Comparison of exact parametric tests for high-dimensional data

S. Kropf, J. Läuter, D. Kose, and D. von Rosen

Research Report
Centre of Biostochastics

Swedish University of
Agricultural Sciences

Report 2007:3
ISSN 1651-8543

Comparison of exact parametric tests for high-dimensional data

Siegfried Kropf, Jürgen Läuter, Daniela Kose
Otto-von-Guericke University Magdeburg, Leipziger Strasse 44, 39120 Magdeburg, Germany

Dietrich von Rosen[1]
Centre of Biostochastics, Swedish University of Agricultural Sciences, Box 7013, 750 07 Uppsala, Sweden

Abstract

In the paper several proposals for exact or at least conservative parametric multivariate tests in the general linear model are considered that are also applicable for high-dimensional data, where the dimension of the observations may exceed the sample size. The common feature is the inclusion of principal component transformations into the test. Whereas a test proposal by Srivastava and von Rosen originally assumes that the multivariate data have a known reduced rank which is used in the construction of the test, several versions of so-called PC tests by Läuter and colleagues accept a reduction of variance and utilize it for a "stabilization" of the test in terms of power. The different tests are compared with respect to their philosophy as well as their performance in two real data examples and in simulation studies. It is shown that the test of Srivastava and von Rosen is a conservative test, even when a rank is assumed to be smaller than the true one. It is, however, less conservative than the conservative version of the PC test derived for the construction of convenient confidence region for the investigated effects. The exact PC tests turned out to have the largest power.

Keywords: parametric tests, high-dimensional data, principal components

[1]E-mail address to the correspondence author: Dietrich.von.Rosen@bt.slu.se

1 Introduction

Classical parametrical multivariate tests such as Wilks' Λ or Hotelling's trace statistic require sample sizes that are larger than the number of variables – an assumption which is often not met in real data. To overcome this problem of high-dimensional analysis, O'Brien (1984) or Lachin (1992) among others presented tests based on condensed constructs (scores) of the high-dimensional data. However, these tests are only asymptotic level α tests. To our knowledge, so far there are only very few proposals for exact parametric tests in this situation. We will consider and discuss several of them based on some kind of principal component decomposition.

A very different approach is used in Kropf et al. (2007), where the analysis is based on pairwise distances of sample elements (as known from cluster analysis) and the tests are carried out as so-called rotation tests (Landsrud, 2005; Läuter et al. 2005). These tests are still restricted to simple test situations and are therefore not considered here in the comparison of methods.

We will discuss the assumptions, the motivation and the performance of three proposals for tests in the general linear model with multivariate normal data given in papers by (i) Srivastava and von Rosen (2002, 2004), (ii) Läuter (1996) and Läuter et al. (1996, 1998) and (iii) Läuter and Glimm (2005).

After the introduction of the mathematical model in Section 2 and two examples in Section 3, the three methods are explained in detail in Sections 4, 5, and 6, respectively. An additional result concerning the proposal of Srivastava and von Rosen in the case of a misspecified rank is given in Section 7 where the test statistics are compared. The results obtained so far are illustrated in some simulation experiments in Section 8. The discussion of pros and cons of the methods in Section 9 completes the paper.

2 The model

We consider a general linear model with fixed effects, where the p-dimensional sample observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ are dependent on the design matrix $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ of size $n \times s$, the parameter matrix $\mathbf{B} = (\beta_{ij})_{s\times p}$, and i.i.d. multivariate normal noise vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$:

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{y}_1' \\ \vdots \\ \boldsymbol{y}_n' \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_n' \end{pmatrix} \mathbf{B} + \begin{pmatrix} \boldsymbol{e}_1' \\ \vdots \\ \boldsymbol{e}_n' \end{pmatrix} = \boldsymbol{X}\mathbf{B} + \boldsymbol{E}, \quad \boldsymbol{E} \sim \mathrm{N}_{n\times p}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}), \quad (1)$$

where $\boldsymbol{I}_n$ denotes the identity matrix of size n. For convenience of notation we assume that the design matrix $\boldsymbol{X}$ is of full rank s ($s < n$), which can usually

be achieved by a suitable reparametrization. The covariance matrix $\mathbf{\Sigma}$ of the noise vectors is positive semidefinite. It may be singular in all procedures considered here later on and in one of them it is even assumed to be singular. The null hypothesis of interest is directed to the parameter matrix $\mathbf{B}$:

$$\mathrm{H}_0 : \boldsymbol{C}\mathbf{B} = \mathbf{0}\,,$$

where the $m \times s$-matrix $\boldsymbol{C}$ $(m < s)$ has to be specified with a full rank m.

In the classical multivariate analysis with positive definite $\mathbf{\Sigma}$ and sufficiently large sample size n one would estimate the parameter matrix by

$$\hat{\mathbf{B}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\,.$$

For the test of the multivariate null hypothesis H_0 one could determine the sums of squares and cross products matrix for the residual errors

$$\begin{aligned} \boldsymbol{G} &= (\boldsymbol{Y} - \boldsymbol{X}\hat{\mathbf{B}})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\mathbf{B}}) \\ &= \left(\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\right)' \\ &= \boldsymbol{Y}'\boldsymbol{Q}_G\boldsymbol{Y} \quad \text{with} \quad \boldsymbol{Q}_G = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}', \end{aligned} \tag{2}$$

and that for the deviations from the null hypothesis

$$\begin{aligned} \boldsymbol{H} &= \hat{\mathbf{B}}'\boldsymbol{C}'\left(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\right)^{-1}\boldsymbol{C}\hat{\mathbf{B}} \\ &= \boldsymbol{Y}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\left(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\right)^{-1}\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \\ &= \boldsymbol{Y}'\boldsymbol{Q}_H\boldsymbol{Y}, \end{aligned} \tag{3}$$

with $\boldsymbol{Q}_H = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\left(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\right)^{-1}\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, and then use one of classical multivariate test statistics such as Wilks' Λ criterion

$$\Lambda = \frac{|\boldsymbol{G}|}{|\boldsymbol{H} + \boldsymbol{G}|} = \frac{\Pi_{i=1}^{p} l_{\boldsymbol{G}}^{(i)}}{\Pi_{i=1}^{p} l_{\boldsymbol{W}}^{(i)}}, \tag{4}$$

where $l_{\boldsymbol{G}}^{(i)}$ and $l_{\boldsymbol{W}}^{(i)}$ $(i = 1, \ldots, p)$ are the eigenvalues of $\boldsymbol{G}$ and $\boldsymbol{W} = \boldsymbol{H} + \boldsymbol{G}$, respectively.

The null hypothesis is rejected if $\Lambda < U_{p,m,f,\alpha}$, where the quantiles $U_{p,m,f,\alpha}$ depend on the number p of variables, the rank m of the contrast matrix $\boldsymbol{C}$, the remaining degrees of freedom $f = n - s$ for the residuals, and the significance level α of the test. Tables for the quantiles are given, e.g. in Pillai and Gupta (1969). In practice one often uses the F approximation

$$F = \frac{uv - 2\gamma}{pm}\,\frac{1 - \Lambda^{1/u}}{\Lambda^{1/u}} \quad \sim \quad F_{pm,uv-2\gamma} \tag{5}$$

with

$$u = \sqrt{\frac{p^2m^2 - 4}{p^2 + m^2 - 5}}\,, \quad v = n - s - \frac{p - m + 1}{2}\,, \quad \gamma = \frac{pm - 2}{4}\,. \tag{6}$$

For $m = 1$ and $m = 2$ these transformations even yield exact F distributions with the given degrees of freedom.

Here our main interest is in situations where the sample size is smaller than the number of variables such that the matrices $\boldsymbol{G}$ and $\boldsymbol{H}$ are singular and the Wilks' test is no longer applicable.

3 Data sets

To illustrate the methods two real data sets will be used. The first one is from Russell et al. (1967), presented in Srivastava and Carter (1983, p.128) and Srivastava and von Rosen (2004) and considers soil characteristics. In a simplified presentation, eight soil samples have been taken from each of two blocks of an area of interest. The samples are taken at four different conditions (two samples per condition and plot) representing a combination of depths and contours. Eight characteristics have been determined for each sample (cf. Table 1). Thus we have $n = 16$ independent observations, $p = 8$ response variables, and two factors ("Block" with 2 levels, and "Condition" with 4 levels) in a complete balanced block design, and we are going to test that there is no block effect in the example.

If the 16 independent observation vectors are arranged in the matrix $\boldsymbol{Y}$ in the same order as in Table 1, then the design matrix $\boldsymbol{X}$ can be set up as

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \end{pmatrix}', \tag{7}$$

where the first column of $\boldsymbol{X}$ corresponds to the total mean (β_1), the second one to the "block" effect with two levels (β_2), and columns 3 to 5 to the "condition" effect with four levels ($\beta_3, \beta_4, \beta_5$). Thus the matrix $\boldsymbol{C}$ consists of one row with the indicator for the parameter β_2,

$$\boldsymbol{C} = (\,0 \quad 1 \quad 0 \quad 0 \quad 0\,)\,. \tag{8}$$

In this example the sample size would be large enough for the application of the classical tests. The variables are, however, highly correlated such that the

Table 1: Soil data example with 16 soil samples from two blocks of an experiment area and at four different conditions. The variables and units are as follows: V1 = pH; V2 = total nitrogen (%); V3 = bulk density (gm/cm^3); V4 = total phosphorus (ppm); V5 = exchangeable + soluble calcium (me/100 gm); V6 = exchangeable + soluble magnesium (me/100 gm); V7 = exchangeable + soluble potassium (me/100 gm); V8 = exchangeable + soluble sodium (me/100 gm).

Block	Condition	V1	V2	V3	V4	V5	V6	V7	V8
1	1	5.40	0.188	0.92	215	16.35	7.65	0.72	1.14
1	1	5.65	0.165	1.04	208	12.25	5.15	0.71	0.94
2	1	5.14	0.260	0.95	300	13.02	5.68	0.68	0.60
2	1	5.14	0.169	1.10	248	11.92	7.88	1.09	1.01
1	2	5.14	0.164	1.12	174	14.17	8.12	0.70	2.17
1	2	5.10	0.094	1.22	129	8.55	6.92	0.81	2.67
2	2	4.70	0.100	1.52	117	8.74	8.16	0.39	3.32
2	2	4.46	0.112	1.47	170	9.49	9.16	0.70	3.76
1	3	4.37	0.112	1.07	121	8.85	10.35	0.74	5.74
1	3	4.39	0.058	1.54	115	4.73	6.91	0.77	5.85
2	3	4.17	0.078	1.26	112	6.29	7.95	0.26	5.30
2	3	3.89	0.070	1.42	117	6.61	9.76	0.41	8.30
1	4	3.88	0.077	1.25	127	6.41	10.96	0.56	9.67
1	4	4.07	0.046	1.54	91	3.82	6.61	0.50	7.67
2	4	3.88	0.055	1.53	91	4.98	8.00	0.23	8.78
2	4	3.74	0.053	1.40	79	5.86	10.14	0.41	11.04

covariance matrix is nearly singular. The eight original variables might – in good approximation – be represented by a smaller number of components.

The second example considers data from the field of occupational medicine. Workers of a metallurgic company (exposed to lead vapour, $n_1 = 26$), car painters (exposed to organic solvents, $n_2 = 45$) and workers with similar jobs but without special exposition (control group, $n_3 = 48$) were investigated with a large scale of physiological, psychological and other tests to detect if a long-termed exposition at a low level induces disturbances (Böckelmann et al. 1998). We neglect group 2 here, such that 74 workers remain in the analysis, and we focus on the data of a psychophysiological investigation with altogether 84 variables describing the performance, physiological data and a subjective assessment of the stress in seven stages of a computer-guided session where the proband had to solve several tasks. Some of the variables were log-transformed to get approximately normal data, missing values were imputed by a regression-like estimator (Kropf, 2000, pp. 157-159). All workers

were adult males (age 25-60 years). The main interest is the *multivariate* comparison of the two groups considered here, so that many small effects in the single parameters may add up in the test. However, traditional multivariate tests are not possible here as $p > n$. The age differed a bit between the two groups and many of the considered parameters are age-dependent to some degree. Therefore, age is included as a covariable.

The design matrix has 74 rows representing the workers of the two groups and three columns corresponding to the expectation in group 1 (β_1), the expectation in group 2 (β_2) and the covariate age (β_3):

$$\boldsymbol{X} = \begin{pmatrix} 1 & 0 & a_1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & a_{26} \\ \cdots & \cdots & \cdots \\ 0 & 1 & a_{27} \\ \vdots & \vdots & \vdots \\ 0 & 1 & a_{74} \end{pmatrix}', \tag{9}$$

where $a_1, \ldots, a_{74}$ denotes the age of the 74 workers. To compare the two groups, we choose

$$\boldsymbol{C} = (\, 1 \quad -1 \quad 0 \,) \,. \tag{10}$$

4 MANOVA with singular covariance matrix

Srivastava and von Rosen (2004) consider the situation where the covariance matrix of the p dimensional residuals has a reduced rank r ($r < p$) which is assumed to be known and to be not larger than the remaining degrees of freedom f. The main aim of their paper was to estimate the parameters $\mathbf{B}$ and $\mathbf{\Sigma}$ but they also proposed a test similar to Wilks' Λ. The authors use a principal component decomposition of the unknown true covariance matrix $\mathbf{\Sigma}$, $\mathbf{\Sigma} = \mathbf{\Gamma\Lambda\Gamma'}$, where $\mathbf{\Gamma}$ is the diagonal matrix of the r positive eigenvalues of $\mathbf{\Sigma}$ and $\mathbf{\Gamma}$ is a $p \times r$-matrix with the corresponding m eigenvectors as columns. Furthermore, a $p \times (p-r)$-matrix $\mathbf{\Gamma}_0$ is chosen such that $(\mathbf{\Gamma}, \mathbf{\Gamma}_0)$ is an orthogonal matrix. This matrix is used to decompose the model (1) into one part containing the "independent" information (random part) of the p variables and a second one presenting the linearly depending portion (dependent part):

$$\boldsymbol{Y}(\mathbf{\Gamma}, \mathbf{\Gamma}_0) = \boldsymbol{X}\mathbf{B}(\mathbf{\Gamma}, \mathbf{\Gamma}_0) + \boldsymbol{E}(\mathbf{\Gamma}, \mathbf{\Gamma}_0) \,.$$

Based on this decomposition, they derive "maximum likelihood type" estimators of the parameters and prove that the "rank reduced" Wilks' statistic

$$\Lambda_{SRr} = \frac{|\hat{\mathbf{\Gamma}}' \boldsymbol{G} \hat{\mathbf{\Gamma}}|}{|\hat{\mathbf{\Gamma}}' (\boldsymbol{H} + \boldsymbol{G}) \hat{\mathbf{\Gamma}}|} \,. \tag{11}$$

can, due to the rank restriction of the data, be computed with the r non-zero eigenvalues of the "original" matrices $\boldsymbol{G}$ and $\boldsymbol{W} = \boldsymbol{H} + \boldsymbol{G}$. Now (11) equals

$$\Lambda_{SRr} = \frac{l_{\boldsymbol{G}}^{(1)} \cdots l_{\boldsymbol{G}}^{(r)}}{l_{\boldsymbol{W}}^{(1)} \cdots l_{\boldsymbol{W}}^{(r)}} \,, \tag{12}$$

and (4) implies that (12) also has the distribution of a Wilks' statistic but with reduced dimension parameter, i.e. with r instead of p, such that the null hypothesis can be rejected if

$$\Lambda_{SRr} < U_{r,m,f,\alpha} \,. \tag{13}$$

Thus the procedure is fairly similar to the non-singular case. One has to calculate the same sums of squares and cross products matrices and to determine their eigenvalues. Only the products in the final statistic are reduced and the reference distribution is changed in one parameter.

A problem in this approach is that the rank of the covariance matrix $\mathbf{\Sigma}$ should be known and that this "real dimension" has to be smaller than the remaining degrees of freedom. But with these restrictions, the formal limitation that the number of variables may not exceed the degrees of freedom is no longer present. In practice one would expect that the empirical covariance matrix usually has the full rank as determined by the minimum of number of variables and degrees of freedom due to some noise in the data. But only the first r eigenvalues should substantially deviate from zero. This discussion is continued in Section 7. It should be noted here, however, that – in contrast to the 'original' Wilks' Λ – the statistic Λ_{SRr} is no longer scale invariant if such noise is present.

Considering the first example, inserting the data matrix $\boldsymbol{Y}$ from Table 1, the design matrix $\boldsymbol{X}$ as given in (7) and the matrix $\boldsymbol{C}$ from (8) into (2) and (3) gives

Table 2: Results of the Srivastava-von Rosen test to detect block effects in the soil data example from Table 1 for varying assumptions of the rank r of the error covariance matrix.

Eigenvalues of $\boldsymbol{G}$	8925	54.75	16.20	4.27	0.17	0.14	0.029	0.001
Eigenvalues of $\boldsymbol{W}$	9105	55.68	24.15	4.40	0.53	0.15	0.035	0.001
Reduced rank r	1	2	3	4	5	6	7	8
Reduced Wilks' Λ	0.98	0.96	0.65	0.63	0.21	0.19	0.16	0.16
F transformations	0.22	0.19	1.64	1.19	5.37	4.16	3.76	2.72
p-values	0.646	0.831	0.248	0.386	0.024	0.053	0.082	0.174

$$\boldsymbol{G} = \begin{pmatrix} 0.20 & -0.00 & -0.01 & -14.51 & -0.38 & -1.42 & -0.09 & -0.80 \\ -0.00 & 0.01 & -0.04 & 7.85 & 0.43 & 0.11 & -0.00 & -0.07 \\ -0.01 & -0.04 & 0.26 & -18.14 & -2.32 & -1.33 & 0.02 & -0.03 \\ -14.51 & 7.85 & -18.14 & 8920.25 & 168.97 & 78.65 & 18.74 & -57.99 \\ -0.38 & 0.43 & -2.32 & 168.97 & 41.74 & 20.45 & -0.73 & 3.95 \\ -1.42 & 0.11 & -1.33 & 78.66 & 20.45 & 26.62 & 0.97 & 10.09 \\ -0.09 & -0.00 & 0.02 & 18.74 & -0.73 & 0.97 & 0.35 & 0.24 \\ -0.80 & -0.07 & -0.03 & -57.99 & 3.95 & 10.09 & 0.24 & 10.79 \end{pmatrix}$$

and

$$\boldsymbol{H} = \begin{pmatrix} 0.52 & 0.00 & -0.17 & -9.72 & 1.48 & -0.73 & 0.24 & -1.13 \\ 0.00 & 0.00 & -0.00 & -0.02 & 0.00 & -0.00 & 0.00 & -0.00 \\ -0.17 & -0.00 & 0.06 & 3.21 & -0.49 & 0.24 & -0.08 & 0.37 \\ -9.72 & -0.02 & 3.21 & 182.25 & -27.74 & 13.70 & -4.52 & 21.13 \\ 1.48 & 0.00 & -0.49 & -27.74 & 4.22 & -2.09 & 0.69 & -3.22 \\ -0.73 & -0.00 & 0.24 & 13.70 & -2.09 & 1.03 & -0.34 & 1.59 \\ 0.24 & 0.00 & -0.08 & -4.52 & 0.69 & -0.34 & 0.11 & -0.52 \\ -1.13 & -0.00 & 0.37 & 21.13 & -3.22 & 1.59 & -0.52 & 2.45 \end{pmatrix}.$$

Table 2 shows the eight eigenvalues of the matrices $\boldsymbol{G}$ and $\boldsymbol{W} = \boldsymbol{H} + \boldsymbol{G}$, the Wilks' statistics (11) as cumulative products from these eigenvalues, and the corresponding F- and p-values according to (12)/(13) with the transformation (5)/(6).

So, if we assume a full rank of 8 then no block effect is seen. The eigenvalues of the matrix $\boldsymbol{G}$ of residuals indicates, however, that the "real" rank might be in the range of 4 to 7. Depending on that choice, the p-value differs considerably, also including values below 0.05. It is interesting to note that the results change dramatically if, e.g., variable V1 is multiplied with the factor 1000 in all samples. Then all p-values for $r \leq 6$ give significant block effects.

In the second example, we have 84 variables and a total sample size of 74 in both groups. The matrices $\boldsymbol{G}$ and $\boldsymbol{W}$ have the reduced rank 71 and 72, respectively. As Figure 1 shows, there are several small jumps in the decreasing eigenvalues of $\boldsymbol{G}$ giving only little hint for the choice of the rank r. One might choose $r = 9$, where the first small gap appears in the dotted curve. The corresponding p-values are smaller than 0.05 in the range from $r = 3$ to $r = 30$ with a minimal p-value of 0.0059, indeed for $r = 9$.

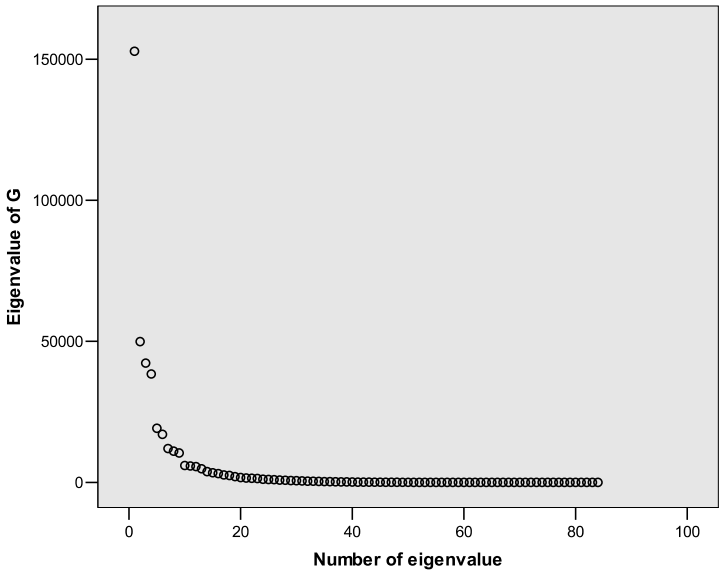

Figure 1: Scree-plot: the eigenvalues of the matrix $\boldsymbol{G}$ in the second example.

5 Principal component test (PC test)

Läuter (1996) and Läuter et al. (1996, 1998) consider a slightly different situation. They also assume that a considerable proportion of variation of the p variables can be expressed by a smaller number q of scores ($1 \leq q \leq \min(p, n - s)$) corresponding to latent variables hidden in the observed data. Therefore, they transform the p-dimensional data vectors $\boldsymbol{y}_i$ ($1 \leq i \leq n$) into q-dimensional scores $\boldsymbol{z}_i = \boldsymbol{D}'\boldsymbol{y}_i$ with a data-dependent weight matrix $\boldsymbol{D}$ of size $p \times q$ in a first step and then carry out univariate or multivariate tests at that reduced dimension q. The transformation can be included in the test statistic by replacing the matrices $\boldsymbol{G}$, $\boldsymbol{H}$ and $\boldsymbol{W}$ by the corresponding matrices $\boldsymbol{D}'\boldsymbol{G}\boldsymbol{D}$,

$\boldsymbol{D}'\boldsymbol{H}\boldsymbol{D}$ and $\boldsymbol{D}'\boldsymbol{W}\boldsymbol{D}$, respectively, and by changing the dimension parameter p into q. Theorems given in the mentioned papers ensure that the tests with the score vectors exactly keep the error of first kind as long as the weight matrix $\boldsymbol{D}$ is dependent on the data only through the total sum of squares and cross products matrix $\boldsymbol{W}$ and the matrix of score vectors $\boldsymbol{Z} = (\,\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\,)' = \boldsymbol{Y}\boldsymbol{D}$ has rank q with probability 1.

Particularly, we consider the PC test here, where the q columns of the weight matrix $\boldsymbol{D}$ are the eigenvectors of the matrix $\boldsymbol{W}$ corresponding to its q largest eigenvalues, i.e., $\boldsymbol{D}$ is a solution of the eigenvalue equation

$$\boldsymbol{W}\boldsymbol{D} = \boldsymbol{D}\boldsymbol{\Lambda}\,, \quad \boldsymbol{D}'\boldsymbol{D} = \boldsymbol{I}_q \tag{14}$$

with the diagonal matrix $\boldsymbol{\Lambda}$ containing the q largest eigenvalues. This yields a scale dependent test. In the above papers, mainly a scale invariant version is considered using the eigenvalue equation

$$\boldsymbol{W}\boldsymbol{D} = \mathrm{Diag}(\boldsymbol{W})\boldsymbol{D}\boldsymbol{\Lambda}\,, \quad \boldsymbol{D}'\mathrm{Diag}(\boldsymbol{W})\boldsymbol{D} = \boldsymbol{I}_q\,. \tag{15}$$

We will use both versions here, the scale dependent version in order to be comparable with the method of Srivastava and von Rosen and the conservative version of the next section, and the scale invariant version as it might be more appropriate to the examples with variables of rather different scales.

As final test statistic one can again use the Wilks' statistic (4) but now with the transformed sums of squares and cross products matrices

$$\Lambda_{LGKq} = \frac{|\boldsymbol{D}'\boldsymbol{G}\boldsymbol{D}|}{|\boldsymbol{D}'(\boldsymbol{H}+\boldsymbol{G})\boldsymbol{D}|} = \frac{l^{(1)}_{\boldsymbol{D}'\boldsymbol{G}\boldsymbol{D}} \cdots l^{(q)}_{\boldsymbol{D}'\boldsymbol{G}\boldsymbol{D}}}{l^{(1)}_{\boldsymbol{D}'\boldsymbol{W}\boldsymbol{D}} \cdots l^{(q)}_{\boldsymbol{D}'\boldsymbol{W}\boldsymbol{D}}}\,, \tag{16}$$

where $l^{(i)}_{\boldsymbol{D}'\boldsymbol{G}\boldsymbol{D}}$ and $l^{(i)}_{\boldsymbol{D}'\boldsymbol{W}\boldsymbol{D}}$ denote the eigenvalues of the matrices given as suffix. Λ_{LGKq} has to be compared to the dimension reduced quantile $U_{q,m,f,\alpha}$, i.e.,

$$\Lambda_{LGKq} < U_{q,m,f,\alpha}\,, \tag{17}$$

for significance. Both Λ_{SRq} and Λ_{LGKq} use transformations of the original sums of squares and cross products matrices, as seen from (11) and (16). But in (11) a reduced rank r of the data is assumed, leading to the effect that each estimator $\hat{\boldsymbol{\Gamma}}$ of full rank r delivers the same value of the final test statistic with the transformed sums of squares and cross products matrices. In contrast, with the PC test we accept that there exists variance in the data beyond the effect of the suspected q latent variables. As a consequence, the weight matrices are not all restricted to the same q-dimensional subspace and do have influence onto the statistic.

Table 3: Results of the scale dependent and scale invariant versions of the PC test to detect block effects in the soil data from Table 1 for different score dimensions q.

Score dimension q	1	2	3	4	5	6	7	8
p-values, scale dep.	.645	.813	.238	.374	.019	.049	.082	.174
p-values, scale invar.	.531	.028	.008	.011	.025	.044	.102	.174

Thus, the score dimension q of the PC test is not restricted to the real rank of the covariance matrix. One even wants to drop ineffective components to gain a high power of the test. The choice of an efficient dimension q can be derived from the data. The proof of the type I error control covers also the situation where q is calculated as function of the matrix $\boldsymbol{W}$. In practice, one uses methods known from factor analysis, particularly in the scale invariant version. A convenient proposal is the Kaiser criterion, where q is chosen as the number of eigenvalues of $\boldsymbol{W}$ exceeding 1.

The results with both test versions for the first example are given in Table 3. The p-values for the scale dependent version in the upper line are similar to those of the Srivastava-von Rosen test. The corresponding eigenvalues of (14) are 9105.9, 55.7, 24.1, 4.40, 0.529, 0.149, 0.035, and 0.001, giving no clear hints for the choice of q because of the very different scales of the variables. In the scale invariant version (15), the eigenvalues are 2.89, 2.19, 1.28, 1.09, 0.24, 0.21, 0.05, and 0.04. Following the Kaiser criterion, the use of four-dimensional scores would be recommended connected with the p-value 0.011 indicating significance.

In the second example, the results for the scale dependent PC test are very similar to the Srivastava-von Rosen test. The p-values are smaller than the threshold 0.05 for q-values in the range from 3 to 31 with a minimal p-value of 0.0017 for $q = 7$. In the scale invariant version, the Kaiser criterion, e.g., would indicate the use of 17 scores ($q = 17$). The final p-values are smaller than 0.05 for the q-values 3, 5, 7 to 45 with the minimum of 0.00057 for $q = 25$.

6 Conservative principal component test

It is an advantage of the PC test that the explicit determination of the weights can also be used to interpret the q scores through their weights or their correlations to the original variables. Just in this setting it would be useful to have confidence regions for the investigated effects, preferably in terms of the derived scores. Basically, confidence regions for the investigated effects $\boldsymbol{C}\mathbf{B}$ can

be derived by an inversion of the PC test (16)/(17) with the weight matrix $\boldsymbol{D}$ from (14) or (15). To do this, we consider shifted null hypotheses $\boldsymbol{C}\mathbf{B} = \mathbf{M}$ by replacing the term $\boldsymbol{C}\hat{\mathbf{B}}$ in the calculation of the matrix $\boldsymbol{H}$ in (3) by the shifted term $\boldsymbol{C}\hat{\mathbf{B}} - \mathbf{M}$. The confidence region for $\boldsymbol{C}\mathbf{B}$ then contains all matrices $\mathbf{M}$ for which the modified test accepts the shifted null hypothesis. Unfortunately, the dependence of the test statistic on the shift $\mathbf{M}$ is sophisticated as each shift already yields modified matrices $\boldsymbol{H}$ and $\boldsymbol{W}$ and thus modified weights $\boldsymbol{D}$. Therefore, even in a very simple test problem as the comparison of two independent samples, the resulting confidence regions have rather irregular and unhandy shapes (Kropf, 2000).

This was the starting for Läuter and Kropf (2002) and in a more general form Läuter and Glimm (2005) to derive a conservative modification of the PC test. It has been shown there that a scale dependent version of the PC test can also be derived with the weight matrix $\tilde{\boldsymbol{D}}$ composed from the eigenvectors of the eigenvalue problem

$$\boldsymbol{G}\tilde{\boldsymbol{D}} = \tilde{\boldsymbol{D}}\tilde{\boldsymbol{\Lambda}}\,, \quad \tilde{\boldsymbol{D}}'\tilde{\boldsymbol{D}} = \boldsymbol{I}_q\,. \tag{18}$$

When the final test is carried out with Wilks' Λ statistic,

$$\Lambda_{LGq} = \frac{|\tilde{\boldsymbol{D}}'\boldsymbol{G}\tilde{\boldsymbol{D}}|}{|\tilde{\boldsymbol{D}}'(\boldsymbol{H}+\boldsymbol{G})\tilde{\boldsymbol{D}}|} = \frac{l^{(1)}_{\tilde{\boldsymbol{D}}'\boldsymbol{G}\tilde{\boldsymbol{D}}} \cdot \ldots \cdot l^{(q)}_{\tilde{\boldsymbol{D}}'\boldsymbol{G}\tilde{\boldsymbol{D}}}}{l^{(1)}_{\tilde{\boldsymbol{D}}'\boldsymbol{W}\tilde{\boldsymbol{D}}} \cdot \ldots \cdot l^{(q)}_{\tilde{\boldsymbol{D}}'\boldsymbol{W}\tilde{\boldsymbol{D}}}}\,, \tag{19}$$

where again $l^{(i)}_{\tilde{\boldsymbol{D}}'\boldsymbol{G}\tilde{\boldsymbol{D}}}$ and $l^{(i)}_{\tilde{\boldsymbol{D}}'\boldsymbol{W}\tilde{\boldsymbol{D}}}$ denote the eigenvalues of the matrices given as suffix, then the test statistic is always larger than that of the scale dependent PC test above. Hence the test is conservative if the test rejects the null hypothesis when

$$\Lambda_{LGq} < U_{q,m,f,\alpha}. \tag{20}$$

The advantage is that the weight matrix $\tilde{\boldsymbol{D}}$ then no longer depends on the shift in the above modified test problem. As a consequence, the confidence regions become much more convenient. In the special case $s = 1$ and $q = 1$, e.g., the confidence region is limited by two parallel $(p-1)$-dimensional hyperplanes of the p-dimensional space.

The resulting test is very close to the Srivastava-von Rosen test insofar as the projections of the sums of squares and cross product matrices with the weight matrix $\tilde{\boldsymbol{D}}$ in (19) are identical to those with $\hat{\boldsymbol{\Gamma}}$ in (11). The difference is that the projection with $\hat{\boldsymbol{\Gamma}}$ is omitted in the final statistic (12) due to the supposed rank restriction, whereas it is maintained in (19), thus accepting the additional noise.

Table 4: Results of the conservative versions of the PC test to detect block effects in the soil data from Table 1 for different score dimensions q.

Score dimension q	1	2	3	4	5	6	7	8
p-values, conserv. test	.646	.846	.260	.398	.050	.058	.082	.174

For the first example with the soil data, the eigenvalues have already been given in Table 2. With this scale dependent version, it is difficult to decide about a suitable choice of q. One might think about values in the range from 1 to 4. According to the results given in Table 4, however, only for $q = 5$, we have a p-value slightly below the 5 %-threshold. Thus we fail to find a significant result.

In the second example, the p-values are below the significance threshold for q in the ranges 2 to 17 and 24 to 29; the minimum is attained for $q = 0.011$. As seen in Figure 1, just $q = 9$ seems to be a well acceptable choice. So we have found a significant result, but compared to the other test versions it is the "least significant" one.

7 Comparison of the different test statistics

As discussed in Section 3, the proposal of Srivastava and von Rosen is designed for the case that the covariance matrix has a reduced and known rank r ($r < \min(p, f+1)$). If this is really the case, then the test will be exact and so it is identical with the scale dependent PC test as well as with the conservative PC test when these use the same score dimension $q = r$. The latter follows from the fact, that with the rank the columns of the theoretical weight matrix $\hat{\mathbf{\Gamma}}$ in (11) and those of $\boldsymbol{D}$ in (16) and of $\tilde{\boldsymbol{D}}$ in (19) span the same r-dimensional subspace with probability 1.

But in practice, the Srivastava-von Rosen test will also be applied in situations with a small amount of additional noise. So the question arises how the test based on statistic (12) behaves in these noisy situations. Here we will show that the same techniques as used in Läuter and Glimm (2005) for the proof that the PC test based on Λ_{LGq} always yields a conservative test can be applied to characterize the behaviour of the test (12)/(13) in the case of an inaccurately specified rank. When the true rank r_0 is larger than the specified rank, then the two representations (11) and (12) of the Wilks' statistic are no longer identical. Instead, the statistic (12), which would be used for the

practical procedure, can be reformulated as

$$\Lambda_{SRr} = \frac{l_{\boldsymbol{G}}^{(1)} \cdots l_{\boldsymbol{G}}^{(r)}}{l_{\boldsymbol{W}}^{(1)} \cdots l_{\boldsymbol{W}}^{(r)}} = \frac{|\tilde{\boldsymbol{D}}' \boldsymbol{G} \tilde{\boldsymbol{D}}|}{|\boldsymbol{D}' \boldsymbol{W} \boldsymbol{D}|} ,$$

where $\boldsymbol{D}$ and $\tilde{\boldsymbol{D}}$ are the solutions of (14) and (18), respectively, each consisting of the r eigenvectors belonging to the r largest eigenvalues of $\boldsymbol{W}$ and $\boldsymbol{G}$, respectively. This can be seen from the decompositions $\boldsymbol{W} = \boldsymbol{D}_p \boldsymbol{\Lambda}_p \boldsymbol{D}_p'$ and $\boldsymbol{G} = \tilde{\boldsymbol{D}}_p \tilde{\boldsymbol{\Lambda}}_p \tilde{\boldsymbol{D}}_p'$, where $\boldsymbol{D}_p$ and $\tilde{\boldsymbol{D}}_p$ are the $p \times p$-matrices of all p eigenvectors of $\boldsymbol{W}$ and $\boldsymbol{G}$, respectively, and $\boldsymbol{\Lambda}_p$ and $\tilde{\boldsymbol{\Lambda}}_p$ are the corresponding diagonal matrices of eigenvalues, such that

$$|\boldsymbol{D}' \boldsymbol{W} \boldsymbol{D}| = \left| \boldsymbol{D}' \boldsymbol{D}_p \boldsymbol{\Lambda}_p \boldsymbol{D}_p' \boldsymbol{D} \right| = \left| \left(\boldsymbol{I}_q, \boldsymbol{0}_{q \times (p-q)} \right) \boldsymbol{\Lambda}_p \begin{pmatrix} \boldsymbol{I}_q \\ \boldsymbol{0}_{(p-q) \times q} \end{pmatrix} \right| = l_{\boldsymbol{W}}^{(1)} \cdots l_{\boldsymbol{W}}^{(q)}$$

and $\left| \tilde{\boldsymbol{D}}' \boldsymbol{G} \tilde{\boldsymbol{D}} \right| = l_{\boldsymbol{G}}^{(1)} \cdot \ldots \cdot l_{\boldsymbol{G}}^{(q)}$ analogously.

Now, basic results of matrix computations state that the matrix $\boldsymbol{E}$ consisting of the eigenvectors belonging to the r largest eigenvalues of a $p \times p$-matrix $\boldsymbol{M}$ maximizes the determinants $|\boldsymbol{Y}' \boldsymbol{M} \boldsymbol{Y}|$ for all $p \times q$-matrices $\boldsymbol{Y}$ with $\boldsymbol{Y}' \boldsymbol{Y} = \boldsymbol{I}_q$ (Rao, 1973). It follows that $\boldsymbol{D}' \boldsymbol{W} \boldsymbol{D} \geq \tilde{\boldsymbol{D}}' \boldsymbol{W} \tilde{\boldsymbol{D}}$ and $\tilde{\boldsymbol{D}}' \boldsymbol{G} \tilde{\boldsymbol{D}} \geq \boldsymbol{D}' \boldsymbol{G} \boldsymbol{D}$. Now remembering that $\Lambda_{LGKq} = |\boldsymbol{D}' \boldsymbol{G} \boldsymbol{D}| / |\boldsymbol{D}' \boldsymbol{W} \boldsymbol{D}|$ and $\Lambda_{LGq} = |\tilde{\boldsymbol{D}}' \boldsymbol{G} \tilde{\boldsymbol{D}}| / |\tilde{\boldsymbol{D}}' \boldsymbol{W} \tilde{\boldsymbol{D}}|$, we can state that

$$\Lambda_{LGr} \geq \Lambda_{SRr} \geq \Lambda_{LKGr} , \tag{21}$$

where Λ_{LKGr} refers to the scale dependent version of the PC test.

As all three test statistics use the same quantile for the test decision and we know that the PC test based on Λ_{LKGr} exactly keeps the pre-specified error level, the other two statistics yield conservative tests. Thus, the test of Srivastava and von Rosen is conservative if the rank restriction of the data is not met exactly, but it is less conservative than the PC test from Section 6. This opens the possibility to use the test (12)/(13) also in situations where the number of variables is larger than the sample size, even if there are no strict dependencies between the variables.

8 Simulation experiments

The results collected so far for a general linear model will now be demonstrated in some simulation experiments in the special case of a comparison of two independent samples $\boldsymbol{y}_{kl} \sim N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$, $l = 1, \ldots, n_k; k = 1, 2$, with respect to

their expectation ($H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$). In this situation, the two matrices $\boldsymbol{G}$ and $\boldsymbol{W}$ are given by $\boldsymbol{W} = \sum_{k=1}^{2} \sum_{l=1}^{n_k} (\boldsymbol{y}_{kl} - \bar{\boldsymbol{y}})(\boldsymbol{y}_{kl} - \bar{\boldsymbol{y}})'$ and $\boldsymbol{G} = \sum_{k=1}^{2} \sum_{l=1}^{n_k} (\boldsymbol{y}_{kl} - \bar{\boldsymbol{y}}_k)(\boldsymbol{y}_{kl} - \bar{\boldsymbol{y}}_k)'$, respectively, where $\bar{\boldsymbol{y}}_1$, $\bar{\boldsymbol{y}}_2$ and $\bar{\boldsymbol{y}}$ denote the group-wise and total means of the two samples. In order to restrict the computing time though reflecting the situation that the dimension of the observation is in a similar order of magnitude as the sample size, we use the dimension $p = 10$ and the sample sizes $n_1 = n_2 = 10$.

Data for the simulation experiments are generated in the following way:

- First two samples of 10 iid three-dimensional vectors $\boldsymbol{z}_{kl}$ ($k = 1, 2; l = 1, \ldots, 10$) are generated representing three independent latent variables following a standard normal distribution. The vectors $\boldsymbol{z}_{kl}$ are put underneath each other and form a matrix $\boldsymbol{Z}$ of size 20×3.

- In the first sample a shift $\delta = 1.5$ is added to the first ν latent variables, where ν is varying in the different simulation runs from 0 to 3.

- Then a transformation matrix $\boldsymbol{A}$ of size 3×10 and a noise matrix $\boldsymbol{U}$ of size 20×10 both with iid standard normal variables are generated. The 10 column vectors of $\boldsymbol{A}$ are divided by the root of their quadratic norm.

- The matrix $\boldsymbol{Y}$ of the 20×10-dimensional "observed" sample vectors is finally calculated as $\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{A} + \sigma\boldsymbol{U}$, where σ varies in the different runs: 0, 0.25, 0.5, 1, 2, 4.

Thus, the observed 10-dimensional data are essentially determined by three latent variables. But in all runs with $\sigma \neq 0$ a noise of varying intensity is added. With these data the four described tests (Srivastava-von Rosen, PC scale dependent, PC scale invariant, PC conservative) are carried out using the ranks/score dimensions $q = 1$ until $q = 5$ in parallel. The experiments used 1,000,000 repetitions per data configuration and the percentage of runs with rejections of the null hypothesis (at the nominal level $\alpha = 0.05$) is determined for each test version.

Figure 2 shows the results for the simulations under the null hypothesis. As can be seen, both versions of the exact PC test really keep the nominal error level in all situations. The Srivastava-von Rosen test and the conservative PC test are both exact in the situation for which the Srivastava-von Rosen test has been constructed, i.e. for the simulations without additional noise ($\sigma = 0$) and with the correct assumption of the rank ($q = 3$). In all other situations, these tests are conservative with increasing degree for increasing noise intensity. The deviations from the nominal level are smaller when the assumed rank/the number of scores used agrees with the number of latent variables used for

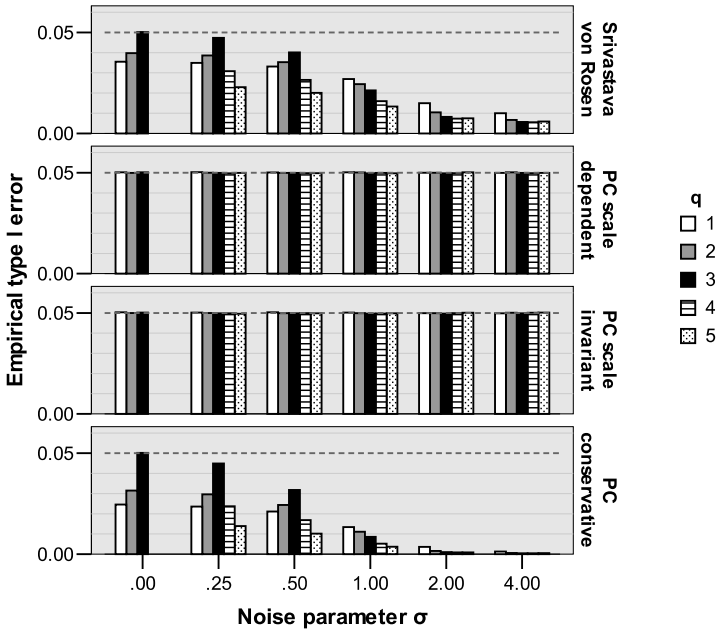

Figure 2: Results of the simulation experiments under the null hypothesis. The bars represent the empirical type I error of the test versions for different choices of the dimension parameter q (corresponding to r for the Srivastava-von Rosen test).

the generation of the observed data vectors. As predicted by the theory, the Sivastava-von Rosen test is less conservative than the conservative PC test. In both cases the degree of conservativeness may become considerable.

Results for the simulated data under a shift in two of the three variables are shown in Figure 3. One can see that the conservative behaviour of some of the test versions under the null hypothesis causes also a smaller power. All test versions, of course, loose power with increasing noise intensity. However, for each fixed value of the noise parameter and for each test dimension q, the scale dependent PC test has the largest power. The scale invariant version is only slightly inferior in these data with equal scale. So it should be used in data with variables of different scale. The Srivastava-von Rosen test has a distinctly smaller power and the conservative PC test has the worst performance. Again all test versions have the largest power if the dimension parameter is chosen

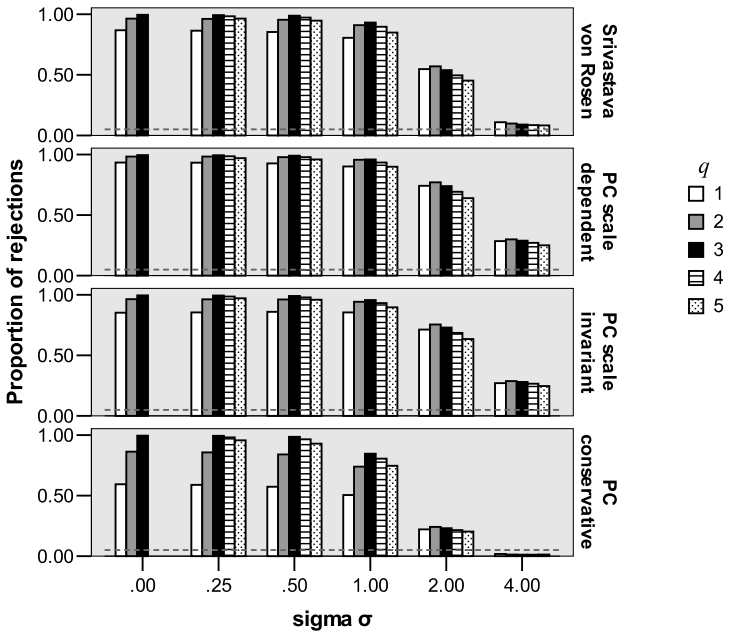

Figure 3: Results of the simulation experiments with a shift in two of the three latent variables. The bars represent the empirical rejection rates of the test versions for different choices of the dimension parameter q (corresponding to r for the Srivastava-von Rosen test).

properly. Only for high noise intensities a smaller q might be favourable which could be better seen in a rearranged graphic (not shown here).

If the constant shift is added only in one of the three latent variables then all test versions loose power. The other relations are the same as above but the power differences between test versions and parameter configurations are larger. Correspondingly, with the shift in all three latent variables, all test versions have a high power and so the differences are smaller but again the relations are the same. Therefore, these results are not given here in more detail.

9 Summary and discussion

We have considered several versions for tests in high-dimensional data and have compared them. All tests considered here are exact parametric tests or at least conservative ones. They consider transformations of the original data by principal component analysis. Whereas this transformation is performed explicitly in the proposals by Läuter and colleagues, it is only the background in the proposal by Srivastava and von Rosen and the transformation itself is omitted because this proposal is based on the assumption of a reduced rank of the residuals (combined with corresponding restrictions of the parameter space). Under this assumption, a reduction of dimension is no longer necessary. In applications, however, the situation with a reduced and known rank will be rare because even in cases with some known influential factors usually further noise will be included in the observed data.

The different assumptions should lead to different choices of the dimension parameters r or q. In the Srivastava-von Rosen test, the dimension r should be pre-specified. If (contrary to the theoretical basis) the rank is derived from the data, then the rank should be large enough to cover almost the whole variation in the data. In the PC test versions some loss of variance is included already in the concept and the data should be reduced to an efficient dimension. The simulation experiments underline that the largest power is attained when the dimension q is near to the number of "essential" input sources. Of course, this dimension is additionally limited by the sample size because the conventional multivariate test with the scores needs positive degrees of freedom. With the results presented above for the behaviour of the Srivastava-von Rosen test in case of an under-specification of the true rank, the situation changes. That test can now also be carried out neglecting a non-ignorable part of the variation and the simulation experiments show that this is also useful in terms of power.

Nevertheless, as the exact versions of the PC test also had the largest power, there is no reason to use one of the other tests as far as the significance statement for the global multivariate test is the main issue. As mentioned in Section 6, the construction of confidence regions for the investigated effect might be another aim of the analysis preferring the conservative PC test.

It should finally be mentioned that the results for the power of the tests would change if the investigated effects are (in contrast to our simulation scheme) assumed *not* to be connected to the main latent variables. This is, however, a rather artificial assumption.

References

[1] Böckelmann, I, Pfister, E., Stefanek, D., Gonschorek, A.S., Hinrichs, H., Hoy, L. and Kropf, S. (1998). Neurophysiologische verfahren zur objektivierung neurotoxischer früheffekten durch chronische blei- und lösemittelexposition, *Zbl Arbeitsmed* **9**, pp 358-367.

[2] Kropf, S. (2000). Confidence regions for stabilized multivariate tests, In: J.G. Bethlehem, P.G.M. van der Heijden (eds.), COMPSTAT. *Proceedings in Computational Statistics*, 14th Symposium held in Utrecht, The Netherlands, Physica-Verlag, Heidelberg New York, pp. 343-348.

[3] Kropf, S. (2000).*Hochdimensionale multivariate Verfahren in der medizinischen Statistik*, Shaker Verlag, Aachen.

[4] Kropf, S., Lux, A., Eszlinger, M., Heuer, H. and Smalla, K. (2007). Comparison of independent samples of high-dimensional data by pairwise distance measures. *Biometrical Journal* **49**, pp. 230-241.

[5] Lachin, J.M.L. (1992). Some large-sample distribution-free estimators and tests for multivariate partially incomplete data from two populations. *Statistics in Medicine* **11**, pp. 1151-1170.

[6] Langsrud, Ø. (2005). Rotation tests. *Statistics and Computing* **15**, pp. 53-60.

[7] Läuter, J. (1996). Exact t and F tests for analyzing studies with multiple endpoints. *Biometrics* **52**, pp. 964-970.

[8] Läuter, J. and Glimm, E. (2005). A theorem on the principal components inference, *Statistics* **39**, pp. 207-209.

[9] Läuter, J. and Kropf, S. (2002). Multivariate inference in clinical trials – tests with model choice and confidence regions. *Biocybernetics and Biomedical Engineering* **22**(1), pp. 5-16.

[10] Läuter, J., Glimm, E. and Kropf, S. (1996). New multivariate tests for data with an inherent structure. *Biometrical Journal* **38**, pp. 5-23. Erratum: *Biometrical Journal* **40**, (1996), p. 1015.

[11] Läuter, J., Glimm, E. and Kropf, S. (1998). Multivariate tests based on left-spherically distributed linear scores. *Annals of Statistics* **26**, pp. 1972-1988. Erratum: *Annals of Statistics* **27**, (1998), p. 1441.

[12] Läuter, J., Glimm, E. and Eszlinger, M. (2005). Search for relevant sets of variables in a high-dimensional setup keeping the familywise error rate. *Statistica Neerlandica* **59**, pp. 298-312.

[13] O'Brien, P.C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics* **40**, pp. 1079-1087.

[14] Pillai, K.C. and Gupta, A.K. (1969). On the exact distribution of Wilks's criterion. *Biometrika* **56**, pp. 109-118.

[15] Rao, C.R. (1973). *Linear Statistical Inference And Its Applications*, 2nd ed., Wiley, New York.

[16] Russell, J.S., Moore, A.W. and Coaldrake, J.E. (1967). Relationships between subtropical semiarid forest of Acacia Narpophylla, micro-relief and chemical properties of gilgai soil. *Austral. J. Botany* **15**, pp. 481-498.

[17] Srivastava, M.S. and Carter, E.M. (1983). *An Introduction to Applied Multivariate Statistics*, North Holland, New York.

[18] Srivastava, M.S. and von Rosen, D. (2002). Regression models with unknown singular covariance matrix. *Linear Algebra Appl.* **354**, pp. 255-273.

[19] Srivastava, M.S. and von Rosen, D. (2004). MANOVA with singular variance matrix. *Acta et Commentationes Universitatis Tartuensis de Mathematica* **8**, pp. 253-269.