Supplementary material for directional testing for one-way MANOVA in divergent dimensions

Caizhu Huang*[1,2], Claudia Di Caterina†[3], and Nicola Sartori‡[2]

[1]School of Digital Economics, Guandong University of Finance and Economics

[2]Department of Statistical Sciences, University of Padova

[3]Department of Economics, University of Verona

Abstract

In this supplementary material, Section S1 reports the additional simulation studies for the homoscedastic one-way MANOVA. Sections S1.1–S1.3 investigate the different values of the number p of variables and the number g of groups. The directional test performs well. The log-likelihood ratio test breaks down even when p is small. The other chi-square approximations, i.e. Bartlett correction and two Skovgaard's modifications, become worse as p increasing, especially in some extreme settings in Section S1.3. The corresponding simulation results for robustness of misspecification are reported in

*caizhu.huang@unipd.it

†claudia.dicaterina@univr.it

‡nicola.sartori@unipd.it

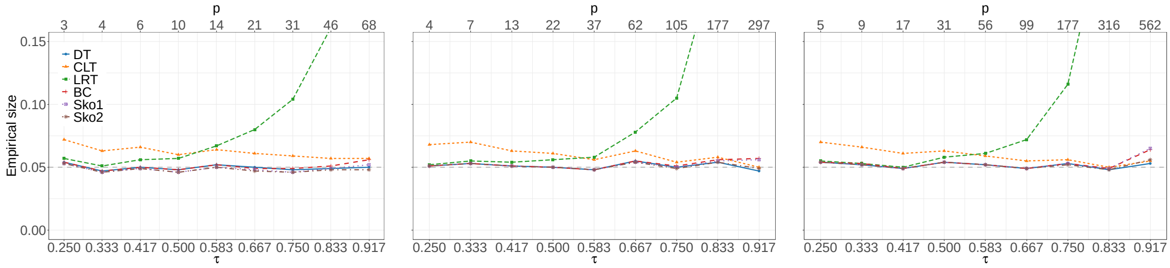

Figure S1: Empirical size of the directional test (DT), central limit theorem test (CLT), likelihood ratio test (LRT), Bartlett correction (BC) and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (8) in the paper, at nominal level $\alpha = 0.05$ given by the gray horizontal line. The left, middle and right panels correspond to $n_i = 100, 500, 1000$, respectively ($g = 3$).

Section S1.4. On the other hand, additional simulation studies for the heteroscedastic one-way MANOVA are showed in Section S2.

S1 Simulation studies for Homoscedastic one-way MANOVA

This section reports additional empirical result for homoscedastic one-way MANOVA in the multivariate normal framework. We compare the performance of exact directional test (DT) with other five approximate approaches: the central limite theorem test (CLT), log-likelihood ratio test (LRT), Bartlett correction (BC) and two Skovgaard's modifications (Sko1 and Sko2). The six tests are evaluated in terms of empirical size. The simulation results are computed via Monte Carlo simulation based on 10,000 replications.

S1.1 Empirical results for moderate setup

Groups of size $n_i, i \in \{1, \ldots, g\}$, are generated from a p-variate standard normal distribution $N_p(0_p, I_p)$ under the null hypothesis. For each simulation experiment, we show results for $p = \lfloor n_i^\tau \rfloor$ with $\tau = j/24$, $j \in \{6, 8, \cdots, 22\}$ and $n_i \in \{100, 500, 1000\}$. Throughout, we set $g = 3$ and $n_1 = n_2 = n_3$.

Table S1: Empirical size of the directional test (DT), central limit theorem test (CLT), likelihood ratio test (LRT), Bartlett correction (BC) and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (8) in the paper with $n_i = 100, 500, 1000$, at nominal level $\alpha = 0.05$.

n_i	$\tau(p)$	DT	CLT	LRT	BC	Sko1	Sko2
100	0.250 (3)	0.054	0.072	0.057	0.054	0.053	0.053
	0.333 (4)	0.047	0.063	0.051	0.047	0.046	0.046
	0.417 (6)	0.050	0.066	0.056	0.050	0.049	0.049
	0.500 (10)	0.048	0.060	0.057	0.048	0.046	0.046
	0.580 (14)	0.052	0.064	0.067	0.052	0.050	0.050
	0.667 (21)	0.050	0.061	0.080	0.049	0.048	0.047
	0.750 (31)	0.048	0.059	0.104	0.049	0.046	0.046
	0.833 (46)	0.049	0.057	0.161	0.051	0.049	0.048
	0.917 (68)	0.050	0.057	0.311	0.056	0.052	0.048
500	0.250 (4)	0.051	0.068	0.052	0.051	0.051	0.051
	0.333 (7)	0.053	0.070	0.055	0.053	0.053	0.053
	0.417 (13)	0.051	0.063	0.054	0.051	0.051	0.051
	0.500 (22)	0.050	0.061	0.056	0.050	0.050	0.050
	0.580 (37)	0.048	0.056	0.058	0.048	0.048	0.048
	0.667 (62)	0.055	0.063	0.078	0.055	0.054	0.054
	0.750 (105)	0.050	0.054	0.105	0.051	0.050	0.049
	0.833 (177)	0.054	0.058	0.222	0.056	0.055	0.054
	0.917 (297)	0.047	0.050	0.609	0.057	0.056	0.049
1000	0.250 (5)	0.054	0.070	0.055	0.054	0.054	0.054
	0.333 (9)	0.052	0.066	0.053	0.053	0.052	0.052
	0.417 (17)	0.049	0.061	0.050	0.049	0.049	0.049
	0.500 (31)	0.054	0.063	0.058	0.054	0.054	0.054
	0.580 (56)	0.052	0.059	0.061	0.052	0.052	0.052
	0.667 (99)	0.049	0.055	0.072	0.049	0.049	0.049
	0.750 (177)	0.053	0.056	0.116	0.053	0.053	0.052
	0.833 (316)	0.048	0.050	0.261	0.049	0.049	0.048
	0.917 (562)	0.053	0.055	0.794	0.064	0.065	0.056

S1.2 Empirical results for large number k of groups

Table S2: Empirical size of the directional test (DT), central limit theorem test (CLT), likelihood ratio test (LRT), Bartlett correction (BC) and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (8) in the paper with $p = n_i^\tau$, $n_i = 100$ and $g = 30$, at nominal level $\alpha = 0.05$

$\tau(p)$	DT	CLT	LRT	BC	Sko1	Sko2
0.250 (3)	0.053	0.060	0.057	0.053	0.053	0.053
0.333 (4)	0.050	0.057	0.055	0.050	0.050	0.050
0.417 (6)	0.050	0.056	0.057	0.050	0.050	0.050
0.500 (10)	0.049	0.053	0.059	0.049	0.049	0.049
0.580 (14)	0.051	0.056	0.064	0.052	0.051	0.051
0.667 (21)	0.050	0.054	0.071	0.050	0.050	0.050
0.750 (31)	0.052	0.054	0.080	0.052	0.052	0.052
0.833 (46)	0.048	0.051	0.095	0.049	0.048	0.048
0.917 (68)	0.054	0.056	0.137	0.054	0.054	0.053

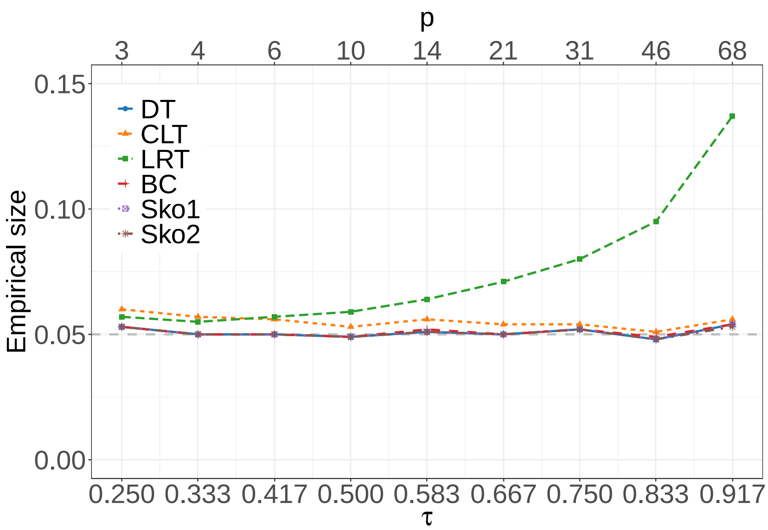

Figure S2: Empirical size of the directional test (DT), central limit theorem test (CLT), likelihood ratio test (LRT), Bartlett correction (BC) and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (8) in the paper with $n_i = 100$ and $g = 30$, at nominal level $\alpha = 0.05$ given by the gray horizontal line.

S1.3 Empirical results for the setup of He et al. (2021, Section A.3)

In each Monte Carlo experiment, we show results for $p = \lfloor n^\tau \rfloor$ with $n = \sum_{i=1}^{g} n_i$. Under the null hypothesis, we set $g = 3$, $n_1 = n_2 = n_3$, $\tau = j/24$ with $j \in \{6, 8, \cdots, 22\}$ and $n_i = 100$.

Table S3: Empirical size of the directional test (DT), central limit theorem test (CLT), likelihood ratio test (LRT), Bartlett correction (BC) and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (8) in the paper with $p = n^\tau$ and $n = \sum_{i=1}^{3} 100 = 300$, at nominal level $\alpha = 0.05$.

τ (p)	DT	CLT	LRT	BC	Sko1	Sko2
0.250 (4)	0.048	0.064	0.052	0.047	0.046	0.046
0.333 (6)	0.049	0.066	0.056	0.049	0.047	0.047
0.417 (10)	0.050	0.064	0.061	0.050	0.049	0.049
0.500 (17)	0.053	0.065	0.074	0.053	0.051	0.051
0.583 (27)	0.046	0.056	0.088	0.047	0.045	0.045
0.667 (44)	0.051	0.060	0.157	0.053	0.050	0.049
0.750 (72)	0.048	0.054	0.347	0.054	0.052	0.047
0.833 (115)	0.050	0.056	0.832	0.078	0.077	0.057
0.917 (186)	0.044	0.049	1.000	0.280	0.278	0.106

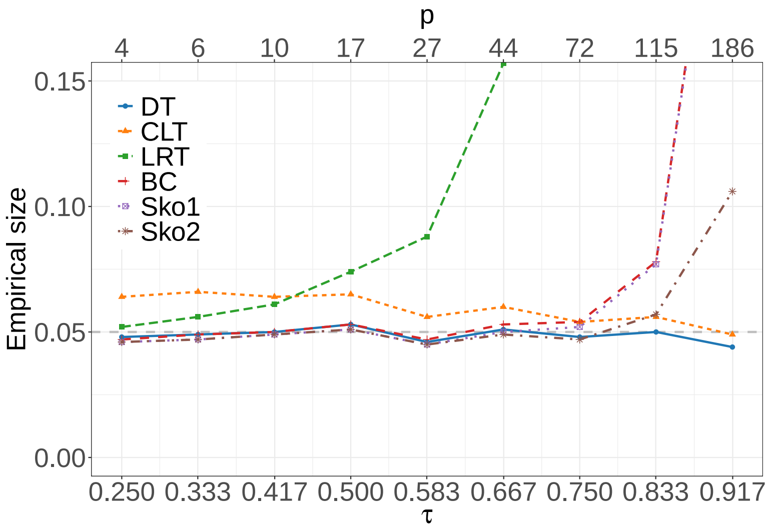

Figure S3: Empirical size of the directional test (DT), central limit theorem test (CLT), likelihood ratio test (LRT), Bartlett correction (BC) and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (8) in the paper with $p = n^\tau$ and $n = \sum_{i=1}^{3} 100 = 300$, at nominal level $\alpha = 0.05$ given by the gray horizontal line.

Table S4: Empirical size of the directional test (DT), central limit theorem test (CLT), likelihood ratio test (LRT), Bartlett correction (BC) and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (8) in the paper with $p = n^\tau$ and $n = \sum_{i=1}^{3} 500 = 1500$, at nominal level $\alpha = 0.05$.

τ (p)	DT	CLT	LRT	BC	Sko1	Sko2
0.250 (6)	0.052	0.068	0.053	0.052	0.052	0.052
0.333 (11)	0.051	0.063	0.053	0.051	0.050	0.050
0.417 (21)	0.050	0.059	0.053	0.050	0.049	0.049
0.500 (38)	0.051	0.060	0.062	0.051	0.051	0.051
0.583 (71)	0.051	0.059	0.080	0.051	0.050	0.050
0.667 (131)	0.051	0.056	0.146	0.052	0.051	0.051
0.750 (241)	0.052	0.057	0.402	0.058	0.058	0.053
0.833 (443)	0.052	0.056	0.976	0.081	0.083	0.063
0.917 (815)	0.049	0.051	1.000	0.420	0.462	0.171

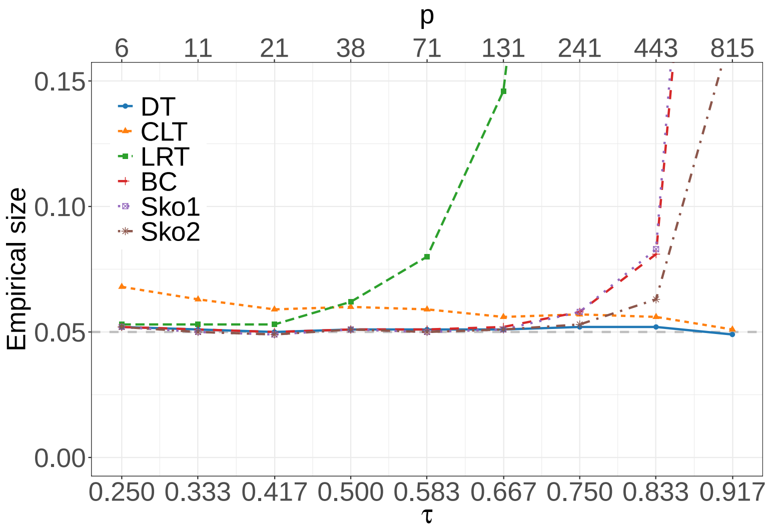

Figure S4: Empirical size of the directional test (DT), central limit theorem test (CLT), likelihood ratio test (LRT), Bartlett correction (BC) and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (8) in the paper with $p = n^\tau$ and $n = \sum_{i=1}^{3} 500 = 1500$, at nominal level $\alpha = 0.05$ given by the gray horizontal line.

S1.4 Robustness to misspecification

In this section we investigate the robustness to misspecification. The true generating processes are multivariate t, multivariate skew-normal or multivariate Laplace distributions. Here we setup $p = n_i^\tau$ with $n_i \in \{100, 500\}$.

More in detail, a multivariate t distribution with location 0_p, scale matrix I_p and degrees of

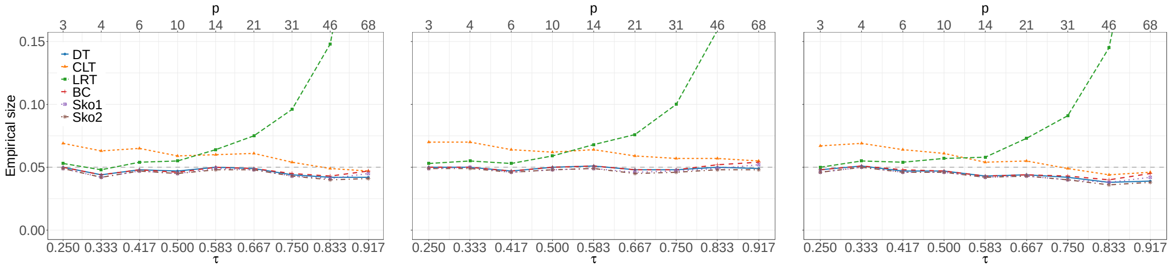

Figure S5: Empirical size for the directional test (DT), central limit theorem test (CLT), likelihood ratio test (LRT), Bartlett correction (BC) and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (8) in the paper, at nominal level $\alpha = 0.05$ given by the gray horizontal line. The left, middle and right panels correspond to multivariate t, multivariate skew-normal, and multivariate Laplace distributions of the true generating process, respectively, with $n_i = 100$ and $g = 3$.

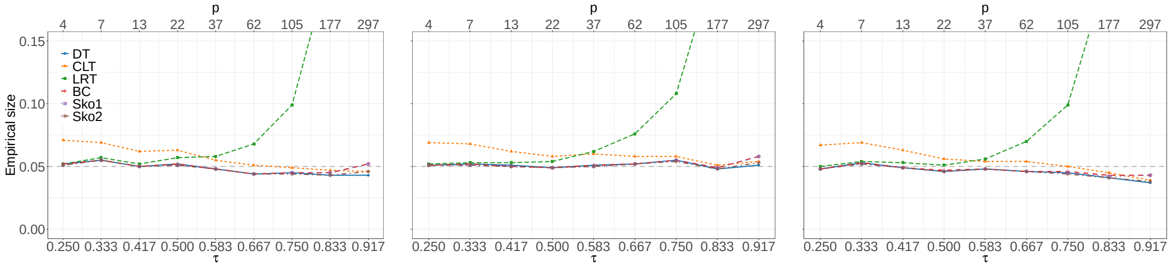

Figure S6: Empirical size for the directional test (DT), central limit theorem test (CLT), likelihood ratio test (LRT), Bartlett correction (BC) and two Skovgaard's modifications (Sko1 and Sko2, respectively) for hypothesis (8) in the paper, at nominal level $\alpha = 0.05$ given by the gray horizontal line. The left, middle and right panels correspond to multivariate t, multivariate skew-normal, and multivariate Laplace distributions of the true generating process, respectively, with $n_i = 500$ and $g = 3$.

freedom 5, a multivariate skew-normal distribution with location 1_p, scale matrix $\Omega = (\omega_{jl}) = (0.2)^{|j-l|}$ and shape parameter 1_p, and a multivariate Laplace distribution with mean vector 1_p and identity covariance matrix.

For hypothesis (8) in the paper, Figures S5–S6 and Tables S5–S6 show the empirical size at the nominal level $\alpha = 0.05$ if the underling distribution is misspecified. We see that the directional test still maintains the hightest accuracy.

Table S5: Empirical size of the directional test (DT), central limit theorem test (CLT), likelihood ratio test (LRT), Bartlett correction (BC) and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (8) in the paper with $p = n_i^\tau$ and $n_i = 100$ and $g = 3$, at nominal level $\alpha = 0.05$.

True Distribution	τ (p)	DT	CLT	LRT	BC	Sko1	Sko2
Multivariate t	0.250 (3)	0.050	0.069	0.053	0.050	0.049	0.049
	0.333 (4)	0.044	0.063	0.048	0.044	0.042	0.042
	0.417 (6)	0.048	0.065	0.054	0.048	0.047	0.047
	0.500 (10)	0.047	0.059	0.055	0.046	0.045	0.045
	0.583 (14)	0.050	0.060	0.064	0.050	0.049	0.048
	0.667 (21)	0.049	0.061	0.075	0.049	0.048	0.048
	0.750 (31)	0.044	0.054	0.096	0.045	0.043	0.043
	0.833 (46)	0.042	0.049	0.148	0.043	0.041	0.040
	0.917 (68)	0.042	0.047	0.302	0.047	0.045	0.041
Multivariate skew-normal	0.250 (3)	0.050	0.070	0.053	0.050	0.049	0.049
	0.333 (4)	0.050	0.070	0.055	0.050	0.050	0.049
	0.417 (6)	0.047	0.064	0.053	0.047	0.046	0.046
	0.500 (10)	0.050	0.062	0.059	0.050	0.048	0.048
	0.583 (14)	0.051	0.064	0.068	0.051	0.049	0.049
	0.667 (21)	0.048	0.059	0.076	0.048	0.046	0.045
	0.750 (31)	0.048	0.057	0.100	0.048	0.047	0.046
	0.833 (46)	0.050	0.057	0.159	0.052	0.049	0.048
	0.917 (68)	0.049	0.055	0.312	0.054	0.052	0.048
Multivariate Laplace	0.250 (3)	0.048	0.067	0.050	0.048	0.046	0.046
	0.333 (4)	0.051	0.069	0.055	0.051	0.050	0.050
	0.417 (6)	0.047	0.064	0.054	0.048	0.046	0.046
	0.500 (10)	0.047	0.061	0.057	0.047	0.046	0.046
	0.583 (14)	0.043	0.054	0.058	0.043	0.042	0.042
	0.667 (21)	0.044	0.055	0.073	0.044	0.043	0.043
	0.750 (31)	0.042	0.049	0.091	0.043	0.040	0.040
	0.833 (46)	0.038	0.044	0.145	0.040	0.038	0.036
	0.917 (68)	0.039	0.046	0.301	0.045	0.042	0.038

Table S6: Empirical size of the directional test (DT), central limit theorem test (CLT), likelihood ratio test (LRT), Bartlett correction (BC) and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (8) in the paper with $p = n_i^{\tau}$ and $n_i = 500$ and $g = 3$, at nominal level $\alpha = 0.05$.

True Distribution	τ (p)	DT	CLT	LRT	BC	Sko1	Sko2
Multivariate t	0.250 (4)	0.052	0.071	0.052	0.052	0.051	0.051
	0.333 (7)	0.055	0.069	0.057	0.055	0.055	0.055
	0.417 (13)	0.050	0.062	0.052	0.050	0.050	0.050
	0.500 (22)	0.052	0.063	0.057	0.052	0.051	0.051
	0.583 (37)	0.048	0.055	0.058	0.048	0.048	0.048
	0.667 (62)	0.044	0.051	0.068	0.044	0.044	0.044
	0.750 (105)	0.045	0.049	0.099	0.045	0.045	0.044
	0.833 (177)	0.043	0.047	0.212	0.045	0.045	0.043
	0.917 (297)	0.043	0.046	0.618	0.052	0.052	0.046
Multivariate skew-normal	0.250 (4)	0.051	0.069	0.052	0.051	0.051	0.051
	0.333 (7)	0.052	0.068	0.053	0.052	0.051	0.051
	0.417 (13)	0.051	0.062	0.053	0.050	0.050	0.050
	0.500 (22)	0.049	0.058	0.054	0.049	0.049	0.049
	0.583 (37)	0.051	0.060	0.062	0.051	0.050	0.050
	0.667 (62)	0.052	0.058	0.076	0.052	0.052	0.052
	0.750 (105)	0.055	0.058	0.108	0.055	0.055	0.054
	0.833 (177)	0.048	0.051	0.218	0.049	0.049	0.048
	0.917 (297)	0.051	0.054	0.612	0.058	0.058	0.053
Multivariate Laplace	0.250 (4)	0.048	0.067	0.050	0.048	0.048	0.048
	0.333 (7)	0.053	0.069	0.054	0.053	0.052	0.052
	0.417 (13)	0.049	0.063	0.053	0.049	0.049	0.049
	0.500 (22)	0.046	0.056	0.051	0.047	0.046	0.046
	0.583 (37)	0.048	0.054	0.056	0.048	0.048	0.048
	0.667 (62)	0.046	0.054	0.070	0.046	0.046	0.046
	0.750 (105)	0.045	0.050	0.099	0.046	0.045	0.044
	0.833 (177)	0.041	0.045	0.209	0.043	0.042	0.041
	0.917 (297)	0.037	0.039	0.609	0.043	0.043	0.038

S2 Simulation studies for Heteroscedastic one-way MANOVA

This section is studied the performance of directional test for heteroscedastic one-way MANOVA, comparing with LRT, Sko1 and Sko1. In particular, when the number of groups $g = 2$, we also consider the F-approximation for the Behrens-Fisher test T^{*2}. The simulation results are computed via Monte Carlo simulation based on 10,000 replications.

S2.1 Empirical results for the moderate setup

Groups of size $n_i, i \in \{1, \ldots, g\}$, are generated from a p-variate standard normal distribution $N_p(0_p, \Lambda_i^{-1})$ under the null hypothesis. We use an autoregressive structure for the covariance matrices. i.e. $\Lambda_i^{-1} = (\sigma_{jl})_{p\times p} = (\rho_i^{|j-l|})_{p\times p}$, with the ρ_i chosen to an equally-distance sequence from 0.1 to 0.9 of length g. For each simulation experiment, we show results for $p = \lceil n_i^\tau \rceil$ with $\tau = j/24$, $j \in \{6, 7, \cdots, 22\}$ and $n_i \in \{100, 500, 1000\}$ and $k = 2$.

Table S7: Empirical size of the directional test (DT), Behrens-Fisher test (BF) (Nel and Merwe, 1986), likelihood ratio test (LRT), and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (9) in the paper with $p = n_i^\tau$ and $n_i = 100, 500, 1000$, at nominal level $\alpha = 0.05$.

n_i	τ (p)	DT	BF	LRT	Sko1	Sko2
100	0.250 (4)	0.053	0.054	0.057	0.053	0.053
	0.333 (5)	0.050	0.052	0.056	0.050	0.050
	0.417 (7)	0.049	0.051	0.057	0.050	0.050
	0.500 (10)	0.051	0.056	0.065	0.052	0.052
	0.583 (15)	0.052	0.061	0.077	0.052	0.052
	0.667 (22)	0.049	0.065	0.094	0.051	0.050
	0.750 (32)	0.051	0.082	0.147	0.055	0.053
	0.833 (47)	0.052	0.115	0.270	0.067	0.061
	0.917 (69)	0.064	0.183	0.594	0.112	0.092
500	0.250 (5)	0.050	0.050	0.051	0.050	0.050
	0.333 (8)	0.051	0.051	0.052	0.051	0.051
	0.417 (14)	0.051	0.052	0.055	0.051	0.051
	0.500 (23)	0.051	0.054	0.059	0.051	0.051
	0.583 (38)	0.054	0.061	0.070	0.054	0.054
	0.667 (63)	0.051	0.068	0.089	0.052	0.052
	0.750 (106)	0.048	0.084	0.143	0.051	0.050
	0.833 (178)	0.051	0.154	0.382	0.062	0.058
	0.917 (298)	0.060	0.392	0.923	0.137	0.106
1000	0.250 (6)	0.052	0.052	0.053	0.052	0.052
	0.333 (10)	0.050	0.050	0.051	0.050	0.050
	0.417 (18)	0.046	0.048	0.048	0.046	0.046
	0.500 (32)	0.052	0.055	0.058	0.052	0.052
	0.583 (57)	0.052	0.058	0.063	0.052	0.052
	0.667 (100)	0.050	0.066	0.083	0.050	0.050
	0.750 (178)	0.047	0.088	0.154	0.050	0.049
	0.833 (317)	0.051	0.179	0.459	0.065	0.059
	0.917 (563)	0.064	0.528	0.987	0.155	0.112

S2.2 Robustness to misspecification

Table S8: Empirical size of the directional test (DT), Behrens-Fisher test (BF) (Nel and Merwe, 1986), likelihood ratio test (LRT), and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (9) in the paper with $p = n_i^\tau$ and $n_i = 100$ and $g = 2$, at nominal level $\alpha = 0.05$.

Distribution	τ (p)	DT	BF	LRT	Sko1	Sko2
Multivariate t	0.250 (4)	0.048	0.048	0.052	0.048	0.048
	0.333 (5)	0.047	0.047	0.053	0.047	0.047
	0.417 (7)	0.048	0.048	0.055	0.048	0.048
	0.500 (10)	0.050	0.049	0.060	0.050	0.050
	0.583 (15)	0.045	0.044	0.064	0.045	0.045
	0.667 (22)	0.047	0.046	0.078	0.047	0.047
	0.750 (32)	0.042	0.040	0.102	0.043	0.042
	0.833 (47)	0.041	0.038	0.166	0.044	0.042
	0.917 (69)	0.043	0.038	0.375	0.054	0.048
Multivariate skew-normal	0.250 (4)	0.046	0.046	0.051	0.046	0.046
	0.333 (5)	0.052	0.052	0.058	0.052	0.052
	0.417 (7)	0.053	0.053	0.061	0.053	0.053
	0.500 (10)	0.054	0.054	0.064	0.054	0.054
	0.583 (15)	0.049	0.049	0.070	0.050	0.049
	0.667 (22)	0.048	0.047	0.080	0.048	0.048
	0.750 (32)	0.050	0.048	0.110	0.051	0.050
	0.833 (47)	0.052	0.051	0.187	0.056	0.053
	0.917 (69)	0.048	0.048	0.379	0.061	0.053
Multivariate Laplace	0.250 (4)	0.049	0.049	0.054	0.049	0.049
	0.333 (5)	0.044	0.044	0.050	0.044	0.044
	0.417 (7)	0.045	0.045	0.051	0.045	0.045
	0.500 (10)	0.046	0.046	0.055	0.046	0.045
	0.583 (15)	0.045	0.045	0.063	0.045	0.045
	0.667 (22)	0.044	0.043	0.075	0.044	0.044
	0.750 (32)	0.040	0.039	0.097	0.042	0.040
	0.833 (47)	0.039	0.038	0.169	0.043	0.040
	0.917 (69)	0.034	0.032	0.372	0.044	0.038

Table S9: Empirical size of the directional test (DT), Behrens-Fisher test (BF) (Nel and Merwe, 1986), likelihood ratio test (LRT), and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (9) in the paper with $p = n_i^\tau$ and $n_i = 500$ and $g = 2$, at nominal level $\alpha = 0.05$.

Distribution	τ (p)	DT	BF	LRT	Sko1	Sko2
Multivariate t	0.250 (5)	0.050	0.050	0.052	0.050	0.050
	0.333 (8)	0.049	0.049	0.050	0.049	0.049
	0.417 (14)	0.047	0.047	0.050	0.047	0.047
	0.500 (23)	0.045	0.045	0.050	0.045	0.045
	0.583 (38)	0.051	0.051	0.063	0.051	0.051
	0.667 (63)	0.044	0.044	0.070	0.044	0.044
	0.750 (106)	0.044	0.043	0.108	0.044	0.044
	0.833 (178)	0.044	0.043	0.236	0.048	0.046
	0.917 (298)	0.047	0.045	0.705	0.064	0.053
Multivariate skew-normal	0.250 (5)	0.051	0.051	0.052	0.051	0.051
	0.333 (8)	0.051	0.051	0.052	0.051	0.051
	0.417 (14)	0.055	0.055	0.059	0.055	0.055
	0.500 (23)	0.054	0.054	0.060	0.054	0.054
	0.583 (38)	0.051	0.051	0.061	0.051	0.051
	0.667 (63)	0.049	0.049	0.074	0.049	0.049
	0.750 (106)	0.051	0.051	0.117	0.052	0.052
	0.833 (178)	0.048	0.048	0.242	0.054	0.050
	0.917 (298)	0.052	0.051	0.699	0.070	0.058
Multivariate Laplace	0.250 (5)	0.054	0.054	0.054	0.053	0.053
	0.333 (8)	0.053	0.053	0.054	0.053	0.053
	0.417 (14)	0.050	0.050	0.052	0.050	0.050
	0.500 (23)	0.051	0.051	0.056	0.051	0.051
	0.583 (38)	0.047	0.047	0.056	0.047	0.047
	0.667 (63)	0.046	0.046	0.070	0.047	0.046
	0.750 (106)	0.042	0.042	0.106	0.043	0.043
	0.833 (178)	0.040	0.040	0.235	0.043	0.040
	0.917 (298)	0.033	0.032	0.708	0.051	0.040

S2.3 Empirical results for large number k of groups

Table S10: Empirical size of the directional test (DT), likelihood ratio test (LRT), and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (9) in the paper with $p = n_i^\tau$ and $n_i = 100$ and $g = 30$, at nominal level $\alpha = 0.05$.

τ (p)	DT	LRT	Sko1	Sko2
0.250 (4)	0.050	0.081	0.050	0.050
0.333 (5)	0.050	0.091	0.050	0.050
0.417 (7)	0.049	0.122	0.050	0.049
0.500 (10)	0.050	0.186	0.051	0.050
0.583 (15)	0.054	0.391	0.058	0.054
0.667 (22)	0.047	0.751	0.056	0.048
0.750 (32)	0.052	0.990	0.087	0.062
0.833 (47)	0.049	1.000	0.223	0.095
0.917 (69)	0.048	1.000	0.932	0.405

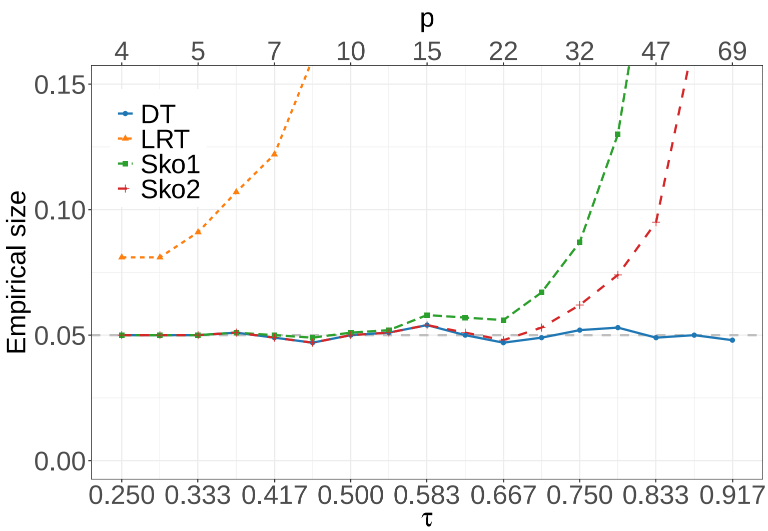

Figure S7: Empirical size of the directional test (DT), likelihood ratio test (LRT), and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (9) in the paper with $p = n_i^\tau$ and $n_i = 100$ and $g = 30$, at nominal level $\alpha = 0.05$ given by the gray horizontal line.

Table S11: Empirical size of the directional test (DT), likelihood ratio test (LRT), and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (9) in the paper with $p = n_i^\tau$ and $n_i = 1000$ and $g = 5$, at nominal level $\alpha = 0.05$

τ (p)	DT	LRT	Sko1	Sko2
0.250 (6)	0.053	0.055	0.053	0.053
0.333 (10)	0.055	0.058	0.055	0.055
0.417 (18)	0.048	0.054	0.048	0.048
0.500 (32)	0.048	0.061	0.048	0.048
0.583 (57)	0.051	0.089	0.052	0.052
0.667 (100)	0.054	0.160	0.055	0.054
0.750 (178)	0.053	0.462	0.058	0.055
0.833 (317)	0.050	0.983	0.083	0.063

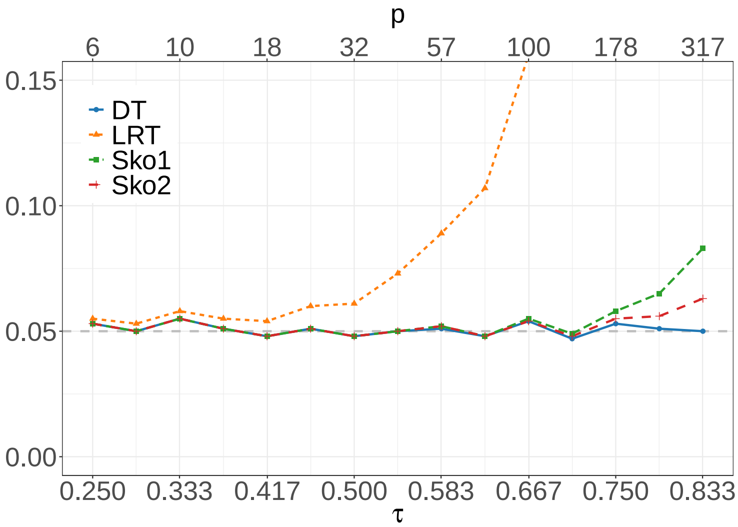

Figure S8: Empirical size of the directional test (DT), likelihood ratio test (LRT), and two Skovgaard's modifications (Sko1 and Sko2) for hypothesis (9) in the paper, with $p = n_i^\tau$ and $n_i = 1000$ and $g = 5$, at nominal level $\alpha = 0.05$ given by the gray horizontal line.

References

He, Y., Meng, B., Zeng, Z., and Xu, G. (2021). On the phase transition of Wilks' phenomenon. Biometrika, 108:741–748.

Nel, D. and Merwe, C. V. D. (1986). A solution to the multivariate Behrens-Fisher problem. Communications in Statistics - Theory and Methods, 15:3719–3735.

Skovgaard, I. (2001). Likelihood asymptotics. Scandinavian Journal of Statistics, 28:3–32.