2$ such that the first two moments coincide [2, 31]. The asymptotic critical value is then given by $k_{D,\alpha} = g_D \chi_{f_D, 1-\alpha}^2$. It is easy to see that

$$g_D = \frac{\text{Var}(\sum_{i=1}^r \lambda_{D,i} Z_{D,i}^2)}{2E(\sum_{i=1}^r \lambda_{D,i} Z_{D,i}^2)}, \quad f_D = \frac{2(E(\sum_{i=1}^r \lambda_{D,i} Z_{D,i}^2))^2}{\text{Var}(\sum_{i=1}^r \lambda_{D,i} Z_{D,i}^2)}. \quad (2.1)$$

The expected value and variance of $\sum_{i=1}^r \lambda_{D,i} Z_{D,i}^2$ are established in the following lemma.

Lemma 2.1. *Under notation of Theorem 2.1, we have $E(\sum_{i=1}^r \lambda_{D,i} Z_{D,i}^2) = d$ and $\text{Var}(\sum_{i=1}^r \lambda_{D,i} Z_{D,i}^2) = 2\text{trace}(\{\{\text{diag}(\mathbf{T}\mathbf{V}\mathbf{T})\}^{-1}\mathbf{T}\mathbf{V}\mathbf{T}\}^2)$.*

By (2.1) and Lemma 2.1, we obtain $g_D = \text{trace}(\{\{\text{diag}(\mathbf{T}\mathbf{V}\mathbf{T})\}^{-1}\mathbf{T}\mathbf{V}\mathbf{T}\}^2) / d$ and $f_D = d^2 / \text{trace}(\{\{\text{diag}(\mathbf{T}\mathbf{V}\mathbf{T})\}^{-1}\mathbf{T}\mathbf{V}\mathbf{T}\}^2)$. Since g_D and f_D involve some unknown quantities, they have to be estimated. The matrix \mathbf{V} is estimated by its consistent estimator $\hat{\mathbf{V}}_N$, so we consider the following simple estimators of g_D , f_D and $k_{D,\alpha}$: $\hat{g}_D = \text{trace}(\{\{\text{diag}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T})\}^{-1}\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}\}^2) / d$, $\hat{f}_D = d^2 / \text{trace}(\{\{\text{diag}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T})\}^{-1}\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}\}^2)$ and $\hat{k}_{D,\alpha} = \hat{g}_D \chi_{\hat{f}_D, 1-\alpha}^2$. The consistency of these estimators as well as the consistency of the test with critical region $\varphi_N^D = \{Q_N^D(\mathbf{T}) > \hat{k}_{D,\alpha}\}$ are established in the following theorem.

Theorem 2.2. *Under assumptions (1.2)–(1.3), $\hat{g}_D \xrightarrow{P} g_D$, $\hat{f}_D \xrightarrow{P} f_D$ and $\hat{k}_{D,\alpha} \xrightarrow{P} k_{D,\alpha}$ as $N \rightarrow \infty$. Moreover, under the alternative hypothesis $H_1 : \mathbf{T}\boldsymbol{\mu} = \mathbf{a} \neq \mathbf{0}_d$, $P(Q_N^D(\mathbf{T}) > \hat{k}_{D,\alpha} | H_1) \rightarrow 1$ as $N \rightarrow \infty$.*

Theorem 2.2 shows that the asymptotic testing procedure based on $Q_N^D(\mathbf{T})$ is consistent for all fixed alternatives, as the WTS test. Unfortunately, the asymptotic diagonal Wald-type test is also similar to the WTS test in the sense that it also requires large sample sizes to obtain a satisfactory approximation (see Section 4). As Pauly et al. [20], we consider a permutation testing procedure based on $Q_N^D(\mathbf{T})$ to improve the small sample behavior of the diagonal Wald-type test.

2.1. Diagonal Wald-type permutation test

Let π be a random permutation of N indices (uniformly distributed on the symmetric group of order N) that is independent from all other occurring random variables. Then, $\mathbf{X}^\pi = \pi(X_{11}, \dots, X_{dn_d})' = (X_{11}^\pi, \dots, X_{dn_d}^\pi)'$ denotes the permutation of the vector of observations \mathbf{X} . Let $\bar{\mathbf{X}}^\pi = (\bar{X}_1^\pi, \dots, \bar{X}_d^\pi)'$ be the vector of the means and $\hat{\mathbf{V}}_N^\pi = N \cdot \text{diag}(\hat{\sigma}_{1,\pi}^2/n_1, \dots, \hat{\sigma}_{d,\pi}^2/n_d)$ be the empirical covariance matrix of $\sqrt{N}\bar{\mathbf{X}}^\pi$. The value of $Q_N^D(\mathbf{T})$ computed from the permuted observations is of the form

$$Q_N^{D,\pi}(\mathbf{T}) = N\bar{\mathbf{X}}^{\pi'}\mathbf{T}\{\text{diag}(\mathbf{T}\hat{\mathbf{V}}_N^\pi\mathbf{T})\}^{-1}\mathbf{T}\bar{\mathbf{X}}^\pi.$$

A diagonal Wald-type permutation test is obtained by comparing $Q_N^D(\mathbf{T})$ with the $(1-\alpha)$ -quantile of the conditional distribution of $Q_N^{D,\pi}(\mathbf{T})$ given the observed data \mathbf{X} . The asymptotic conditional permutation distribution of $Q_N^{D,\pi}(\mathbf{T})$ is presented in the following theorem. The permutation distribution is the empirical distribution of a given test statistic recomputed over all permutations of the data.

Theorem 2.3. *Under assumptions (1.2)–(1.3), if $r = \text{rank}(\mathbf{T})$, then the permutation distribution of $Q_N^{D,\pi}(\mathbf{T})$ conditioned on the observed data \mathbf{X} weakly converges to $\sum_{i=1}^r \lambda_{D,\pi,i} Z_{D,\pi,i}^2$ as $N \rightarrow \infty$ in probability, where $\lambda_{D,\pi,1}, \dots, \lambda_{D,\pi,r}$ are the nonzero eigenvalues of $\{\text{diag}(\mathbf{T}\mathbf{D}\mathbf{T})\}^{-1}\mathbf{T}\mathbf{D}\mathbf{T}$, $\mathbf{D} = \text{diag}(1/\kappa_1, \dots, 1/\kappa_d)$, and $Z_{D,\pi,1}, \dots, Z_{D,\pi,r}$ are the independent standard normal random variables.*

From Theorem 2.3, it follows that the asymptotic conditional permutation distribution of $Q_N^{D,\pi}(\mathbf{T})$ is independent of $\mathbf{T}\boldsymbol{\mu}$. Moreover, it does not depend on the variances, unlike the asymptotic null distribution of $Q_N^D(\mathbf{T})$ (see Theorem 2.1). For this reason, these distributions are not generally the same. The following result shows the case where they are the same. It is a direct consequence of Theorems 2.1 and 2.3.

Corollary 2.1. *Under assumptions and notation of Theorems 2.1 and 2.3, if $\sigma_1^2 = \dots = \sigma_d^2$, then the unconditional null distribution of $Q_N^D(\mathbf{T})$ is the same as the permutation distribution of $Q_N^{D,\pi}(\mathbf{T})$ conditioned on the observed data \mathbf{X} , as $N \rightarrow \infty$.*

Remark 2.1.

1. Under the assumptions of Corollary 2.1, the asymptotic conditional permutation distribution of $Q_N^{D,\pi}(\mathbf{T})$ always approximates the asymptotic null distribution of $Q_N^D(\mathbf{T})$. This property is desirable for resampling procedures (see [4, 13, 14, 20, 23, 24], for other examples of such results). Thus, the test based on $Q_N^{D,\pi}(\mathbf{T})$ asymptotically maintains the nominal type I error level under the null hypothesis and is consistent for any fixed alternative (by Theorem 2.2 and Corollary 2.1).
2. Pauly et al. [20] proved the conditional distribution of the Wald-type permutation statistic (WTPS) $Q_N^\pi(\mathbf{T}) = N\bar{\mathbf{X}}^{\pi'}\mathbf{T}(\mathbf{T}\hat{\mathbf{V}}_N^\pi\mathbf{T})^+\mathbf{T}\bar{\mathbf{X}}^\pi$ always approximates the null distribution of $Q_N(\mathbf{T})$. Unfortunately, for the diagonal

Wald-type statistic, such result does not hold. The asymptotic conditional distribution of $Q_N^{D,\pi}(\mathbf{T})$ is the same as the asymptotic unconditional null distribution of $Q_N^D(\mathbf{T})$ for homoscedastic designs only. Although these distributions are not the same for heteroscedastic designs, simulations suggest that they are quite close to each other when the variances are not extremely different. As we will see in Section 4, for small sample sizes, the permutation test based on $Q_N^{D,\pi}(\mathbf{T})$ behaves quite well (even better than the $Q_N^\pi(\mathbf{T})$ test) in the case of different variances. For these reasons, it seems sensible to consider this test as a possible testing procedure in general framework introduced in Section 1.

2.2. Diagonal Wald-type tests based on standardized test statistic

Although the $Q_N^{D,\pi}(\mathbf{T})$ test works quite well for small sample sizes, we consider another procedure, which controls the nominal type I error level better than this test under very different variances, and both testing procedures have very similar empirical power (see Section 4). More precisely, we consider the standardized version of $Q_N^{D,\mathbf{V}_N}(\mathbf{T}) = N\bar{\mathbf{X}}'\mathbf{T}\{\text{diag}(\mathbf{T}\mathbf{V}_N\mathbf{T})\}^{-1}\mathbf{T}\bar{\mathbf{X}}.$, where $\mathbf{V}_N = N \cdot \text{diag}(\sigma_1^2/n_1, \dots, \sigma_d^2/n_d)$ (see, for instance, [21]), namely

$$\frac{Q_N^{D,\mathbf{V}_N}(\mathbf{T}) - E_{H_0}(Q_N^{D,\mathbf{V}_N}(\mathbf{T}))}{\sqrt{\text{Var}_{H_0}(Q_N^{D,\mathbf{V}_N}(\mathbf{T}))}}.$$

To simplify the analysis, we assume normality of the observations, i.e., $\varepsilon_{i1} \sim N(0, \sigma_i)$, $i = 1, \dots, d$. Then, $\sqrt{N}\bar{\mathbf{X}}. \sim N_d(\sqrt{N}\boldsymbol{\mu}, \mathbf{V}_N)$. Hence, under $H_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$, $\sqrt{N}\mathbf{T}\bar{\mathbf{X}}. \sim N_d(\mathbf{0}_d, \mathbf{T}\mathbf{V}_N\mathbf{T})$. By analysis similar to that in the proof of Lemma 2.1, we conclude that $E_{H_0}(Q_N^{D,\mathbf{V}_N}(\mathbf{T})) = d$ and $\text{Var}_{H_0}(Q_N^{D,\mathbf{V}_N}(\mathbf{T})) = 2\text{trace}(\{\{\text{diag}(\mathbf{T}\mathbf{V}_N\mathbf{T})\}^{-1}\mathbf{T}\mathbf{V}_N\mathbf{T}\}^2)$. To estimate \mathbf{V}_N , we use $\hat{\mathbf{V}}_N$, which is an L_2 -consistent estimator of it in the sense that $\hat{\mathbf{V}}_N\mathbf{V}_N^{-1} \xrightarrow{L_2} \mathbf{I}_d$ (see [20]). Therefore, we consider the following test statistic:

$$Q_N^{D,s}(\mathbf{T}) = \frac{Q_N^D(\mathbf{T}) - d}{\sqrt{2\text{trace}(\{\{\text{diag}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T})\}^{-1}\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}\}^2)}}.$$

By Zhang [30], we approximated the distribution of $Q_N^{D,s}(\mathbf{T})$ by a sequence of standardized χ^2 -distributions $(\chi_{h_{D,s}}^2 - h_{D,s})/\sqrt{2h_{D,s}}$. Unfortunately the $Q_N^{D,s}$ test works similar to or even worse than the Q_N^D test for small sample sizes, i.e., it is too liberal (see Section 4). This follows from the simulation results not included in the paper, but available from the author. The reason for this is that the speed of convergence of distribution of the $Q_N^{D,s}$ statistic to its asymptotic distribution is too slow. So, we do not consider this test in the paper, but the construction and asymptotic properties of it are given in the Appendix B for completeness. For this reason, we consider a permutation test based on

$$Q_N^{D,s,\pi}(\mathbf{T}) = \frac{Q_N^{D,\pi}(\mathbf{T}) - d}{\sqrt{2\text{trace}(\{\{\text{diag}(\mathbf{T}\hat{\mathbf{V}}_N^\pi\mathbf{T})\}^{-1}\mathbf{T}\hat{\mathbf{V}}_N^\pi\mathbf{T}\}^2)}}$$

only. This test performs better than the $Q_N^{D,\pi}(\mathbf{T})$ test for very small sample sizes or under very different variances. The asymptotic validity of this test in homoscedastic case is shown in the following theorem, which follows immediately from Theorems 2.1 and 2.3 and from $\hat{\mathbf{V}}_N \xrightarrow{P} \mathbf{V}$ and $\hat{\mathbf{V}}_N^\pi \xrightarrow{P} \sigma^2 \mathbf{D}$ (see Lemma 2 in the supplement to Pauly et al. [20]) as $N \rightarrow \infty$, where σ^2 is given by (3.1). Although we assumed normality to construct the standardized diagonal Wald-type statistic (more precisely to establish the expected value and variance of $Q_N^{D,\mathbf{V}^N}(\mathbf{T})$), the asymptotic properties of the tests based on it are proved without this assumption.

Theorem 2.4. *Under assumptions (1.2)–(1.3):*

1. *Under the null hypothesis $H_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$, if $r = \text{rank}(\mathbf{T})$, then*

$$Q_N^{D,s}(\mathbf{T}) \xrightarrow{d} \frac{\sum_{i=1}^r \lambda_{D,i} Z_{D,i}^2 - d}{\sqrt{2\text{trace}(\{\{\text{diag}(\mathbf{T}\mathbf{V}\mathbf{T})\}^{-1}\mathbf{T}\mathbf{V}\mathbf{T}\}^2)}}$$

as $N \rightarrow \infty$, where $\lambda_{D,1}, \dots, \lambda_{D,r}$, \mathbf{V} and $Z_{D,1}, \dots, Z_{D,r}$ are as in Theorem 2.1.

2. *If $r = \text{rank}(\mathbf{T})$, then the permutation distribution of $Q_N^{D,s,\pi}(\mathbf{T})$ conditioned on the observed data \mathbf{X} weakly converges to*

$$\frac{\sum_{i=1}^r \lambda_{D,\pi,i} Z_{D,\pi,i}^2 - d}{\sqrt{2\text{trace}(\{\{\text{diag}(\mathbf{T}\mathbf{D}\mathbf{T})\}^{-1}\mathbf{T}\mathbf{D}\mathbf{T}\}^2)}}$$

as $N \rightarrow \infty$ in probability, where $\lambda_{D,\pi,1}, \dots, \lambda_{D,\pi,r}$, \mathbf{D} and $Z_{D,\pi,1}, \dots, Z_{D,\pi,r}$ are as in Theorem 2.3.

3. *If $\sigma_1^2 = \dots = \sigma_d^2$, then the unconditional null distribution of $Q_N^{D,s}(\mathbf{T})$ is the same as the permutation distribution of $Q_N^{D,s,\pi}(\mathbf{T})$ conditioned on the observed data \mathbf{X} , as $N \rightarrow \infty$.*

Apart from the diagonal Wald-type testing procedures, in the next section, we also consider the unscaled Wald-type tests where we use the identity matrix as a weight matrix.

3. Unscaled Wald-type test

The other idea to improve the performance of the WTS is removing from it the Moore-Penrose inverse of $\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}$, i.e., we consider the unscaled Wald-type statistic of the form

$$Q_N^U(\mathbf{T}) = N\bar{\mathbf{X}}'\mathbf{T}\bar{\mathbf{X}}.$$

Bai and Saranadasa [1] considered this idea as the first for the Hotelling's T^2 statistic. Chen and Qin [3] and Zhang and Xu [32] extended this method for high-dimensional data. Recently, Duchesne and Francq [7] and Pauly et al. [21] used

this idea for multivariate hypothesis testing and to analyze high-dimensional one sample repeated measures designs, respectively.

The following result gives the asymptotic distribution of the unscaled Wald-type statistic under H_0 .

Theorem 3.1. *Under assumptions (1.2)–(1.3) and the null hypothesis $H_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$, if $r = \text{rank}(\mathbf{T})$, then $Q_N^U(\mathbf{T}) \xrightarrow{d} \sum_{i=1}^r \lambda_{U,i} Z_{U,i}^2$ as $N \rightarrow \infty$, where $\lambda_{U,1}, \dots, \lambda_{U,r}$ are the nonzero eigenvalues of $\mathbf{T}\mathbf{V}\mathbf{T}$, $\mathbf{V} = \text{diag}(\sigma_1^2/\kappa_1, \dots, \sigma_d^2/\kappa_d)$, and $Z_{U,1}, \dots, Z_{U,r}$ are the independent standard normal variables.*

To construct the asymptotic test based on $Q_N^U(\mathbf{T})$, we use the approximation by scaled χ^2 -distribution similarly as in Section 2, say $g_U \chi_{f_U}^2$. Then, $k_{U,\alpha} = g_U \chi_{f_U, 1-\alpha}^2$ is the asymptotic critical value. In much the same way as in the proof of Lemma 2.1, we obtain $E(\sum_{i=1}^r \lambda_{U,i} Z_{U,i}^2) = \text{trace}(\mathbf{T}\mathbf{V}\mathbf{T})$ and $\text{Var}(\sum_{i=1}^r \lambda_{U,i} Z_{U,i}^2) = 2\text{trace}([\mathbf{T}\mathbf{V}\mathbf{T}]^2)$ (by Theorem 3.1). Hence, $g_U = \text{trace}([\mathbf{T}\mathbf{V}\mathbf{T}]^2)/\text{trace}(\mathbf{T}\mathbf{V}\mathbf{T})$ and $f_U = [\text{trace}(\mathbf{T}\mathbf{V}\mathbf{T})]^2/\text{trace}([\mathbf{T}\mathbf{V}\mathbf{T}]^2)$. The estimators of g_U , f_U and $k_{U,\alpha}$ are of the form $\hat{g}_U = \text{trace}([\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}]^2)/\text{trace}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T})$, $\hat{f}_U = [\text{trace}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T})]^2/\text{trace}([\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}]^2)$ and $\hat{k}_{U,\alpha} = \hat{g}_U \chi_{\hat{f}_U, 1-\alpha}^2$. In the following result, the properties of them and of the asymptotic test based on $Q_N^U(\mathbf{T})$ are given. Its proof is similar to that of Theorem 2.2, and therefore it is omitted.

Theorem 3.2. *Under assumptions (1.2)–(1.3), $\hat{g}_U \xrightarrow{P} g_U$, $\hat{f}_U \xrightarrow{P} f_U$ and $\hat{k}_{U,\alpha} \xrightarrow{P} k_{U,\alpha}$ as $N \rightarrow \infty$. Furthermore, under the alternative hypothesis $H_1 : \mathbf{T}\boldsymbol{\mu} = \mathbf{a} \neq \mathbf{0}_d$, $P(Q_N^U(\mathbf{T}) > \hat{k}_{U,\alpha} | H_1) \rightarrow 1$ as $N \rightarrow \infty$.*

The finite sample behavior of the asymptotic unscaled Wald-type test is much better than that of the WTS test and the testing procedure based on $Q_N^D(\mathbf{T})$. However, it shows a tendency of conservativity in some situations. To overcome this problem, the first our idea was to consider a permutation test based on $Q_N^{U,\pi}(\mathbf{T}) = N\bar{\mathbf{X}}^{\pi'}\mathbf{T}\bar{\mathbf{X}}^\pi$. In the following theorem, the asymptotic conditional permutation distribution of $Q_N^{U,\pi}(\mathbf{T})$ is established.

Theorem 3.3. *Under assumptions (1.2)–(1.3), if $r = \text{rank}(\mathbf{T})$, then the permutation distribution of $Q_N^{U,\pi}(\mathbf{T})$ conditioned on the observed data \mathbf{X} weakly converges to $\sum_{i=1}^r \lambda_{U,\pi,i} Z_{U,\pi,i}^2$ as $N \rightarrow \infty$ in probability, where $\lambda_{U,\pi,1}, \dots, \lambda_{U,\pi,r}$ are the nonzero eigenvalues of the matrix $\sigma^2\mathbf{T}\mathbf{D}\mathbf{T}$,*

$$\sigma^2 = \sum_{i=1}^d \kappa_i \sigma_i^2 + \sum_{i=1}^d \kappa_i \left(\mu_i - \sum_{m=1}^d \kappa_m \mu_m \right)^2, \tag{3.1}$$

$\mathbf{D} = \text{diag}(1/\kappa_1, \dots, 1/\kappa_d)$, and $Z_{U,\pi,1}, \dots, Z_{U,\pi,r}$ are the independent $N(0, 1)$ random variables.

Unfortunately, Theorems 3.1 and 3.3 indicate that the $Q_N^{U,\pi}(\mathbf{T})$ test may not work well. First of all, the asymptotic permutation distribution of $Q_N^{U,\pi}(\mathbf{T})$ depends on the vector $\boldsymbol{\mu}$. Secondly, this distribution is rarely the same as the asymptotic null distribution of $Q_N^U(\mathbf{T})$. For example, they are the same when

$\sigma_1^2 = \dots = \sigma_d^2$ and $\mu_1 = \dots = \mu_d$. In fact, the simulations in Section 4 suggest that the unscaled Wald-type permutation test is too conservative or too liberal in certain cases. So, we further tried to improve this test. For this purpose, we consider the standardized version of $Q_N^U(\mathbf{T})$, i.e.,

$$\frac{Q_N^U(\mathbf{T}) - E_{H_0}(Q_N^U(\mathbf{T}))}{\sqrt{\text{Var}_{H_0}(Q_N^U(\mathbf{T}))}},$$

similarly as in Section 2. Under normality assumption, theorem on the moments of quadratic forms (see, for instance, [15], p. 55), shows that $E_{H_0}(Q_N^U(\mathbf{T})) = \text{trace}(\mathbf{T}\mathbf{V}_N\mathbf{T})$ and $\text{Var}_{H_0}(Q_N^U(\mathbf{T})) = 2\text{trace}([\mathbf{T}\mathbf{V}_N\mathbf{T}]^2)$. Estimating \mathbf{V}_N by $\hat{\mathbf{V}}_N$, we define

$$Q_N^{U,s}(\mathbf{T}) = \frac{Q_N^U(\mathbf{T}) - \text{trace}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T})}{\sqrt{2\text{trace}([\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}]^2)}}.$$

By simulation results not included in the paper, but available from the author, the approximation by a sequence of standardized χ^2 -distributions for $Q_N^{U,s}(\mathbf{T})$ (see Appendix B) resulted in test that has similar finite sample behavior to the $Q_N^U(\mathbf{T})$ test and is even slightly more conservative than this test (see Section 4). So we do not consider it here. However, the permutation test based on $Q_N^{U,s}(\mathbf{T})$ behaves very well (better than the $Q_N^U(\mathbf{T})$ and $Q_N^{U,\pi}(\mathbf{T})$ tests) for small sample sizes. The permutation version of $Q_N^{U,s}(\mathbf{T})$ is denoted by $Q_N^{U,s,\pi}(\mathbf{T})$. Theoretically, it is also better than the $Q_N^{U,\pi}(\mathbf{T})$ test in the sense of the following theorem (compare with Corollary 2.1, Remark 2.1 and Theorem B.2 in the Appendix B). This result was obtained similarly to Theorem 2.4.

Theorem 3.4. *Under assumptions (1.2)–(1.3):*

1. *Under the null hypothesis $H_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$, if $r = \text{rank}(\mathbf{T})$, then*

$$Q_N^{U,s}(\mathbf{T}) \xrightarrow{d} \frac{\sum_{i=1}^r \lambda_{U,i} Z_{U,i}^2 - \text{trace}(\mathbf{T}\mathbf{V}\mathbf{T})}{\sqrt{2\text{trace}([\mathbf{T}\mathbf{V}\mathbf{T}]^2)}}$$

as $N \rightarrow \infty$, where $\lambda_{U,1}, \dots, \lambda_{U,r}$, \mathbf{V} and $Z_{U,1}, \dots, Z_{U,r}$ are as in Theorem 3.1.

2. *If $r = \text{rank}(\mathbf{T})$, then the permutation distribution of $Q_N^{U,s,\pi}(\mathbf{T})$ conditioned on the observed data \mathbf{X} weakly converges to*

$$\frac{\sum_{i=1}^r \lambda_{U,\pi,i}^* Z_{U,\pi,i}^2 - \text{trace}(\mathbf{T}\mathbf{D}\mathbf{T})}{\sqrt{2\text{trace}([\mathbf{T}\mathbf{D}\mathbf{T}]^2)}}$$

as $N \rightarrow \infty$ in probability, where $\lambda_{U,\pi,1}^, \dots, \lambda_{U,\pi,r}^*$ are the nonzero eigenvalues of the matrix $\mathbf{T}\mathbf{D}\mathbf{T}$, and \mathbf{D} and $Z_{U,\pi,1}, \dots, Z_{U,\pi,r}$ are as in Theorem 3.3.*

3. *If $\sigma_1^2 = \dots = \sigma_d^2$, then the unconditional null distribution of $Q_N^{U,s}(\mathbf{T})$ is the same as the permutation distribution of $Q_N^{U,s,\pi}(\mathbf{T})$ conditioned on the observed data \mathbf{X} , as $N \rightarrow \infty$.*

Almost all of the new solutions considered have very similar asymptotic properties at least under homoscedastic designs. However as we will see in the next section they behave very differently for finite sample.

4. Simulation experiments

In this section, a simulation study is carried out in order to evaluate the finite sample performance of the testing procedures proposed in Sections 2 and 3, i.e., an idea of the size control and power of these tests is provided. The new methods are compared with the WTS and WTPS tests by Pauly et al. [20]. The simulation experiments as well as the illustrative examples of Section 5 were performed in the R programming language [22]. For the asymptotic and permutation tests proposed by Pauly et al. [20], we used the functions implemented in the R package GFD [8].

4.1. Simulation design

Similarly to Pauly et al. [20] and Smaga [23], we restrict our simulation studies to the one-way layout. We consider factor A with $a = 8$ levels. The data were generated from (1.1), i.e., $X_{ij} = \mu_i + \varepsilon_{ij}$, where $i = 1, \dots, 8, j = 1, \dots, n_i$, $\varepsilon_i = \sigma_i(Y_{ij} - E(Y_{i1}))/\sqrt{\text{Var}(Y_{i1})}$, $\sigma_i > 0$ and the distributions of random variables Y_{ij} are taken from normal, Laplace, χ_{10}^2 , log-normal, different symmetric ($Y_{ij}, i = 1, \dots, 4$ are normally distributed, and $Y_{ij}, i = 5, \dots, 8$ are of Laplace distribution) and different skewed ($Y_{ij}, i = 1, \dots, 4$ are of χ_3^2 -distribution, and $Y_{ij}, i = 5, \dots, 8$ are of log-normal one) distributions. To investigate balanced, unbalanced, homoscedastic and heteroscedastic settings, we chose the following vectors of sample sizes $\mathbf{n} = (n_1, \dots, n_8)'$ and standard deviations $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_8)'$: $\mathbf{n}_l + m\mathbf{1}_8$, $l = 1, 2$, $\mathbf{n}_1 = 5\mathbf{1}_8$, $\mathbf{n}_2 = (5, 6, 8, 10, 12, 14, 16, 17)'$, $m = 5, 10, 15, 20, 25$, $\boldsymbol{\sigma}_1 = \mathbf{1}_8$, $\boldsymbol{\sigma}_2 = (1, 1.1, 1.3, 1.5, 1.7, 1.9, 2.1, 2.2)'$ and $\boldsymbol{\sigma}_3 = (2.2, 2.1, 1.9, 1.7, 1.5, 1.3, 1.1, 1)'$. These settings contain the so-called positive (increasing sample sizes are combined with increasing variances) and negative (increasing sample sizes are combined with decreasing variances) pairings (see, for example, [20]). The behavior of the procedures under these two settings is a major assessment criterion for the accuracy of them.

We consider $\boldsymbol{\mu} = \mathbf{0}_8$ (resp. the alternatives $\boldsymbol{\mu}_1 = (1, \mathbf{0}'_7)'$, $\boldsymbol{\mu}_2 = (1, 1, \mathbf{0}'_6)'$, $\boldsymbol{\mu}_3 = (1, 1, 1, \mathbf{0}'_5)'$ and $\mathbf{n} = \mathbf{n}_l + 10\mathbf{1}_8$, $l = 1, 2$) to investigate the type I error level (resp. power) of the tests. To estimate empirical sizes and powers, we used simulation consisting of 1000 runs. The p -values for the permutation testing procedures were estimated from $n_{perm} = 10000$ replications. We also used the significance level $\alpha = 0.05$ for simplicity.

For the nominal level $\alpha = 5\%$, the binomial proportion confidence interval implies the empirical size over the 1000 independent replications should belong to the interval [3.6%, 6.4%] (resp. [3.2%, 6.8%]) with probability 95% (resp. 99%) (see [7]). Thus the reported empirical sizes are displayed in bold (resp.

underlined) when they are outside the 95% (resp. 99%) significance limits (see Tables 3–7 and 10 in the Appendix D).

4.2. Simulation results

Now, we discuss the behavior of the empirical sizes and powers of the tests under consideration. The following observations are confirmed by the results of multiple comparisons of tests using the Nemenyi post hoc test [5, 17], which are available from the author (see [10], for a similar statistical comparison of tests). For readability, in the remainder of the paper, we omit the indication of the test statistic's dependency on the projection matrix \mathbf{T} , e.g., we write Q_N instead of $Q_N(\mathbf{T})$.

4.2.1. Size control

Figure 1 and Tables 3–5 in the Appendix D depict the empirical sizes of the testing procedures considered in Sections 2 and 3. We immediately observe that the asymptotic tests based on Q_N and Q_N^D do not keep the preassigned type I error in almost all settings except the Laplace model with $\mathbf{n} = \mathbf{n}_l + 25\mathbf{1}_8$, $l = 1, 2$. They are usually more or less liberal even with larger sample sizes than considered ones. Therefore, these methods can not be recommended either. On the other hand, the asymptotic unscaled Wald-type test usually demonstrates quite accurate control of the nominal type I error level under symmetric and moderately skewed (χ_{10}^2 -distribution in our simulations) distributions. However, for extremely skewed (the log-normal and χ_3^2 -log-normal models) and also sometimes for Laplace distributions, this method seems to have conservative character. When the sample sizes or standard deviations are equal, the unscaled permutation test based on $Q_N^{U,\pi}$ works quite well and its behavior seems to be only a little worse than that of the best test among the other permutation tests. Nevertheless, this testing procedure is conservative in the case of positive pairing, and it does not maintain the nominal type I error level in the case of negative one. Our simulations confirm the results of Pauly et al. [20] and Smaga [23] about the WTPS test, which seems to be an adequate testing procedure except for the extremely skewed distribution with unequal variances, where it tends to highly over-reject the null hypothesis. In all investigated situations, the tests based on $Q_N^{U,s,\pi}$, $Q_N^{D,\pi}$ and $Q_N^{D,s,\pi}$ demonstrate the most accurate control of the preassigned type I error level.

Summarising the above results, we conclude that: The Q_N and Q_N^D tests are much more liberal than the other testing procedures in almost all scenarios. Under the null, the test based on Q_N^U seems to perform well similarly as the permutation tests under normal and χ_{10}^2 -distributions. However, the asymptotic unscaled Wald-type testing procedure is conservative in the other settings. The permutation tests except eventually the test based on $Q_N^{U,\pi}$ behave quite similarly under symmetric and χ_{10}^2 -distributions. It is also true under extremely skewed distributions when the standard deviations are equal. However, under extremely skewed distributions and unequal variances, the WTPS test and pos-

sibly the test based on $Q_N^{U,\pi}$ do not keep the nominal type I error level in contrast to the other permutation testing procedures.

Since the asymptotic unconditional null and conditional permutation distributions of the diagonal and unscaled test statistics are not in general the same for heteroscedastic designs (see Sections 2 and 3), we additionally investigated the behavior of the tests under the null when the differences between variances are greater. For this purpose, we conducted the simulations in the case of negative pairing with the following vectors of standard deviations: $\sigma_4 = (4.4, 4.2, 3.8, 3.5, 3.2, 2.9, 2.6, 2.4)'$ and $\sigma_5 = (5.0, 4.5, 4.0, 3.5, 2.0, 1.7, 1.3, 1.0)'$. The resulting empirical sizes are given in Tables 6 and 7 in the Appendix D. For $\sigma = \sigma_4$, the conclusions are similar to those given when $\sigma = \sigma_3$. The WTPS (except under extremely skewed distributions), $Q_N^{D,\pi}$, $Q_N^{D,s,\pi}$ and $Q_N^{U,s,\pi}$ tests behave very satisfactorily. For $\sigma = \sigma_5$, the situation changes a little bit. The tests based on Q_N^π and $Q_N^{D,\pi}$ have a more or less liberal character in almost all scenarios. The $Q_N^{D,s,\pi}$, $Q_N^{U,s,\pi}$ and Q_N^U tests seem to work best. However, under log-normal distribution, the first two tests are too liberal and the last one is conservative. Nevertheless, the $Q_N^{D,s,\pi}$ and $Q_N^{U,s,\pi}$ tests decrease their liberality with increasing sample sizes much faster than the WTPS and diagonal permutation tests. Therefore, under very different variances, we may observe the advantage for the $Q_N^{D,s,\pi}$ and $Q_N^{U,s,\pi}$ tests over the other permutation ones.

4.2.2. Power

In Figure 2 and Tables 8–9 in the Appendix D, we present the empirical powers of the tests under consideration. Since the Q_N and Q_N^D tests (resp. the test based on Q_N^π and $Q_N^{U,\pi}$) in all settings (resp. in some cases discussed in the paper) are too liberal, their empirical powers are not really comparable (resp. in those cases). However, they are included for illustration and completeness. In fact, the Q_N and Q_N^D tests appear to have the best power in most cases owing to their extremely liberal behavior.

Under homoscedastic settings, the empirical powers of the asymptotic unscaled Wald-type test are comparable with those of the permutation tests for symmetric and χ_{10}^2 -distributions, while for the other ones they are a few percent smaller. Moreover, the empirical powers of the permutation testing procedures are very similar. However, the $Q_N^{U,\pi}$ and $Q_N^{U,s,\pi}$ tests may be more powerful under symmetric and χ_{10}^2 -distributions, while the opposite is true under skewed ones.

Under heteroscedastic designs and symmetric and χ_{10}^2 -distributions, the Q_N^U and $Q_N^{U,s,\pi}$ tests have quasi identical power. This power is much less than that of the other permutation tests except the $Q_N^{U,\pi}$ test for $\sigma = \sigma_2$. However, the situation may change a little for $\sigma = \sigma_3$ and some alternatives.

Under extremely skewed distributions and heteroscedastic settings, the $Q_N^{U,s,\pi}$ test is evidently more powerful than the unscaled Wald-type test. However, it has still less power than the permutation tests based on Q_N^π , $Q_N^{D,\pi}$ and $Q_N^{D,s,\pi}$ for $\sigma = \sigma_2$. The empirical powers of the unscaled Wald-type permutation test are very similar to these of the $Q_N^{U,s,\pi}$ test for $\sigma = \sigma_2$ and $\mathbf{n}_l = \mathbf{n}_1$. Moreover,

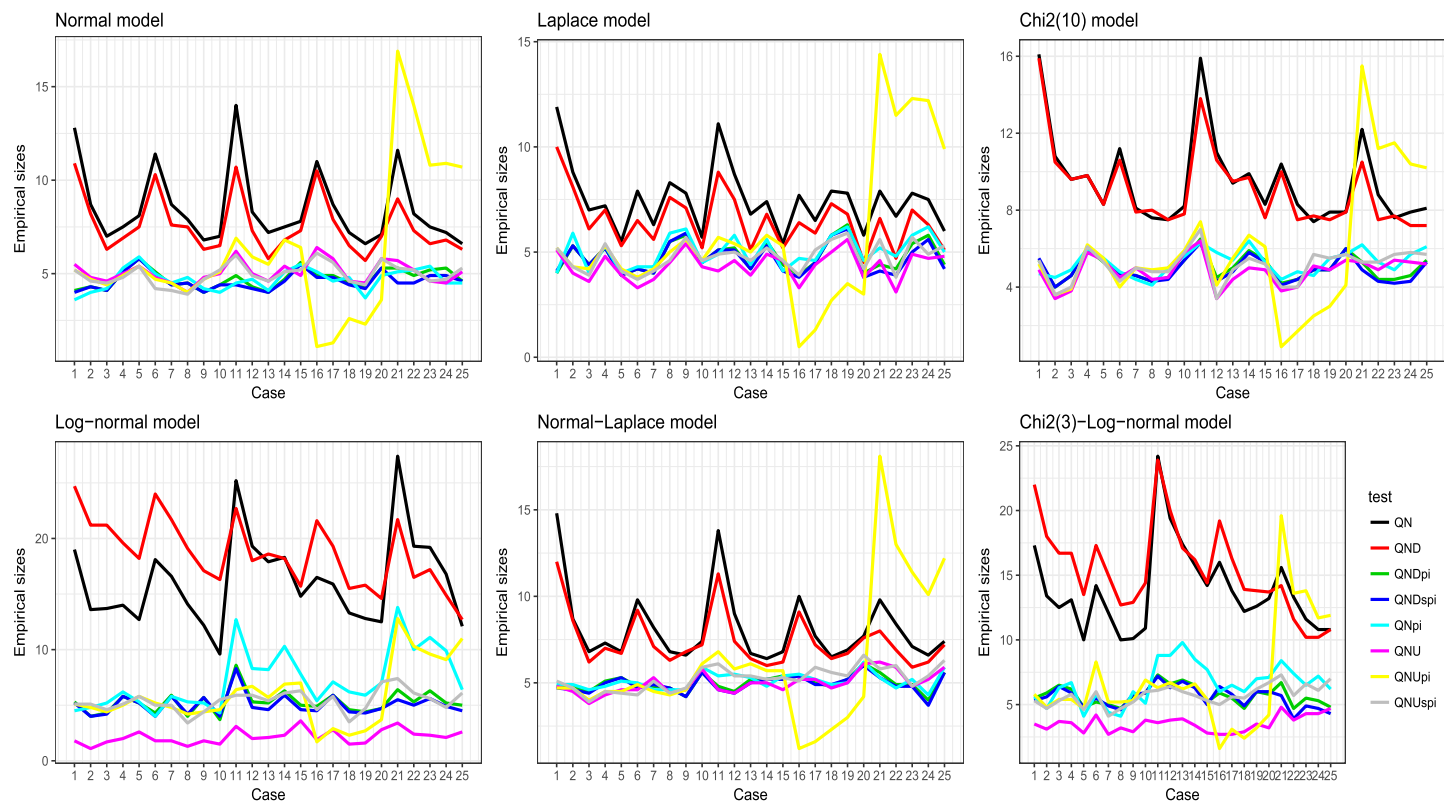


FIG 1. Empirical sizes (as percentages) of the tests obtained for cases of (σ, \mathbf{n}, m) : 1. $(\sigma_1, \mathbf{n}_1, 5)$, 2. $(\sigma_1, \mathbf{n}_1, 10)$, 3. $(\sigma_1, \mathbf{n}_1, 15)$, 4. $(\sigma_1, \mathbf{n}_1, 20)$, 5. $(\sigma_1, \mathbf{n}_1, 25)$, 6. $(\sigma_1, \mathbf{n}_2, 5)$, 7. $(\sigma_1, \mathbf{n}_2, 10)$, 8. $(\sigma_1, \mathbf{n}_2, 15)$, 9. $(\sigma_1, \mathbf{n}_2, 20)$, 10. $(\sigma_1, \mathbf{n}_2, 25)$, 11. $(\sigma_2, \mathbf{n}_1, 5)$, 12. $(\sigma_2, \mathbf{n}_1, 10)$, 13. $(\sigma_2, \mathbf{n}_1, 15)$, 14. $(\sigma_2, \mathbf{n}_1, 20)$, 15. $(\sigma_2, \mathbf{n}_1, 25)$, 16. $(\sigma_2, \mathbf{n}_2, 5)$, 17. $(\sigma_2, \mathbf{n}_2, 10)$, 18. $(\sigma_2, \mathbf{n}_2, 15)$, 19. $(\sigma_2, \mathbf{n}_2, 20)$, 20. $(\sigma_2, \mathbf{n}_2, 25)$, 21. $(\sigma_3, \mathbf{n}_2, 5)$, 22. $(\sigma_3, \mathbf{n}_2, 10)$, 23. $(\sigma_3, \mathbf{n}_2, 15)$, 24. $(\sigma_3, \mathbf{n}_2, 20)$, 25. $(\sigma_3, \mathbf{n}_2, 25)$.

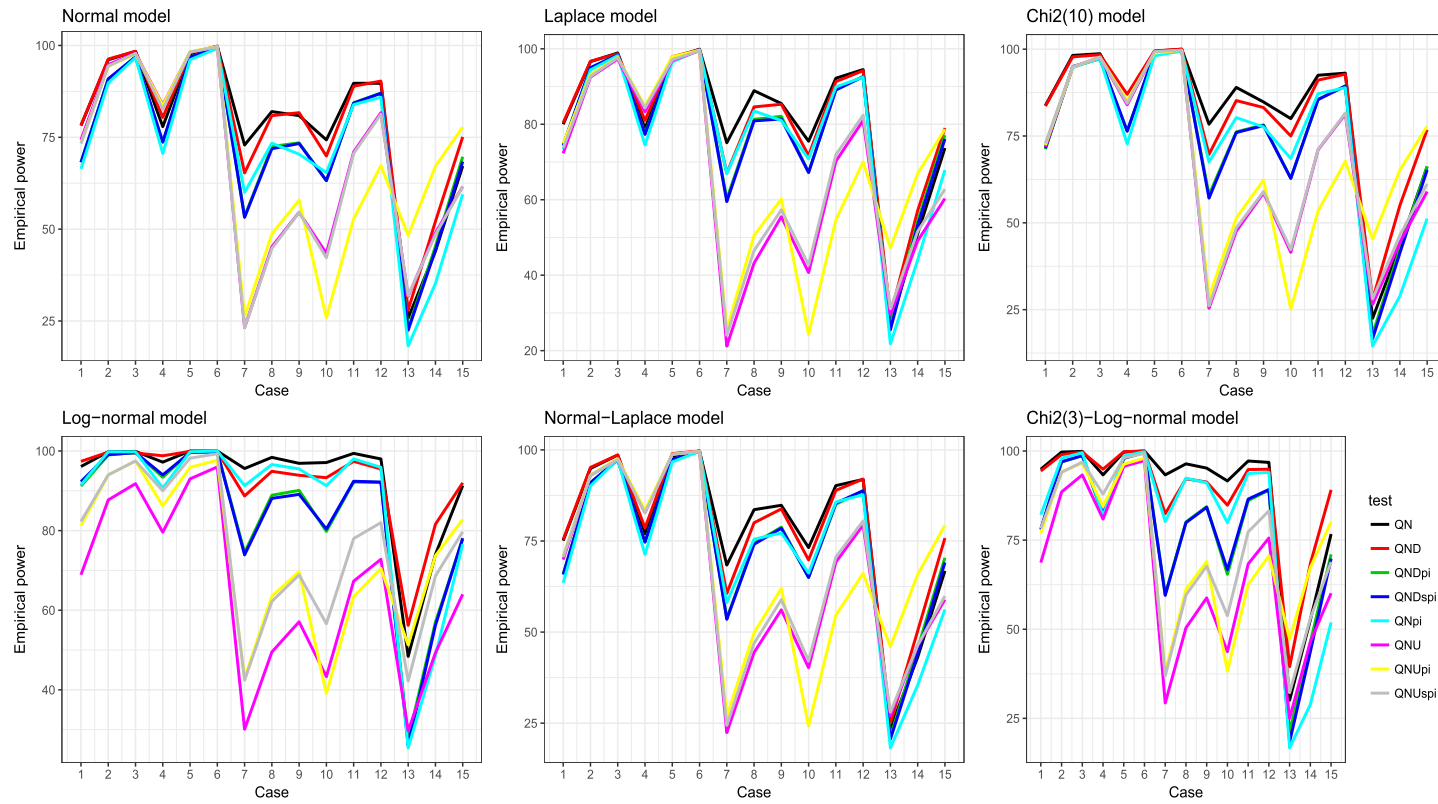


FIG 2. Empirical powers (as percentages) of the tests obtained for cases of $(\sigma, \mathbf{n}, \mu)$: 1. $(\sigma_1, \mathbf{n}_1, \mu_1)$, 2. $(\sigma_1, \mathbf{n}_1, \mu_2)$, 3. $(\sigma_1, \mathbf{n}_1, \mu_3)$, 4. $(\sigma_1, \mathbf{n}_2, \mu_1)$, 5. $(\sigma_1, \mathbf{n}_2, \mu_2)$, 6. $(\sigma_1, \mathbf{n}_2, \mu_3)$, 7. $(\sigma_2, \mathbf{n}_1, \mu_1)$, 8. $(\sigma_2, \mathbf{n}_1, \mu_2)$, 9. $(\sigma_2, \mathbf{n}_1, \mu_3)$, 10. $(\sigma_2, \mathbf{n}_2, \mu_1)$, 11. $(\sigma_2, \mathbf{n}_2, \mu_2)$, 12. $(\sigma_2, \mathbf{n}_2, \mu_3)$, 13. $(\sigma_3, \mathbf{n}_2, \mu_1)$, 14. $(\sigma_3, \mathbf{n}_2, \mu_2)$, 15. $(\sigma_3, \mathbf{n}_2, \mu_3)$ ($\mathbf{n} = \mathbf{n}_l + 101\mathbf{1}$, $l = 1, 2$, $\mu_1 = (1, \mathbf{0}'_7)'$, $\mu_2 = (1, 1, \mathbf{0}'_6)'$, $\mu_3 = (1, 1, 1, \mathbf{0}'_5)'$).

TABLE 1

Recommended test statistics for different scenarios and small sample sizes. Star * (resp. LP) denotes that given test controls the nominal type I error (resp. the lower power of the test was observed).

Distribution	Setting		Test					
			Q_N^U	Q_N^π	$Q_N^{D,\pi}$	$Q_N^{D,s,\pi}$	$Q_N^{U,\pi}$	$Q_N^{U,s,\pi}$
symmetric	homoscedastic	balanced	*	*	*	*	*	*
		unbalanced	*	*	*	*	*	*
	heteroscedastic	balanced	* LP	*	*	*	* LP	* LP
		unbalanced	* LP	*	*	*	*	* LP
modetately skewed	homoscedastic	balanced	*	*	*	*	*	*
		unbalanced	*	*	*	*	*	*
	heteroscedastic	balanced	* LP	*	*	*	* LP	* LP
		unbalanced	* LP	*	*	*	*	* LP
extremely skewed	homoscedastic	balanced		*	*	*	*	*
		unbalanced		*	*	*	*	*
	heteroscedastic	balanced			*	*	* LP	* LP
		unbalanced			*	*	*	* LP

they also emphasize the conservativity and liberality of the $Q_N^{U,\pi}$ test under positive and negative pairing, respectively.

For $\sigma = \sigma_2$ and symmetric and χ_{10}^2 -distributions, the Q_N^π , $Q_N^{D,\pi}$ and $Q_N^{D,s,\pi}$ tests have similar power, possibly except for $\mu = \mu_1$, where the WTPS test may be a little more powerful. Since the Q_N^π test tends to highly over-reject the null hypothesis in the extremely skewed distributions, its empirical powers are greater than those of the diagonal Wald-type permutation tests for $\sigma = \sigma_2$. Interestingly, under negative pairing, the $Q_N^{D,\pi}$ and $Q_N^{D,s,\pi}$ tests are more powerful than the WTPS test (even for extremely skewed distributions where the WTPS test is too liberal). It seems that there are no significant differences in the empirical powers of the tests based on $Q_N^{D,\pi}$ and $Q_N^{D,s,\pi}$ in all scenarios.

4.2.3. Summary

Summarizing, it seems that the diagonal Wald-type permutation testing procedures perform best in size control and power for small sample sizes ($m \leq 10$). The WTPS test and the test based on $Q_N^{U,s,\pi}$ also work well in many scenarios, but are too liberal or have less power in some cases. For the convenience of the Reader, we provide Table 1 to indicate the recommended test statistics for different scenarios and small sample sizes based on our simulation studies. The Q_N and Q_N^D tests are not considered, since they do not keep the type I error for small sample sizes. For large sample sizes, all testing procedures seem to work equally well, but the permutation tests may be time-consuming.

In simulations of this section, we considered only $a = 8$. The comparison of the tests under different numbers of levels and observations is presented in the Appendix C, as a supplement to the above considerations.

5. Real data illustrative examples

In this section, we express three known experimental designs in terms of the general framework of factorial designs presented in Section 1 (see also Section 4 in [20]). For each design, an application of the tests to certain real data examples is also given.

5.1. One-way layout

In the one-way layout where factor A has a levels, $\mathbf{X} = (X_{11}, \dots, X_{an_a})'$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_a)'$, we are interested in testing the hypothesis of no treatment effect $H_0 : \mu_1 = \dots = \mu_a$. This hypothesis is equivalent to $H_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$, where $\mathbf{T} = \mathbf{P}_a = \mathbf{I}_a - (1/a)\mathbf{1}_a\mathbf{1}'_a$. As an example we consider the *startup* data from the R package GFD [8]. This dataset contains the startup costs (in thousands of dollars) of five companies. The company is treated as a factor with five levels ($a = 5$): bakery, gifts, pets, pizza and shoes ($n_1 = 11, n_2 = 10, n_3 = 16, n_4 = 13, n_5 = 10$). We would like to check statistically whether the type of the company has an effect on the startup costs. To solve this problem, we applied all of the tests considered in Section 4, obtaining the results given in Table 2. For $\alpha = 5\%$, all of the tests reject the null hypothesis that the startup costs under the five types of companies do not differ significantly. However, the p -values differ from one another. These of the Q_N and Q_N^D tests are the smallest, while these of the unscaled Wald-type permutation tests are the largest (compare with the results of Section 4).

5.2. Two-way cross-classification design

In the cross-classification with two factors A (with a levels) and B (with b levels), we have $\mathbf{X} = (X_{111}, \dots, X_{abn_{ab}})'$ and $\boldsymbol{\mu} = (\mu_{11}, \dots, \mu_{ab})'$. We consider the following hypotheses: of no main effect A , i.e., $H_0^A : \bar{\mu}_{1.} = \dots = \bar{\mu}_{a.}$; of no main effect B , i.e., $H_0^B : \bar{\mu}_{.1} = \dots = \bar{\mu}_{.b}$; and of no interaction between A and B . To express these hypotheses in the model of Section 1, we use the following contrast matrices $\mathbf{P}_a \otimes (1/b)\mathbf{1}'_b$, $(1/a)\mathbf{1}'_a \otimes \mathbf{P}_b$ and $\mathbf{P}_a \otimes \mathbf{P}_b$, respectively, where \otimes denotes the Kronecker product of matrices. For illustrative purposes, we use the *batteries* data ([16], Table 5-1 p. 176). In this dataset, the life (in hours) of batteries is compared by three material types and three operating temperatures (low - 15°F, medium - 70°F, high - 125°F). From each material type, twelve batteries are randomly selected. Then, they are randomly allocated to each temperature level. The research questions concern possible difference in mean life of the batteries for differing material types and operating temperature levels. To answer these questions, the new tests and those from Pauly et al. [20] were used (p -values are given in Table 2). All testing procedures, except possibly the $Q_N^{U,\pi}$ test, reject the null hypotheses, and hence we conclude that there are significant differences in mean battery life for the three material types and temperature levels, and there is interaction between them. As we observed

TABLE 2
P-values (as percentages) of the tests for the startup, batteries and curdies data.

Data set	H_0	Test							
		Q_N	Q_N^D	Q_N^U	Q_N^π	$Q_N^{D,\pi}$	$Q_N^{D,s,\pi}$	$Q_N^{U,\pi}$	$Q_N^{U,s,\pi}$
startup	$\mu_1 = \dots = \mu_5$	0.46	0.70	4.00	2.83	2.96	2.94	4.63	4.91
batteries	$\bar{\mu}_1 = \bar{\mu}_2 = \bar{\mu}_3$	0.14	0.08	0.06	1.04	0.46	0.46	8.52	0.24
	$\bar{\mu}_{\cdot 1} = \bar{\mu}_{\cdot 2} = \bar{\mu}_{\cdot 3}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	interaction	0.26	0.61	1.19	4.91	2.96	3.00	38.08	2.39
curdies	$\bar{\mu}_1 = \bar{\mu}_2$	2.00	2.00	2.00	0.00	0.00	0.00	0.00	0.00
	$\mu_{i1} = \mu_{i2} = \mu_{i3}$	26.73	40.92	24.53	33.12	51.82	51.98	27.30	19.62
	$i = 1, 2$								

in Section 4, the unscaled Wald-type permutation test is conservative in some situations, what is evident here.

5.3. Nested hierarchical design

We can also consider nested hierarchical designs. As an example, we present that with two fixed factors. Suppose that the factor A (categories) has a levels and the factor B (subcategories) has b_i levels within level i of factor A . The vector of observations is $\mathbf{X} = (X_{111}, \dots, X_{1b_1 n_{1b_1}}, \dots, X_{ab_a 1}, \dots, X_{ab_a n_{ab_a}})'$ and $\boldsymbol{\mu} = (\mu_{11}, \dots, \mu_{1b_1}, \dots, \mu_{a1}, \dots, \mu_{ab_a})'$ is the vector of expected values. The hypothesis of no category effect A , i.e., $H_0^A : \bar{\mu}_1 = \dots = \bar{\mu}_a$ and that of no subcategory effect $B(A)$, i.e., $H_0^{B(A)} : \mu_{i1} = \dots = \mu_{ib_i}, i = 1, \dots, a$, are equivalently written as $H_0^A : \mathbf{P}_a \mathbf{Q} \boldsymbol{\mu} = \mathbf{0}$ and $H_0^{B(A)} : \mathbf{B} \boldsymbol{\mu} = \mathbf{0}$, respectively, where $\mathbf{Q} = \text{diag}((1/b_1)\mathbf{1}'_{b_1}, \dots, (1/b_a)\mathbf{1}'_{b_a})$ and $\mathbf{B} = \text{diag}(\mathbf{P}_{b_1}, \dots, \mathbf{P}_{b_a})$. Here, we consider the *curdies* data containing the number of flatworms (*dugesia*) sampled in two seasons at different sites in the Curdies River in Western Victoria. This dataset is available in the R package `GFD` [8]. Season is a factor A with levels “summer” and “winter”, while site is a factor B with levels 1 to 6, nested under A . The total number of observations is 36. To test the hypotheses H_0^A and $H_0^{B(A)}$ for this data, we used the testing procedures under consideration. Table 2 contains the obtained p -values. For H_0^A , the p -values of all tests are less than the significance level $\alpha = 5\%$, and hence we reject this null hypothesis. On the other hand, the opposite situation is true for $H_0^{B(A)}$. So, site of the river does not have a significant effect on the number of flatworms.

6. Concluding remarks

Compared to other test statistics, the Wald-type one has the advantage that it is applicable in general factorial designs without assuming homoscedasticity or a particular error distribution. We have proposed the asymptotic and permutation tests based on the modified Wald-type statistics where a weight matrix is a certain diagonal matrix, i.e., diagonal and unscaled Wald-type tests. The new methods do not perform equally well when small sample sizes are apparent.

By extensive simulation studies, we conclude that the diagonal and standardized unscaled Wald-type permutation testing procedures perform best. These tests are comparable with or superior to the WTPS test of Pauly et al. [20] under finite samples. Interestingly, the best of our new solutions perform even better than the WTPS test under small sample sizes and heteroscedastic designs, where the new permutation methods are not in general asymptotically valid. So, the resampling procedures may perform well for finite sample without property of asymptotic validity. Except possibly the unscaled Wald-type permutation test, the other new permutation methods are asymptotically valid under homoscedastic designs, which means that they maintain the same asymptotic properties as the corresponding tests based on asymptotic null distribution of their test statistics.

In the case of very different variances in samples, the new tests and the WTPS test of Pauly et al. [20] may fail to keep the prescribed type I error at least under extremely skewed distributions and small sample sizes. When the number of observations is small, the sample variances are perhaps not reliable estimators, which has a negative effect on the behavior of the tests considered. Therefore, an improved variance estimation may result in better performance of the testing procedures [6, 28]. This seems to be an interesting direction of the future research.

Appendix A: Proofs

Proof of Theorem 2.1. From Pauly et al. [20], we obtain

$$\sqrt{N}\mathbf{TX} \xrightarrow{d} \mathbf{Y} \sim \mathcal{N}_d(\mathbf{0}_d, \mathbf{TVT}) \tag{A.1}$$

as $N \rightarrow \infty$, under the null. The consistency of the estimators $\hat{\sigma}_i^2, i = 1, \dots, d$ and the assumption (1.3) imply $\text{diag}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}) \xrightarrow{P} \text{diag}(\mathbf{TVT})$ as $N \rightarrow \infty$. As we noticed in Section 2, the diagonal elements of $\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}$ are positive with probability one. By (1.2), the diagonal elements of \mathbf{TVT} are also positive. Hence, $\{\text{diag}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T})\}^{-1} \xrightarrow{P} \{\text{diag}(\mathbf{TVT})\}^{-1}$ as $N \rightarrow \infty$. Therefore, Slutsky’s theorem and the continuous mapping theorem yield $Q_N^D(\mathbf{T}) \xrightarrow{d} \mathbf{Y}'\{\text{diag}(\mathbf{TVT})\}^{-1}\mathbf{Y}$ as $N \rightarrow \infty$. From the representation theorem of the quadratic forms in normal variables, it follows that $\mathbf{Y}'\{\text{diag}(\mathbf{TVT})\}^{-1}\mathbf{Y} = \sum_{i=1}^r \lambda_{D,i}Z_{D,i}^2$, where $\lambda_{D,1}, \dots, \lambda_{D,r}$ are the nonzero eigenvalues of $\mathbf{V}^{1/2}\mathbf{T}\{\text{diag}(\mathbf{TVT})\}^{-1}\mathbf{TV}^{1/2}$, and $Z_{D,1}, \dots, Z_{D,r}$ are the independent standard normal variables (see, for example, [15], p. 90). Theorem 21.10.1 in [11] implies $\lambda_{D,1}, \dots, \lambda_{D,r}$ are also the nonzero eigenvalues of $\{\text{diag}(\mathbf{TVT})\}^{-1}\mathbf{TVT}$, which completes the proof. \square

Proof of Lemma 2.1. From the proof of Theorem 2.1, we obtain

$$\sum_{i=1}^r \lambda_{D,i}Z_{D,i}^2 = \mathbf{Y}'\{\text{diag}(\mathbf{TVT})\}^{-1}\mathbf{Y},$$

where $\mathbf{Y} \sim \mathcal{N}_d(\mathbf{0}_d, \mathbf{TVT})$. Hence, by theorem on the moments of quadratic forms (see, for instance, [15], p. 55), we conclude that

$$E \left(\sum_{i=1}^r \lambda_{D,i} Z_{D,i}^2 \right) = \text{trace} \left(\{\text{diag}(\mathbf{TVT})\}^{-1} \mathbf{TVT} \right) = d,$$

$$\text{Var} \left(\sum_{i=1}^r \lambda_{D,i} Z_{D,i}^2 \right) = 2 \text{trace} \left(\left[\{\text{diag}(\mathbf{TVT})\}^{-1} \mathbf{TVT} \right]^2 \right). \quad \square$$

Proof of Theorem 2.2. The consistency of the estimators follows immediately from the consistency of $\hat{\mathbf{V}}_N$ for \mathbf{V} . Under $H_1 : \mathbf{T}\boldsymbol{\mu} = \mathbf{a} \neq \mathbf{0}$, we have $\mathbf{TX} \xrightarrow{P} \mathbf{a}$ as $N \rightarrow \infty$. From the proof of Theorem 2.1, it follows that $\{\text{diag}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T})\}^{-1} \xrightarrow{P} \{\text{diag}(\mathbf{TVT})\}^{-1}$ as $N \rightarrow \infty$. Thus, we obtain $\bar{\mathbf{X}}'\mathbf{T}\{\text{diag}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T})\}^{-1}\mathbf{TX} \xrightarrow{P} \mathbf{a}'\{\text{diag}(\mathbf{TVT})\}^{-1}\mathbf{a} > 0$ as $N \rightarrow \infty$. Hence $Q_N^D(\mathbf{T}) \xrightarrow{P} \infty$ as $N \rightarrow \infty$. Since $\hat{k}_{D,\alpha} \xrightarrow{P} k_{D,\alpha}$, the proof is complete. \square

Proof of Theorem 2.3. From Lemma 1 in the supplement to Pauly et al. [20], it follows that conditional convergence in distribution given \mathbf{X} , $\sqrt{N}\mathbf{TX}^\pi \xrightarrow{d} \mathcal{N}_d(\mathbf{0}_d, \sigma^2\mathbf{TDT})$ as $N \rightarrow \infty$, in probability, where σ^2 is given by (3.1). Moreover, Lemma 2 in that supplement shows that $\mathbf{T}\hat{\mathbf{V}}_N^\pi\mathbf{T} \xrightarrow{P} \sigma^2\mathbf{TDT}$ as $N \rightarrow \infty$. Therefore, $\{\text{diag}(\mathbf{T}\hat{\mathbf{V}}_N^\pi\mathbf{T})\}^{-1} \xrightarrow{P} (1/\sigma^2)\{\text{diag}(\mathbf{TDT})\}^{-1}$ as $N \rightarrow \infty$. The rest of the proof runs as in the proof of Theorem 2.1. \square

Proof of Theorem 3.1. Under the null hypothesis, we have $\sqrt{N}\mathbf{TX} \xrightarrow{d} \mathbf{Y} \sim \mathcal{N}_d(\mathbf{0}_d, \mathbf{TVT})$ as $N \rightarrow \infty$ [20]. From the continuous mapping theorem and the representation theorem of the quadratic forms in normal variables, it follows that $Q_N^U(\mathbf{T}) \xrightarrow{d} \mathbf{Y}'\mathbf{Y} = \sum_{i=1}^r \lambda_{U,i} Z_{U,i}^2$ as $N \rightarrow \infty$. \square

Proof of Theorem 3.3. The result may be proved in much the same way as Theorem 3.1 using the fact that conditional convergence in distribution given \mathbf{X} , $\sqrt{N}\mathbf{TX}^\pi \xrightarrow{d} \mathcal{N}_d(\mathbf{0}_d, \sigma^2\mathbf{TDT})$ as $N \rightarrow \infty$, in probability (see the proof of Theorem 2.3). \square

Remark A.1. *It is worth noting that the theoretical results of Sections 2 and 3 (Theorems 2.1–2.4 and 3.1–3.4, Lemma 2.1 and Corollary 2.1) are also valid under the assumptions of finite variances. This is also true for the results of Pauly et al. [20] and Smaga [23]. It follows from that the results, which we mainly use, i.e., (A.1), the consistency of $\hat{\mathbf{V}}_N$ and Lemmas 1 and 2 in the supplement to Pauly et al. [20] (see Remark 8.1 in this supplement), hold under these weaker assumptions than those of existing fourth moments. We present the results under stronger assumptions, since we want to be consistent with the results of Pauly et al. [20] and Smaga [23].*

Appendix B: Asymptotic standardized diagonal and unscaled Wald-type tests

In this section, we briefly present the constructions of the asymptotic tests based on the standardized diagonal and unscaled Wald-type statistics. Their asymptotic properties are also investigated.

Theorem 2.4 shows that under the null hypothesis $H_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$, if $r = \text{rank}(\mathbf{T})$, then

$$Q_N^{D,s}(\mathbf{T}) = \frac{Q_N^D(\mathbf{T}) - d}{\sqrt{2\text{trace}([\{\text{diag}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T})\}^{-1}\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}]^2)}} \xrightarrow{d} \frac{\sum_{i=1}^r \lambda_{D,i} Z_{D,i}^2 - d}{\sqrt{2\text{trace}([\{\text{diag}(\mathbf{T}\mathbf{V}\mathbf{T})\}^{-1}\mathbf{T}\mathbf{V}\mathbf{T}]^2)}}$$

as $N \rightarrow \infty$, where $\lambda_{D,1}, \dots, \lambda_{D,r}$ are the nonzero eigenvalues of the matrix $\{\text{diag}(\mathbf{T}\mathbf{V}\mathbf{T})\}^{-1}\mathbf{T}\mathbf{V}\mathbf{T}$, $\mathbf{V} = \text{diag}(\sigma_1^2/\kappa_1, \dots, \sigma_d^2/\kappa_d)$, and $Z_{D,1}, \dots, Z_{D,r}$ are the independent standard normal variables. By Zhang [30], the asymptotic null distribution of the standardized diagonal Wald-type statistic can be approximated by that of $(\chi_{h_{D,s}}^2 - h_{D,s})/\sqrt{2h_{D,s}}$, where

$$h_{D,s} = \frac{\text{trace}^3([\{\text{diag}(\mathbf{T}\mathbf{V}\mathbf{T})\}^{-1}\mathbf{T}\mathbf{V}\mathbf{T}]^2)}{\text{trace}^2([\{\text{diag}(\mathbf{T}\mathbf{V}\mathbf{T})\}^{-1}\mathbf{T}\mathbf{V}\mathbf{T}]^3)}.$$

The value $h_{D,s}$ is selected such that the first three moments of the variables coincide. The asymptotic critical value is then given by $k_{D,s,\alpha} = (\chi_{h_{D,s},1-\alpha}^2 - h_{D,s})/\sqrt{2h_{D,s}}$. We estimate \mathbf{V} by $\hat{\mathbf{V}}_N$, and hence we have the following simple estimators of $h_{D,s}$ and $k_{D,s,\alpha}$:

$$\hat{h}_{D,s} = \frac{\text{trace}^3([\{\text{diag}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T})\}^{-1}\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}]^2)}{\text{trace}^2([\{\text{diag}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T})\}^{-1}\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}]^3)},$$

$$\hat{k}_{D,s,\alpha} = (\chi_{\hat{h}_{D,s},1-\alpha}^2 - \hat{h}_{D,s})/\sqrt{2\hat{h}_{D,s}}.$$

Theorem B.1. *Under assumptions (1.2)–(1.3), $\hat{h}_{D,s} \xrightarrow{P} h_{D,s}$ and $\hat{k}_{D,s,\alpha} \xrightarrow{P} k_{D,s,\alpha}$ as $N \rightarrow \infty$. Moreover, under the alternative hypothesis $H_1 : \mathbf{T}\boldsymbol{\mu} = \mathbf{a} \neq \mathbf{0}_d$, $P(Q_N^{D,s}(\mathbf{T}) > \hat{k}_{D,s,\alpha} | H_1) \rightarrow 1$ as $N \rightarrow \infty$.*

The proof of this result follows from similar arguments to those in the proof of Theorem 2.2.

By Theorem 3.4, under the null hypothesis $H_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$, if $r = \text{rank}(\mathbf{T})$, then

$$Q_N^{U,s}(\mathbf{T}) = \frac{Q_N^U(\mathbf{T}) - \text{trace}(\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T})}{\sqrt{2\text{trace}([\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}]^2)}} \xrightarrow{d} \frac{\sum_{i=1}^r \lambda_{U,i} Z_{U,i}^2 - \text{trace}(\mathbf{T}\mathbf{V}\mathbf{T})}{\sqrt{2\text{trace}([\mathbf{T}\mathbf{V}\mathbf{T}]^2)}}$$

as $N \rightarrow \infty$, where $\lambda_{U,1}, \dots, \lambda_{U,r}$ are the nonzero eigenvalues of the matrix $\mathbf{T}\mathbf{V}\mathbf{T}$, $\mathbf{V} = \text{diag}(\sigma_1^2/\kappa_1, \dots, \sigma_d^2/\kappa_d)$, and $Z_{U,1}, \dots, Z_{U,r}$ are the independent standard normal variables. The asymptotic null distribution of the standardized unscaled Wald-type statistic can be approximated by that of $(\chi_{h_{U,s}}^2 - h_{U,s})/\sqrt{2h_{U,s}}$, where $h_{U,s} = \text{trace}^3([\mathbf{T}\mathbf{V}\mathbf{T}]^2)/\text{trace}^2([\mathbf{T}\mathbf{V}\mathbf{T}]^3)$ [30]. The asymptotic critical value is

then given by $k_{U,s,\alpha} = (\chi_{\hat{h}_{U,s},1-\alpha}^2 - h_{U,s})/\sqrt{2\hat{h}_{U,s}}$. The estimators of $h_{U,s}$ and $k_{U,s,\alpha}$ are of the form

$$\hat{h}_{U,s} = \frac{\text{trace}^3([\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}]^2)}{\text{trace}^2([\mathbf{T}\hat{\mathbf{V}}_N\mathbf{T}]^3)}, \quad \hat{k}_{U,s,\alpha} = (\chi_{\hat{h}_{U,s},1-\alpha}^2 - \hat{h}_{U,s})/\sqrt{2\hat{h}_{U,s}}.$$

Theorem B.2. *Under assumptions (1.2)–(1.3), $\hat{h}_{U,s} \xrightarrow{P} h_{U,s}$ and $\hat{k}_{U,s,\alpha} \xrightarrow{P} k_{U,s,\alpha}$ as $N \rightarrow \infty$. Moreover, under the alternative hypothesis $H_1 : \mathbf{T}\boldsymbol{\mu} = \mathbf{a} \neq \mathbf{0}_d$, $P(Q_N^{U,s}(\mathbf{T}) > \hat{k}_{U,s,\alpha} | H_1) \rightarrow 1$ as $N \rightarrow \infty$.*

Proof. The consistency of $\hat{\mathbf{V}}_N$ for \mathbf{V} implies the consistency of the estimators $\hat{h}_{U,s}$ and $\hat{k}_{U,s,\alpha}$. Under $H_1 : \mathbf{T}\boldsymbol{\mu} = \mathbf{a} \neq \mathbf{0}$, $\mathbf{T}\bar{\mathbf{X}} \xrightarrow{P} \mathbf{a}$ as $N \rightarrow \infty$. Hence, we have $\bar{\mathbf{X}}'\mathbf{T}\bar{\mathbf{X}} \xrightarrow{P} \mathbf{a}'\mathbf{a} > 0$ as $N \rightarrow \infty$. Since $\hat{\mathbf{V}}_N \xrightarrow{P} \mathbf{V}$, $Q_N^{U,s}(\mathbf{T}) \xrightarrow{P} \infty$ as $N \rightarrow \infty$. By the result that $\hat{k}_{U,s,\alpha} \xrightarrow{P} k_{U,s,\alpha}$, the proof is complete. \square

Appendix C: Additional simulation studies

The simulations presented were performed to analyse how the power of the tests jointly depends on the sample sizes and the number of levels of treatment factor. Simulation design was very similar to that of Subsection 4.1. We considered the normal, χ_{10}^2 and log-normal models, and $\mathbf{n} = (m+5)\mathbf{1}_a$, $\boldsymbol{\sigma} = \mathbf{1}_a$, $m = 5, 10, 15, 20$ and $a = 6, 8, 10, 12$. We chose balanced and homoscedastic settings to fairly show the behavior of the power under increasing number of levels and observations. To investigate the type I error level (resp. power) of the tests, we considered $\boldsymbol{\mu} = \mathbf{0}_a$ (resp. the alternative $\boldsymbol{\mu}_1 = (1, \mathbf{0}'_{a-1})'$).

The resulting empirical sizes and powers are depicted in Tables 10 and 11 in the Appendix D, respectively. For given sample sizes, the Q_N and Q_N^D tests become more and more liberal with growth of the number of levels. The empirical sizes of the other testing procedures are very similar for all numbers of observations and levels. Moreover, they are quite close to significance level, except these for the Q_N^U test in case of the extremely skewed distribution. The empirical powers of the Q_N and Q_N^D tests are usually at the same level for different values of a . However, they slightly increase or decrease in some cases, when the number of levels increases. This can be explained by unacceptable behavior of these tests under the null hypothesis. The empirical powers of the other testing procedures decrease as the number of levels increases, which seems to be natural as most of these tests keep the nominal type I error level. This decrease, however, usually decreases as the number of observations increases. Of course, for given number of levels, the empirical powers of these testing procedures increase quite fast as the number of observations increases.

Appendix D: Simulation results tables

Tables 3–11 contain the results of simulation studies considered in Section 4 and the Appendix C.

TABLE 3
 Empirical sizes (as percentages) of the tests obtained in the normal and Laplace models.

Model	σ	n_l	m	Test									
				Q_N	Q_N^D	Q_N^U	Q_N^π	$Q_N^{D,\pi}$	$Q_N^{D,s,\pi}$	$Q_N^{U,\pi}$	$Q_N^{U,s,\pi}$		
Normal	σ_1	n_1	5	12.8	10.9	5.5	3.6	4.1	4.0	5.2	5.2		
			10	8.7	8.2	4.8	4.0	4.3	4.3	4.7	4.6		
			15	7.0	6.3	4.6	4.2	4.1	4.1	4.4	4.5		
			20	7.5	6.9	5.0	5.3	5.1	5.1	4.9	4.8		
			25	8.1	7.5	5.4	5.9	5.8	5.8	5.4	5.4		
		n_2	5	11.4	10.3	4.8	5.0	5.1	5.0	4.7	4.2		
			10	8.7	7.6	4.5	4.5	4.4	4.4	4.5	4.1		
			15	7.9	7.5	4.0	4.8	4.5	4.5	4.0	3.9		
			20	6.8	6.3	4.8	4.2	4.0	4.0	4.7	4.7		
			25	7.0	6.5	5.0	4.0	4.4	4.4	5.1	5.2		
		σ_2	n_1	5	14.0	10.7	6.2	4.5	4.9	4.4	6.9	6.0	
				10	8.3	7.3	5.0	4.7	4.3	4.2	5.9	4.9	
	15			7.2	5.8	4.6	4.1	4.0	4.0	5.5	4.6		
	20			7.5	6.8	5.4	5.1	4.7	4.6	6.8	5.1		
	25			7.8	7.3	4.9	5.5	5.6	5.5	6.4	5.2		
	n_2		5	11.0	10.5	6.4	5.1	4.9	4.8	1.1	6.1		
			10	8.7	7.9	5.8	4.6	4.9	4.8	1.3	5.6		
			15	7.2	6.5	4.6	4.8	4.5	4.4	2.6	4.6		
			20	6.6	5.7	4.5	3.7	4.3	4.2	2.3	4.4		
			25	7.1	7.0	5.8	4.9	5.3	5.3	3.6	5.8		
			σ_3	n_2	5	11.6	9.0	5.7	5.1	5.3	4.5	16.9	5.3
					10	8.2	7.3	5.2	5.2	4.9	4.5	14.0	5.2
	15	7.5			6.6	4.6	5.4	5.2	4.9	10.8	4.6		
	20	7.2			6.8	4.5	4.5	5.3	4.9	10.9	4.7		
25	6.6	6.3			5.1	4.5	4.6	4.6	10.7	5.3			
Laplace	σ_1	n_1	5	11.9	10.0	5.1	4.0	4.0	4.1	5.2	5.2		
			10	8.8	8.1	4.0	5.9	5.3	5.3	4.3	4.3		
			15	7.0	6.1	3.6	4.0	4.4	4.4	4.2	3.9		
			20	7.2	7.0	4.8	5.3	5.2	5.2	5.3	5.4		
			25	5.5	5.3	3.9	4.0	3.9	3.9	4.2	4.1		
		n_2	5	7.9	6.5	3.3	4.3	4.2	4.2	3.8	3.7		
			10	6.3	5.6	3.7	4.3	4.0	4.0	4.2	4.1		
			15	8.3	7.6	4.5	5.9	5.5	5.5	5.0	4.6		
			20	7.8	7.1	5.4	6.1	5.8	5.9	5.6	5.7		
			25	5.7	5.2	4.3	4.5	4.5	4.5	4.6	4.6		
			σ_2	n_1	5	11.1	8.8	4.1	4.9	5.1	5.1	5.7	4.9
					10	8.7	7.5	4.6	5.8	5.2	5.1	5.4	5.0
					15	6.8	5.1	3.9	4.4	4.2	4.2	5.0	4.6
		20			7.4	6.8	4.9	5.6	5.5	5.3	5.8	5.2	
		25			5.4	5.2	4.6	4.1	4.4	4.1	5.3	4.6	
		n_2		5	7.7	6.4	3.3	4.7	3.9	3.8	0.5	3.9	
				10	6.5	5.9	4.4	4.6	4.5	4.5	1.3	5.1	
				15	7.9	7.3	5.0	5.8	5.8	5.8	2.7	5.6	
	20			7.8	6.8	5.6	6.2	6.3	6.1	3.5	5.9		
	σ_3	n_2	5	7.9	6.6	4.6	5.2	4.4	4.1	14.4	5.6		
			10	6.7	4.7	3.1	4.8	4.2	3.9	11.5	3.8		
			15	7.8	7.0	4.9	5.8	5.4	5.0	12.3	5.7		
			20	7.5	6.3	4.7	6.2	5.8	5.6	12.2	4.9		
			25	6.0	5.1	4.8	5.0	4.4	4.2	9.9	5.3		

TABLE 4
 Empirical sizes (as percentages) of the tests obtained in the χ^2_{10} and log-normal models.

Model	σ	n_i	m	Test								
				Q_N	Q_N^D	Q_N^U	Q_N^π	$Q_N^{D,\pi}$	$Q_N^{D,s,\pi}$	$Q_N^{U,\pi}$	$Q_N^{U,s,\pi}$	
χ^2_{10}	σ_1	n_1	5	16.1	15.9	4.9	4.7	5.4	5.5	5.2	5.4	
			10	10.8	10.5	3.4	4.5	4.0	4.0	3.6	3.6	
			15	9.6	9.6	3.8	4.9	4.6	4.6	3.9	4.0	
			20	9.8	9.8	5.8	6.0	6.0	6.0	6.2	6.1	
			25	8.3	8.3	5.4	5.5	5.3	5.4	5.5	5.3	
		n_2	5	11.2	10.6	4.5	4.8	4.7	4.7	4.0	4.3	
			10	8.1	7.9	5.0	4.4	4.6	4.6	5.0	5.0	
			15	7.6	8.0	4.4	4.1	4.3	4.3	4.9	4.8	
			20	7.5	7.5	4.5	5.0	4.5	4.4	5.0	4.8	
			25	8.2	7.8	5.8	5.6	5.4	5.4	5.9	5.8	
	σ_2	n_1	5	15.9	13.8	6.4	6.3	6.5	6.3	7.4	7.0	
			10	11.0	10.6	3.4	5.8	4.5	4.3	4.1	3.4	
			15	9.4	9.5	4.4	5.4	5.0	4.8	5.5	4.9	
			20	9.9	9.7	5.0	6.4	5.9	5.8	6.7	5.5	
			25	8.3	7.6	4.9	5.3	5.4	5.3	6.1	5.2	
		n_2	5	10.4	10.0	3.8	4.4	4.2	4.1	0.9	4.0	
			10	8.3	7.5	4.0	4.8	4.4	4.4	1.7	4.0	
			15	7.4	7.7	5.1	4.6	4.9	4.9	2.5	5.7	
			20	7.9	7.5	4.9	5.5	4.9	4.9	3.0	5.5	
			25	7.9	7.9	5.4	5.7	6.0	6.0	4.1	5.7	
σ_3	n_2	5	12.2	10.5	5.3	6.2	5.3	4.9	15.5	5.3		
		10	8.8	7.5	4.9	5.3	4.4	4.3	11.2	5.3		
		15	7.6	7.7	5.4	4.9	4.4	4.2	11.5	5.7		
		20	7.9	7.2	5.3	5.7	4.6	4.3	10.4	5.8		
		25	8.1	7.2	5.2	6.1	5.4	5.3	10.2	5.7		
	Log-normal	σ_1	n_1	5	19.0	24.7	1.8	4.5	5.3	5.2	5.1	5.1
				10	13.6	21.2	1.1	4.8	4.0	4.0	4.8	5.1
				15	13.7	21.2	1.7	5.2	4.3	4.2	4.4	4.6
				20	14.0	19.6	2.0	6.2	5.8	5.8	5.0	5.1
				25	12.7	18.2	2.6	5.3	5.2	5.2	5.8	5.8
n_2			5	18.1	24.0	1.8	4.0	4.3	4.0	5.2	5.0	
			10	16.6	21.7	1.8	5.8	5.9	5.8	4.8	5.0	
			15	14.1	19.1	1.3	5.3	4.0	4.2	4.2	3.4	
			20	12.2	17.1	1.8	5.2	5.7	5.7	4.4	4.4	
			25	9.6	16.3	1.5	4.3	3.7	4.0	4.6	5.5	
σ_2		n_1	5	25.2	22.7	3.1	12.7	8.6	8.3	6.4	5.9	
			10	19.3	18.0	2.0	8.3	5.3	4.8	6.7	5.9	
			15	17.9	18.6	2.1	8.2	5.2	4.6	5.7	5.4	
			20	18.3	18.2	2.3	10.3	6.3	5.9	6.9	6.1	
			25	14.8	15.7	3.6	7.8	5.0	4.6	7.0	6.3	
		n_2	5	16.5	21.6	1.9	5.4	4.9	4.5	1.7	4.6	
			10	15.9	19.3	2.8	7.1	5.9	5.9	2.9	5.8	
			15	13.3	15.5	1.5	6.2	4.6	4.4	2.3	3.5	
			20	12.8	15.8	1.6	5.9	4.4	4.3	2.7	4.7	
			25	12.5	14.6	2.8	7.0	4.7	4.7	3.7	7.1	
σ_3	n_2	5	27.4	21.7	3.4	13.8	6.4	5.5	12.8	7.4		
		10	19.3	16.5	2.4	10.0	5.3	5.0	10.3	6.1		
		15	19.2	17.2	2.3	11.1	6.3	5.6	9.6	5.6		
		20	16.8	14.9	2.1	9.9	5.2	4.9	9.1	4.8		
		25	12.1	12.7	2.6	6.4	5.0	4.5	11.0	6.1		

TABLE 5
 Empirical sizes (as percentages) of the tests obtained in the normal-Laplace and χ_3^2 -log-normal models.

Model	σ	n_l	m	Test								
				Q_N	Q_N^D	Q_N^U	Q_N^π	$Q_N^{D,\pi}$	$Q_N^{D,s,\pi}$	$Q_N^{U,\pi}$	$Q_N^{U,s,\pi}$	
Normal-Laplace	σ_1	n_1	5	14.8	12.0	4.8	4.9	4.8	4.7	4.7	5.1	
			10	8.7	8.6	4.5	4.9	4.8	4.6	4.7	4.7	
			15	6.8	6.2	3.8	4.6	4.5	4.4	4.0	3.9	
			20	7.3	7.0	4.3	4.8	5.1	5.0	4.5	4.5	
			25	6.8	6.7	4.6	5.1	5.3	5.3	4.5	4.4	
		n_2	5	9.8	9.2	4.6	5.0	4.8	4.7	4.9	4.3	
			10	8.2	7.1	5.3	4.6	4.8	4.9	4.5	5.1	
			15	6.8	6.3	4.4	4.6	4.7	4.7	4.3	4.4	
			20	6.6	6.8	4.6	4.6	4.2	4.2	4.6	4.7	
			25	7.4	7.2	5.8	5.9	5.6	5.6	6.1	5.9	
	σ_2	n_1	5	13.8	11.3	4.6	5.4	4.8	4.6	6.8	6.1	
			10	9.0	7.4	4.4	5.5	4.5	4.4	5.8	5.4	
			15	6.7	6.4	5.0	5.3	5.2	5.0	6.1	5.4	
			20	6.4	6.0	5.0	4.8	5.2	5.2	5.7	5.2	
			25	6.8	6.2	4.6	5.4	5.4	5.2	5.7	5.3	
		n_2	5	10.0	9.1	5.2	5.5	5.4	5.4	1.2	5.1	
			10	7.7	7.2	5.2	5.2	5.0	4.9	1.6	5.9	
			15	6.5	6.4	4.7	4.9	4.8	4.9	2.3	5.6	
			20	6.9	6.7	5.0	5.0	5.2	5.2	3.0	5.4	
			25	7.7	7.6	6.1	6.2	6.1	6.0	4.2	6.6	
σ_3	n_2	5	9.8	8.0	6.2	5.3	5.6	5.3	18.1	5.8		
		10	8.4	6.9	5.9	4.7	4.9	4.8	13.0	6.0		
		15	7.1	5.9	4.8	5.2	4.9	4.8	11.4	4.8		
		20	6.6	6.2	5.2	4.3	4.0	3.7	10.1	5.4		
		25	7.4	7.2	5.9	5.9	5.6	5.6	12.2	6.3		
	χ_3^2 -Log-normal	σ_1	n_1	5	17.3	22.0	3.5	5.6	5.5	5.3	5.8	5.3
				10	13.4	18.0	3.1	4.7	5.9	5.6	4.8	4.7
				15	12.5	16.7	3.7	6.3	6.5	6.4	5.4	5.3
				20	13.1	16.7	3.6	6.7	6.2	5.9	5.4	5.8
				25	10.0	13.5	2.8	4.1	4.8	4.8	4.7	4.6
n_2			5	14.2	17.3	4.2	5.8	5.2	5.5	8.3	6.0	
			10	12.1	15.0	2.7	4.4	5.0	4.9	5.2	4.1	
			15	10.0	12.7	3.2	4.1	4.7	4.6	5.2	4.7	
			20	10.1	12.9	2.9	6.0	5.7	5.8	5.3	5.2	
			25	10.9	14.4	3.8	5.1	5.9	5.9	6.9	6.0	
σ_2			n_1	5	24.2	23.9	3.6	8.8	7.3	7.2	6.3	6.1
				10	19.4	20.0	3.8	8.8	6.6	6.3	6.7	6.4
				15	17.4	17.1	3.9	9.8	6.9	6.8	6.2	6.0
				20	15.8	16.2	3.4	8.5	6.5	6.3	6.6	5.7
				25	14.2	14.4	2.8	7.7	5.2	5.0	5.4	5.3
		n_2	5	16.0	19.2	2.7	6.0	5.9	6.4	1.6	5.0	
			10	13.8	16.3	2.7	6.5	5.5	5.8	3.1	5.6	
			15	12.2	13.9	2.9	6.0	4.7	4.9	2.4	5.5	
			20	12.6	13.8	3.5	7.0	6.0	6.0	3.2	6.2	
			25	13.2	13.7	3.2	7.1	5.8	6.0	4.2	6.7	
		σ_3	n_2	5	15.6	14.2	4.8	8.4	6.7	5.7	19.6	7.3
				10	13.3	11.6	3.8	7.4	4.7	3.9	13.6	5.7
				15	11.6	10.2	4.3	6.5	5.5	4.9	13.8	6.6
				20	10.8	10.2	4.3	7.2	5.3	4.7	11.7	6.1
				25	10.8	10.8	4.7	6.2	4.8	4.3	11.9	7.0

TABLE 6
 Empirical sizes (as percentages) of the tests obtained for $\sigma = \sigma_4$ and $\mathbf{n}_1 = \mathbf{n}_2$.

Model	m	Test							
		Q_N	Q_N^D	Q_N^U	Q_N^π	$Q_N^{D,\pi}$	$Q_N^{D,s,\pi}$	$Q_N^{U,\pi}$	$Q_N^{U,s,\pi}$
Normal	5	11.5	9.4	5.7	4.8	5.2	4.8	14.2	5.3
	10	8.5	7.9	4.8	4.8	4.4	4.3	11.8	4.8
	15	7.6	7.2	4.1	5.0	5.1	5.0	9.4	4.3
	20	7.0	6.9	4.5	4.4	4.8	4.4	9.6	4.6
	25	6.6	6.5	5.3	4.2	4.7	4.4	9.1	5.1
Laplace	5	7.8	6.9	4.8	5.0	4.3	4.0	12.5	5.6
	10	6.5	5.1	3.4	4.8	4.1	4.0	9.1	3.7
	15	8.0	6.9	4.8	5.8	5.5	5.3	10.2	5.4
	20	7.5	6.6	4.8	6.4	5.9	5.7	10.6	5.0
	25	5.9	4.8	4.7	4.8	4.3	4.3	8.5	5.2
χ_{10}^2	5	11.7	10.6	5.2	5.6	5.3	5.0	12.6	5.1
	10	8.8	7.8	5.0	5.1	4.6	4.5	9.2	5.1
	15	7.3	7.4	5.3	4.7	4.4	4.1	9.6	5.7
	20	7.7	7.1	5.5	5.0	4.6	4.4	8.7	5.7
	25	8.1	7.8	5.5	6.0	5.5	5.4	9.1	5.8
Log-normal	5	24.8	22.2	2.9	10.8	5.2	4.8	10.5	6.9
	10	19.9	20.1	2.6	10.9	6.7	6.7	10.1	6.3
	15	17.7	17.7	1.9	8.7	5.7	5.4	7.9	5.0
	20	15.7	15.7	2.0	7.6	4.6	4.3	7.8	4.5
	25	10.9	13.4	2.3	5.5	4.6	4.2	9.0	6.1
Normal-Laplace	5	10.1	8.8	5.8	5.3	5.1	5.1	13.8	5.6
	10	8.1	6.8	5.8	4.8	4.7	4.6	11.0	5.9
	15	6.8	6.2	4.6	4.9	4.9	4.8	9.3	4.4
	20	6.7	6.4	5.0	4.1	3.9	3.7	8.7	5.1
	25	7.4	7.3	5.9	5.8	5.4	5.3	11.0	6.0
χ_3^2 -Log-normal	5	14.7	14.3	4.6	7.3	6.2	5.7	15.9	7.0
	10	12.7	12.1	3.1	6.4	4.7	4.0	11.3	5.1
	15	10.7	10.6	4.3	5.5	5.0	4.6	11.3	6.1
	20	11.0	10.6	4.0	6.8	5.3	4.9	9.8	5.5
	25	10.4	11.4	4.6	5.8	4.9	4.5	10.8	6.9

TABLE 7
 Empirical sizes (as percentages) of the tests obtained for $\sigma = \sigma_5$ and $\mathbf{n}_1 = \mathbf{n}_2$.

Model	m	Test							
		Q_N	Q_N^D	Q_N^U	Q_N^π	$Q_N^{D,\pi}$	$Q_N^{D,s,\pi}$	$Q_N^{U,\pi}$	$Q_N^{U,s,\pi}$
Normal	5	11.6	7.8	5.4	7.2	7.3	4.8	27.9	6.2
	10	8.7	6.4	4.9	5.9	6.6	4.8	19.8	5.4
	15	8.1	6.1	5.0	6.0	6.5	4.9	17.2	5.2
	20	7.2	6.3	4.9	4.9	6.4	5.0	16.5	5.4
	25	6.7	5.6	5.3	5.2	6.0	5.0	15.1	5.6
Laplace	5	8.0	6.2	4.7	6.0	6.9	4.7	25.5	6.3
	10	7.1	4.5	3.6	6.0	5.7	3.9	17.4	4.5
	15	7.4	6.1	5.0	6.2	7.3	6.0	18.8	6.1
	20	7.4	6.0	5.2	6.8	7.1	5.7	17.9	6.3
	25	6.5	5.6	5.2	5.5	6.6	5.1	14.7	6.0
χ_{10}^2	5	13.5	8.6	6.0	9.3	7.9	6.3	26.3	6.8
	10	9.8	7.5	4.9	6.5	6.7	5.5	18.1	5.9
	15	7.9	7.2	5.2	5.9	6.8	5.1	17.4	6.0
	20	8.2	6.0	5.1	6.1	5.6	4.7	17.1	6.0
	25	8.5	6.8	5.3	6.6	6.4	5.0	14.8	6.1
Log-normal	5	35.8	19.5	5.2	28.5	13.6	11.1	25.4	10.2
	10	27.6	16.4	4.0	22.2	11.3	9.8	20.2	8.5
	15	26.6	14.3	3.6	20.8	9.7	8.2	18.7	8.0
	20	22.2	13.0	3.5	16.5	8.6	7.3	16.5	6.9
	25	18.0	10.2	3.6	11.2	6.2	5.0	16.3	7.2
Normal-Laplace	5	9.8	8.2	7.1	6.3	7.4	5.7	27.1	7.7
	10	7.9	7.1	5.8	5.3	6.8	5.4	19.8	6.3
	15	7.2	5.9	4.8	5.4	6.1	4.6	17.3	5.3
	20	7.2	5.5	5.5	5.2	6.0	4.9	14.9	5.8
	25	7.4	6.4	5.9	6.4	7.4	5.7	17.1	6.6
χ_3^2 -Log-normal	5	18.7	11.0	5.6	15.2	9.8	7.6	29.0	8.1
	10	17.0	9.4	4.0	11.4	7.8	5.3	20.4	6.6
	15	14.5	9.8	5.1	11.0	7.1	5.2	19.6	7.2
	20	13.1	9.0	5.1	9.5	6.8	5.3	18.1	6.7
	25	13.2	8.3	5.0	8.2	6.3	4.4	17.6	7.1

TABLE 8
 Empirical powers (as percentages) of the tests obtained in the normal, Laplace and χ^2_{10} models ($\mathbf{n} = \mathbf{n}_l + 10\mathbf{1}_8$, $l = 1, 2$, $\boldsymbol{\mu}_1 = (1, \mathbf{0}'_7)'$, $\boldsymbol{\mu}_2 = (1, 1, \mathbf{0}'_6)'$, $\boldsymbol{\mu}_3 = (1, 1, 1, \mathbf{0}'_5)'$).

Model	σ	\mathbf{n}_l	$\boldsymbol{\mu}$	Test								
				Q_N	Q_N^D	Q_N^U	Q_N^π	$Q_N^{D,\pi}$	$Q_N^{D,s,\pi}$	$Q_N^{U,\pi}$	$Q_N^{U,s,\pi}$	
Normal	σ_1	\mathbf{n}_1	$\boldsymbol{\mu}_1$	78.2	78.3	74.3	66.4	68.2	68.2	73.4	73.3	
			$\boldsymbol{\mu}_2$	96.1	96.3	94.8	89.8	90.8	90.9	94.3	94.4	
			$\boldsymbol{\mu}_3$	98.4	98.4	97.9	96.7	96.9	96.9	97.7	97.8	
		\mathbf{n}_2	$\boldsymbol{\mu}_1$	77.9	80.4	83.8	70.6	73.7	73.7	83.6	82.6	
			$\boldsymbol{\mu}_2$	97.7	98.0	98.2	96.1	96.8	96.8	98.3	98.2	
			$\boldsymbol{\mu}_3$	99.8	99.8	99.6	99.2	99.5	99.5	99.7	99.6	
		σ_2	\mathbf{n}_1	$\boldsymbol{\mu}_1$	72.9	65.3	23.2	60.0	53.4	53.2	26.1	23.2
				$\boldsymbol{\mu}_2$	82.0	80.9	45.2	73.4	72.5	71.9	48.7	44.7
				$\boldsymbol{\mu}_3$	80.9	81.7	54.6	70.4	73.5	73.3	57.9	54.6
	\mathbf{n}_2		$\boldsymbol{\mu}_1$	74.3	69.9	43.4	65.4	63.2	63.2	25.9	42.2	
			$\boldsymbol{\mu}_2$	89.7	88.9	71.0	83.9	84.4	84.3	52.8	70.6	
			$\boldsymbol{\mu}_3$	89.7	90.3	81.7	85.9	87.0	87.0	67.3	81.3	
	σ_3	\mathbf{n}_2	$\boldsymbol{\mu}_1$	25.8	28.0	31.7	18.2	23.0	22.5	48.3	31.9	
			$\boldsymbol{\mu}_2$	44.0	52.1	49.0	35.2	45.3	44.3	67.1	48.5	
			$\boldsymbol{\mu}_3$	67.2	75.1	61.6	59.5	69.7	68.4	77.8	61.6	
Laplace	σ_1	\mathbf{n}_1	$\boldsymbol{\mu}_1$	80.0	80.4	72.3	73.4	74.3	74.5	74.9	73.7	
			$\boldsymbol{\mu}_2$	96.6	96.7	92.5	94.4	94.9	95.0	93.2	92.7	
			$\boldsymbol{\mu}_3$	98.9	98.7	97.3	98.2	98.4	98.4	97.7	97.6	
		\mathbf{n}_2	$\boldsymbol{\mu}_1$	78.9	81.0	83.3	74.5	77.5	77.3	84.8	84.3	
			$\boldsymbol{\mu}_2$	97.4	97.9	96.9	96.5	97.2	97.2	97.9	97.1	
			$\boldsymbol{\mu}_3$	99.9	99.8	99.5	99.7	99.7	99.7	99.6	99.5	
		σ_2	\mathbf{n}_1	$\boldsymbol{\mu}_1$	75.1	67.0	21.2	66.9	60.2	59.5	25.6	24.1
				$\boldsymbol{\mu}_2$	88.9	84.6	43.2	83.5	81.3	80.9	50.5	46.5
				$\boldsymbol{\mu}_3$	85.5	85.3	55.5	81.0	82.1	81.4	60.2	57.3
	\mathbf{n}_2		$\boldsymbol{\mu}_1$	75.5	71.7	40.7	70.8	67.3	67.2	24.3	42.7	
			$\boldsymbol{\mu}_2$	92.2	91.3	70.2	89.9	89.1	89.1	54.6	71.9	
			$\boldsymbol{\mu}_3$	94.5	94.2	81.0	92.5	92.7	92.6	70.0	82.4	
	σ_3	\mathbf{n}_2	$\boldsymbol{\mu}_1$	26.9	29.4	29.5	21.8	26.6	25.5	47.2	31.1	
			$\boldsymbol{\mu}_2$	51.0	57.1	49.1	44.0	54.3	53.2	66.9	51.6	
			$\boldsymbol{\mu}_3$	73.7	79.0	60.3	67.9	77.1	76.1	78.7	62.8	
χ^2_{10}	σ_1	\mathbf{n}_1	$\boldsymbol{\mu}_1$	83.8	83.6	71.9	72.3	71.2	71.5	72.4	72.9	
			$\boldsymbol{\mu}_2$	98.2	97.8	94.8	94.7	94.9	95.0	94.8	94.9	
			$\boldsymbol{\mu}_3$	98.7	98.4	97.7	97.3	97.3	97.3	97.7	97.7	
		\mathbf{n}_2	$\boldsymbol{\mu}_1$	84.3	87.0	84.0	72.7	76.4	76.4	85.1	84.4	
			$\boldsymbol{\mu}_2$	99.4	99.5	99.1	98.1	98.8	98.8	98.9	99.3	
			$\boldsymbol{\mu}_3$	100	100	99.7	99.5	99.4	99.5	99.7	99.7	
		σ_2	\mathbf{n}_1	$\boldsymbol{\mu}_1$	78.4	69.7	25.4	67.5	58.0	57.1	28.7	26.1
				$\boldsymbol{\mu}_2$	89.0	85.2	47.6	80.3	76.2	76.0	51.4	48.5
				$\boldsymbol{\mu}_3$	84.8	83.2	58.6	77.6	78.2	78.0	62.3	59.0
	\mathbf{n}_2		$\boldsymbol{\mu}_1$	80.0	75.0	41.5	68.5	62.8	62.8	25.2	42.3	
			$\boldsymbol{\mu}_2$	92.5	91.1	71.0	87.2	85.5	85.5	53.4	71.1	
			$\boldsymbol{\mu}_3$	93.1	92.8	81.3	88.9	89.4	89.4	67.7	81.6	
	σ_3	\mathbf{n}_2	$\boldsymbol{\mu}_1$	22.5	27.9	26.5	14.5	17.9	16.6	45.3	28.1	
			$\boldsymbol{\mu}_2$	43.4	55.0	44.4	28.9	42.5	41.3	65.2	46.4	
			$\boldsymbol{\mu}_3$	66.1	76.8	59.0	51.2	66.3	65.2	77.8	61.2	

TABLE 9
 Empirical powers (as percentages) of the tests obtained in the log-normal, normal-Laplace and χ^2_3 -log-normal models ($\mathbf{n} = \mathbf{n}_l + 10\mathbf{1}_8$, $l = 1, 2$, $\boldsymbol{\mu}_1 = (1, \mathbf{0}'_7)'$, $\boldsymbol{\mu}_2 = (1, 1, \mathbf{0}'_6)'$, $\boldsymbol{\mu}_3 = (1, 1, 1, \mathbf{0}'_5)'$).

Model	σ	\mathbf{n}_l	$\boldsymbol{\mu}$	Test							
				Q_N	Q_N^D	Q_N^U	Q_N^π	$Q_N^{D,\pi}$	$Q_N^{D,s,\pi}$	$Q_N^{U,\pi}$	$Q_N^{U,s,\pi}$
Log-normal	σ_1	\mathbf{n}_1	$\boldsymbol{\mu}_1$	96.1	97.4	68.9	91.5	91.2	92.3	81.3	82.3
			$\boldsymbol{\mu}_2$	99.8	99.7	87.7	99.8	99.0	99.1	94.0	94.0
			$\boldsymbol{\mu}_3$	99.9	99.5	91.8	99.8	99.6	99.6	97.5	97.5
		\mathbf{n}_2	$\boldsymbol{\mu}_1$	97.2	98.8	79.6	90.7	93.4	94.0	86.2	90.0
			$\boldsymbol{\mu}_2$	100	99.9	93.0	99.8	99.7	99.6	95.9	98.2
			$\boldsymbol{\mu}_3$	100	99.9	96.0	100	99.9	99.8	97.7	99.3
	σ_2	\mathbf{n}_1	$\boldsymbol{\mu}_1$	95.6	88.7	30.1	91.3	74.5	73.9	42.9	42.4
			$\boldsymbol{\mu}_2$	98.4	94.9	49.5	96.6	88.9	88.1	63.5	62.2
			$\boldsymbol{\mu}_3$	96.9	93.9	57.1	95.5	90.1	89.1	69.7	68.9
		\mathbf{n}_2	$\boldsymbol{\mu}_1$	97.1	93.3	43.3	91.3	79.8	80.4	39.1	56.6
			$\boldsymbol{\mu}_2$	99.4	97.4	67.3	98.0	92.2	92.4	63.4	78.0
			$\boldsymbol{\mu}_3$	98.0	95.5	72.8	95.9	92.1	92.2	70.5	82.0
σ_3	\mathbf{n}_2	$\boldsymbol{\mu}_1$	48.4	56.2	29.7	25.4	27.5	26.0	51.3	42.2	
		$\boldsymbol{\mu}_2$	74.0	81.6	49.3	48.6	57.0	56.2	73.7	68.6	
		$\boldsymbol{\mu}_3$	91.3	92.0	64.0	76.6	77.8	78.1	82.8	79.7	
Normal-Laplace	σ_1	\mathbf{n}_1	$\boldsymbol{\mu}_1$	75.1	75.4	69.9	63.5	65.9	65.8	70.8	70.6
			$\boldsymbol{\mu}_2$	94.9	95.2	93.0	90.1	91.0	91.0	93.1	93.0
			$\boldsymbol{\mu}_3$	98.6	98.6	97.4	97.4	97.3	97.3	97.6	97.4
		\mathbf{n}_2	$\boldsymbol{\mu}_1$	77.0	78.6	82.8	71.3	74.7	74.7	83.5	82.7
			$\boldsymbol{\mu}_2$	98.3	99.0	98.8	96.7	97.4	97.4	98.8	98.8
			$\boldsymbol{\mu}_3$	99.7	99.7	99.6	99.6	99.6	99.6	99.6	99.6
	σ_2	\mathbf{n}_1	$\boldsymbol{\mu}_1$	68.4	60.6	22.4	58.1	53.8	53.5	26.9	24.4
			$\boldsymbol{\mu}_2$	83.6	80.0	44.4	75.5	74.4	74.1	49.8	46.9
			$\boldsymbol{\mu}_3$	84.8	83.8	56.1	77.2	78.8	78.5	62.0	58.9
		\mathbf{n}_2	$\boldsymbol{\mu}_1$	73.2	69.8	40.2	66.2	65.0	65.0	24.2	42.1
			$\boldsymbol{\mu}_2$	90.2	88.9	69.2	85.7	85.1	85.4	54.6	70.6
			$\boldsymbol{\mu}_3$	91.9	92.0	79.4	87.7	88.8	88.8	66.1	80.5
σ_3	\mathbf{n}_2	$\boldsymbol{\mu}_1$	23.6	25.0	27.1	18.2	21.3	20.6	46.0	28.0	
		$\boldsymbol{\mu}_2$	43.1	50.3	46.3	35.4	45.5	44.0	65.8	46.4	
		$\boldsymbol{\mu}_3$	66.8	75.8	58.9	56.2	70.4	69.1	79.3	59.9	
χ^2_3 -Log-normal	σ_1	\mathbf{n}_1	$\boldsymbol{\mu}_1$	94.9	94.4	68.7	82.1	77.9	78.1	76.9	78.3
			$\boldsymbol{\mu}_2$	99.7	99.0	88.5	98.0	96.9	97.0	94.2	94.0
			$\boldsymbol{\mu}_3$	99.9	99.7	93.3	99.5	98.7	98.7	96.6	96.9
		\mathbf{n}_2	$\boldsymbol{\mu}_1$	93.3	94.9	80.9	81.8	81.6	82.3	84.5	87.9
			$\boldsymbol{\mu}_2$	99.7	99.8	95.7	98.6	97.9	98.0	96.3	98.1
			$\boldsymbol{\mu}_3$	100	100	97.3	99.8	99.6	99.6	98.0	99.3
	σ_2	\mathbf{n}_1	$\boldsymbol{\mu}_1$	93.3	82.4	29.3	80.2	59.8	59.5	37.9	36.9
			$\boldsymbol{\mu}_2$	96.4	92.2	50.4	92.4	80.1	79.9	61.2	59.6
			$\boldsymbol{\mu}_3$	95.2	91.3	58.8	91.1	84.4	84.2	69.0	67.7
		\mathbf{n}_2	$\boldsymbol{\mu}_1$	91.6	84.8	43.7	79.9	65.4	66.7	38.2	53.8
			$\boldsymbol{\mu}_2$	97.2	94.8	68.3	93.7	86.1	86.5	62.5	77.3
			$\boldsymbol{\mu}_3$	96.8	94.9	75.6	94.0	89.0	89.2	70.6	83.2
σ_3	\mathbf{n}_2	$\boldsymbol{\mu}_1$	30.1	39.5	25.0	16.7	21.1	18.4	47.2	32.2	
		$\boldsymbol{\mu}_2$	52.6	68.4	46.5	28.9	46.3	43.7	67.3	53.0	
		$\boldsymbol{\mu}_3$	76.7	89.1	60.1	51.9	71.0	69.8	80.1	68.9	

TABLE 10
 Empirical sizes (as percentages) of the tests obtained in the normal, χ_{10}^2 and log-normal models ($\mathbf{n} = (m + 5)\mathbf{1}_a$, $\boldsymbol{\sigma} = \mathbf{1}_a$, $m = 5, 10, 15, 20$, $a = 6, 8, 10, 12$).

Model	m	a	Test							
			Q_N	Q_N^D	Q_N^U	Q_N^π	$Q_N^{D,\pi}$	$Q_N^{D,s,\pi}$	$Q_N^{U,\pi}$	$Q_N^{U,s,\pi}$
Normal	5	6	11.2	9.9	5.6	4.2	4.4	4.4	5.1	4.9
		8	14.5	13.1	5.6	5.7	5.4	5.4	5.3	5.1
		10	14.1	12.6	4.8	3.9	4.1	4.1	4.7	4.7
		12	17.8	16.0	4.9	5.0	5.0	5.0	4.4	4.4
	10	6	9.6	8.4	6.0	5.2	5.2	5.2	5.6	5.6
		8	8.7	8.2	4.8	4.0	4.3	4.3	4.7	4.6
		10	11.7	11.0	5.3	5.0	4.9	4.9	4.9	5.0
		12	10.7	9.8	4.8	4.7	4.8	4.8	5.0	4.8
	15	6	6.8	6.4	5.4	4.9	4.6	4.7	5.1	5.2
		8	7.0	6.3	4.6	4.2	4.1	4.1	4.4	4.5
		10	9.9	9.5	5.6	6.1	5.9	5.9	5.4	5.3
		12	10.1	9.8	6.1	5.1	4.9	4.9	6.0	5.9
	20	6	6.0	5.4	4.4	4.7	4.5	4.5	4.3	4.2
		8	7.5	6.9	5.0	5.3	5.1	5.1	4.9	4.8
		10	7.2	6.8	4.4	4.0	4.0	4.0	4.5	4.5
		12	8.1	7.3	4.7	4.3	4.2	4.2	4.8	4.7
χ_{10}^2	5	6	12.0	10.4	4.9	5.4	5.4	5.4	5.0	4.8
		8	16.1	15.9	4.9	4.7	5.4	5.5	5.2	5.4
		10	18.4	18.2	5.1	4.4	5.0	5.0	5.7	5.7
		12	20.8	21.9	5.2	5.0	4.8	4.8	5.6	5.6
	10	6	10.3	9.6	5.0	5.7	5.2	5.2	5.3	5.2
		8	10.8	10.5	3.4	4.5	4.0	4.0	3.6	3.6
		10	12.1	13.4	4.3	4.4	4.3	4.3	4.6	4.5
		12	14.1	14.8	5.1	5.5	5.6	5.6	5.6	5.7
	15	6	8.1	8.3	5.4	4.9	5.0	5.0	5.3	5.5
		8	9.6	9.6	3.8	4.9	4.6	4.6	3.9	4.0
		10	10.7	11.6	4.4	5.3	5.2	5.2	4.7	4.8
		12	10.2	10.8	4.4	5.0	5.0	5.0	4.9	4.6
	20	6	8.7	8.1	5.9	5.5	5.5	5.5	5.9	5.8
		8	9.8	9.8	5.8	6.0	6.0	6.0	6.2	6.1
		10	9.9	10.4	5.6	5.9	5.6	5.6	5.9	5.7
		12	10.4	10.6	5.6	5.7	5.6	5.6	6.1	6.1
Log-normal	5	6	12.7	13.3	2.6	4.2	4.7	4.6	5.1	5.5
		8	19.0	24.7	1.8	4.5	5.3	5.2	5.1	5.1
		10	21.2	35.1	1.8	4.4	4.7	4.7	5.6	5.0
		12	23.2	43.9	1.2	4.2	4.6	4.7	3.8	3.7
	10	6	13.6	14.1	1.8	6.1	4.6	4.8	4.0	4.0
		8	13.6	21.2	1.1	4.8	4.0	4.0	4.8	5.1
		10	20.2	32.6	2.2	6.3	5.3	5.6	6.4	6.8
		12	21.2	41.2	0.8	5.3	4.7	4.6	4.2	3.8
	15	6	10.9	11.8	1.8	4.9	4.6	4.5	4.1	4.4
		8	13.7	21.2	1.7	5.2	4.3	4.2	4.4	4.6
		10	15.2	27.5	1.5	4.5	4.1	4.2	4.4	4.3
		12	17.9	35.7	1.7	4.1	4.5	4.3	5.6	5.6
	20	6	10.5	13.0	2.2	5.7	4.7	5.0	5.4	5.2
		8	14.0	19.6	2.0	6.2	5.8	5.8	5.0	5.1
		10	12.4	23.2	1.5	5.0	5.3	5.3	4.3	4.0
		12	16.9	31.1	1.4	4.9	4.6	4.6	4.9	5.1

TABLE 11
 Empirical powers (as percentages) of the tests obtained in the normal, χ^2_{10} and log-normal models ($\mathbf{n} = (m + 5)\mathbf{1}_a$, $\boldsymbol{\sigma} = \mathbf{1}_a$, $\boldsymbol{\mu}_1 = (1, \mathbf{0}'_{a-1})'$, $m = 5, 10, 15, 20$, $a = 6, 8, 10, 12$).

Model	m	a	Test								
			Q_N	Q_N^D	Q_N^U	Q_N^π	$Q_N^{D,\pi}$	$Q_N^{D,s,\pi}$	$Q_N^{U,\pi}$	$Q_N^{U,s,\pi}$	
Normal	5	6	65.1	64.1	56.9	47.0	50.7	50.7	53.2	53.5	
		8	61.9	60.7	51.3	41.2	43.4	43.4	49.9	50.1	
		10	62.7	61.4	48.1	37.2	39.3	39.2	47.4	47.6	
		12	60.3	59.9	44.6	34.4	36.1	36.3	43.4	43.6	
	10	6	78.8	78.8	74.2	70.0	71.3	71.3	73.6	73.5	
		8	78.1	77.8	72.5	66.4	68.8	68.8	72.0	71.9	
		10	77.1	77.3	71.0	62.4	64.2	64.3	70.8	70.6	
		12	73.1	73.2	65.6	56.8	58.5	58.5	65.4	65.5	
	15	6	90.8	91.0	90.2	86.1	87.5	87.5	89.3	89.1	
		8	87.1	87.1	84.9	80.8	82.3	82.3	84.9	84.6	
		10	87.5	87.6	84.8	79.2	81.0	81.0	84.8	84.6	
		12	85.7	85.5	81.9	73.8	75.1	75.1	81.3	81.5	
	20	6	95.9	96.0	95.4	93.6	93.9	93.9	95.1	95.1	
		8	95.5	95.7	95.1	91.7	93.0	93.0	94.9	95.0	
		10	92.4	92.8	91.6	89.0	89.7	89.7	91.6	91.5	
		12	91.5	92.0	90.6	87.0	87.6	87.6	90.4	90.5	
	χ^2_{10}	5	6	70.9	68.1	54.6	48.9	49.9	50.1	53.5	54.2
			8	69.9	68.1	48.9	44.1	44.5	44.8	49.2	49.2
			10	70.0	69.3	45.8	42.6	42.9	43.0	46.9	47.0
			12	68.9	69.2	43.0	36.2	36.1	36.1	43.7	44.2
		10	6	86.4	86.4	77.1	77.0	77.2	77.3	77.2	77.1
			8	83.8	83.6	71.9	72.3	71.2	71.5	72.4	72.9
			10	84.5	84.6	69.8	66.5	65.9	66.1	70.9	70.5
			12	80.1	81.5	65.8	60.1	59.1	59.2	66.3	66.6
15		6	94.3	94.6	88.9	90.4	90.4	90.4	88.7	88.9	
		8	94.5	94.3	88.2	89.0	87.4	87.4	88.7	88.7	
		10	93.7	93.5	85.1	84.6	84.0	84.0	85.1	85.4	
		12	92.2	91.7	82.9	82.1	81.0	81.0	83.1	83.3	
20		6	98.6	98.6	96.6	96.8	96.6	96.7	96.0	96.0	
		8	98.1	98.1	95.1	96.3	96.1	96.1	95.4	95.3	
		10	97.1	96.8	93.2	93.4	93.2	93.2	93.5	93.6	
		12	97.0	97.3	92.7	93.2	92.6	92.7	93.0	93.1	
Log-normal		5	6	92.8	90.6	60.6	84.7	83.6	83.9	70.8	71.4
			8	95.1	95.1	53.6	79.5	78.0	78.6	64.7	64.9
			10	93.6	96.1	44.5	76.5	72.0	72.7	58.2	58.8
			12	92.9	98.5	41.0	71.7	66.7	67.8	57.5	57.9
		10	6	96.9	96.2	73.1	93.3	93.6	94.0	82.2	83.2
			8	96.1	97.4	68.9	91.5	91.2	92.3	81.3	82.3
			10	95.9	98.6	61.9	90.3	88.4	89.0	76.2	77.1
			12	96.9	98.4	56.0	86.3	84.1	85.0	73.9	74.9
	15	6	98.5	99.0	85.9	97.1	97.8	98.3	92.4	92.8	
		8	98.0	99.1	81.1	94.8	96.0	96.2	88.8	89.5	
		10	97.8	99.5	78.4	93.7	94.4	95.0	88.6	89.4	
		12	96.9	99.0	72.0	92.2	90.9	91.8	85.8	86.6	
	20	6	98.4	99.2	89.2	97.6	99.0	99.3	94.7	95.1	
		8	98.4	99.6	87.9	97.4	97.9	98.0	94.9	95.2	
		10	98.6	99.8	85.0	96.5	97.8	98.1	93.6	93.8	
		12	97.7	99.7	82.1	94.9	95.2	95.5	92.6	93.1	

Acknowledgements

The author is thankful to four anonymous referees and the editors for helpful suggestions and comments for the improvement of this manuscript.

References

- [1] BAI, Z. D. and SARANADASA, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statist. Sinica* **6** 311–329. [MR1399305](#)
- [2] BOX, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics* **25** 290–302. [MR0061787](#)
- [3] CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. [MR2604697](#)
- [4] CHUNG, E. Y. and ROMANO, J. P. (2013). Exact and asymptotically robust permutation tests. *Ann. Statist.* **41** 484–507. [MR3099111](#)
- [5] DEMŠAR, J. (2006). Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **7** 1–30. [MR2274360](#)
- [6] DONG, K., PANG, H., TONG, T. and GENTON, M. G. (2016). Shrinkage-based diagonal Hotelling's tests for high-dimensional small sample size data. *J. Multivariate Anal.* **143** 127–142. [MR3431423](#)
- [7] DUCHESNE, P. and FRANCO, C. (2015). Multivariate hypothesis testing using generalized and $\{2\}$ -inverses - with applications. *Statistics* **49** 475–496. [MR3349073](#)
- [8] FRIEDRICH, S., KONIETSCHKE, F. and PAULY, M. (2015). GFD: Tests for general factorial designs. R package version 0.1.2. <http://CRAN.R-project.org/package=GFD>
- [9] GETSON, A. J. and HSUAN, F. C. (1988). $\{2\}$ -Inverses and Their Statistical Application. Lecture Notes in Statistics **47**. Springer-Verlag, New York. [MR0971746](#)
- [10] GÓRECKI, T. and SMAGA L. (2015). A comparison of tests for the one-way ANOVA problem for functional data. *Comput. Stat.* **30** 987–1010. [MR3433439](#)
- [11] HARVILLE, D. A. (2008). *Matrix Algebra from a Statistician's Perspective*. Springer, New York. [MR1874239](#)
- [12] JANSSEN, A. (1997). Studentized permutation tests for non-i.i.d. hypotheses and the generalized Behrens-Fisher problem. *Statist. Probab. Lett.* **36** 9–21. [MR1491070](#)
- [13] KONIETSCHKE, F., BATHKE, A. C., HARRAR, S. W. and PAULY, M. (2015). Parametric and nonparametric bootstrap methods for general MANOVA. *J. Multivariate Anal.* **140** 291–301. [MR3372569](#)
- [14] KONIETSCHKE, F. and PAULY, M. (2012). A studentized permutation test

- for the non-parametric Behrens-Fisher problem in paired data. *Electron. J. Stat.* **6** 1358–1372. [MR2988450](#)
- [15] MATHAI, A. M. and PROVOST, S. B. (1992). *Quadratic Forms in Random Variables*. Marcel Dekker Inc., New York. [MR1192786](#)
- [16] MONTGOMERY, D. C. (2001). *Design and Analysis of Experiments*. 5th Ed. John Wiley & Sons Inc., New York. [MR0428633](#)
- [17] NEMENYI, P. B. (1963). Distribution-free multiple comparisons. Ph.D. dissertation, Princeton University. [MR2614171](#)
- [18] NEUHAUS, G. (1993). Conditional rank tests for the two-sample problem under random censorship. *Ann. Statist.* **21** 1760–1779. [MR1245767](#)
- [19] PARK, J. and NAG AYYALA, D. (2013). A test for the mean vector in large dimension and small samples. *J. Statist. Plann. Inference* **143** 929–943. [MR3011304](#)
- [20] PAULY, M., BRUNNER, E. and KONIETSCHKE, F. (2015). Asymptotic permutation tests in general factorial designs. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **77** 461–473. [MR3310535](#)
- [21] PAULY, M., ELLENBERGER, D. and BRUNNER, E. (2015). Analysis of high-dimensional one group repeated measures designs. *Statistics* **49** 1243–1261. [MR3415218](#)
- [22] R CORE TEAM (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- [23] SMAGA, Ł. (2015). Wald-type statistics using $\{2\}$ -inverses for hypothesis testing in general factorial designs. *Statist. Probab. Lett.* **107** 215–220. [MR3412779](#)
- [24] SMAGA, Ł. (2016). Bootstrap methods for multivariate hypothesis testing. *Comm. Statist. Simulation Comput.* <http://dx.doi.org/10.1080/03610918.2016.1248573>
- [25] SRIVASTAVA, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *J. Multivariate Anal.* **100** 518–532. [MR2483435](#)
- [26] SRIVASTAVA, M. S. and DU, M. (2008). A test for the mean vector with fewer observations than the dimension. *J. Multivariate Anal.* **99** 386–402. [MR2396970](#)
- [27] SRIVASTAVA, M. S., KATAYAMA, S. and KANO, Y. (2013). A two sample test in high dimensional data. *J. Multivariate Anal.* **114** 349–358. [MR2993891](#)
- [28] TONG, T. and WANG, Y. (2007). Optimal shrinkage estimation of variances with applications to microarray data analysis. *J. Amer. Statist. Assoc.* **102** 113–122. [MR2293304](#)
- [29] WU, Y., GENTON, M. G. and STEFANSKI, L. A. (2006). A multivariate two-sample mean test for small sample size and missing data. *Biometrics* **62** 877–885. [MR2247217](#)
- [30] ZHANG, J. T. (2005). Approximate and asymptotic distributions of chi-squared-type mixtures with applications. *J. Amer. Statist. Assoc.* **100** 273–285. [MR2156837](#)

- [31] ZHANG, J. T. (2013). *Analysis of Variance for Functional Data*. Chapman & Hall, London.
- [32] ZHANG, J. and XU, J. (2009). On the k-sample Behrens-Fisher problem for high-dimensional data. *Science in China Series A: Mathematics* **52** 1285–1304. [MR2520575](#)