Estimate and test the general multivariate linear model in high dimensions

Martin Singull

Department of Mathematics
Linköping university, Sweden

LINKÖPING UNIVERSITY

Content

Two sample tests

Hotelling's T^2-test & High dimensional tests

Multivariate Linear Model

General linear hypothesis – LRT & High dimensional tests

General Multivariate Linear Model / Growth Curve Model

Estimators

Testing the mean – LRT & High dimensional tests

Testing two covariance structures – LRT & High dimensional tests

Introduction – Two sample tests

Let $\boldsymbol{x}_{ij}$ be independent and identically distributed vectors with p-variate normal distribution $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, where $i = 1, 2$ and $j = 1, ..., N_i$.

The sample mean vectors are, respectively, given by

$$\bar{\boldsymbol{x}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \boldsymbol{x}_{ij}, \quad i = 1, 2,$$

and the sample covariance matrices are, respectively, given by

$$\boldsymbol{S}_i = \frac{1}{n_i} \boldsymbol{X}_i (\boldsymbol{I} - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}') \boldsymbol{X}_i', \quad n_i = N_i - 1, \quad i = 1, 2,$$

where $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iN_i}) : p \times N_i$.

When $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, an unbiased estimator of $\boldsymbol{\Sigma}$ is given by

$$\boldsymbol{S} = \frac{n_1 \boldsymbol{S}_1 + n_2 \boldsymbol{S}_2}{n}, \quad n = n_1 + n_2 = N_1 + N_2 - 2.$$

Hotelling's T^2-test

We wish to test the hypothesis

$$H : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{vs.} \quad A : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

As generalization of the Student's t-test statistic in the univariate case, we get Hotelling's T^2-test statistic, given as

$$T^2 = \left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' \boldsymbol{S}^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$$

The null hypothesis H is rejected at level α, if

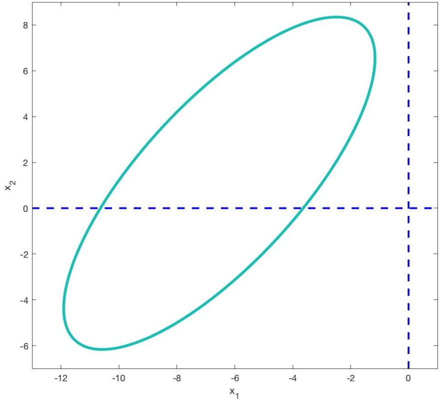
$$\frac{n-p+1}{np} T^2 \geq F_\alpha(p, n-p+1).$$

Here we must have $n \geq p$.

An $100(1-\alpha)\%$ confidence region will be an ellipsoidal region given by

$$R_{\boldsymbol{\mu}} = \left\{ \boldsymbol{\mu} : \frac{N_1 N_2}{N_1 + N_2} \Big(\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}} - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \Big)' \boldsymbol{S}^{-1} \Big(\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}} - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \Big) \leq \frac{np}{n-p+1} F_{1-\alpha}(p, n-p) \right\}$$

Example: $p = 2$



What happens if $n < p$?

Then the sample covariance matrix $\boldsymbol{S}$ is singular.

We can not use $\boldsymbol{S}^{-1}$,

but, $\mathrm{tr}\boldsymbol{S}$, $\boldsymbol{S}^{+}$ or $\boldsymbol{D}_{\boldsymbol{S}}$, ...

In many applications, high dimensional data, when the dimension is often comparable to or even (much) larger than the sample size, are given.

Examples include

- genomics,
- Electroencephalograph (EEG)
- medical imaging,
- risk management,
- web search problems,

to mention a few.

In such high dimensional settings, classical methods designed for the low-dimensional case either perform poorly or are no longer applicable.

For example, the performance of Hotelling's T^2 test is unsatisfactory when the dimension is high relative to the sample sizes.

Dempster's test

When $n < p$, a test if two mean vectors are equal, $H : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ has been proposed by Dempster (1958), under the assumption that the two distributions have the same covariance matrix.

Dempster's test statistic is given by

$$T_D = \left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{-1} \frac{(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)}{\mathrm{tr}\boldsymbol{S}}$$

If $\boldsymbol{\Sigma} = \gamma^2 \boldsymbol{I}_p$ one can show that under the null hypothesis

$$T_D \underset{H}{\sim} F(p, np).$$

It may be noted that when $\boldsymbol{\Sigma} = \gamma^2 \boldsymbol{I}_p$, and under the assumption of normality Dempster's test T_D is uniformly most powerful among all tests whose power depends on $\boldsymbol{\mu}'\boldsymbol{\mu}/\gamma^2$.

For a general $\boldsymbol{\Sigma}$, under the assumption of normality and assuming

$$(\star) \qquad 0 < \lim_{p \to \infty} a_i < \infty,\ i = 1, ..., 4, \text{ where } a_i = \frac{\mathrm{tr}\boldsymbol{\Sigma}^i}{p}$$

one can show that, under the null hypothesis,

$$T_D \approx F([\hat{r}], [n\hat{r}]),$$

where $[a]$ denotes the largest integer value $\leq a$, $\hat{r} = p\hat{b}$, $\hat{b} = \frac{\hat{a}_1^2}{\hat{a}_2}$,

$$\hat{a}_1 = \frac{\mathrm{tr}\boldsymbol{S}}{p}, \quad \text{and} \quad \hat{a}_2 = \frac{1}{p}\left(\mathrm{tr}\boldsymbol{S}^2 - \frac{1}{n}(\mathrm{tr}\boldsymbol{S})^2\right).$$

Bai and Saranadasa's test

Bai and Saranadasa (1996) proposed another asymptotically equivalent test which does not require the assumption of normality but have asymptotically the same power as the one proposed by Dempster (1958).

The statistic testing the hypothesis about equal means given by Bai and Saranadasa is

$$T_{BS} = \frac{\left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - \mathrm{tr}\boldsymbol{S}}{\sqrt{2\left(\mathrm{tr}\boldsymbol{S}^2 - \frac{1}{n}(\mathrm{tr}\boldsymbol{S})^2\right)}} \underset{H}{\sim} N(0,1)$$

Srivastavas' test

Srivastava (2007) proposed a Hotelling's T^2 type test, by using Moore-Penrose inverse of the sample covariance matrix $\boldsymbol{S}^+$ instead of the inverse when N is smaller than p.

The test statistic given by Srivastava (2007) is

$$T^{+2} = \left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' \boldsymbol{S}^+ (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$$

and the asymptotic distribution, assuming $(\star)$, is proved to be

$$\frac{\hat{b}p}{n} T^{+2} \approx \chi^2(n).$$

Srivastava and Du's test

It may be noted that all the above discussed tests are invariant under the group of orthogonal matrices.

A test that is invariant under the group of non-singular diagonal matrices has recently been proposed by Srivastava and Du (2008) under the normal distribution and Srivastava (2009) under non-normality.

It may be noted that this test is not invariant under the transformation by orthogonal matrices.

The test statistic given by Srivastava and Du is

$$T_{SD} = \frac{\left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' \boldsymbol{D}_{\boldsymbol{S}}^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - p}{\sqrt{2\left(\mathrm{tr}\widehat{\boldsymbol{R}}^2 - \frac{p^2}{n}\right) c_{p,n}}}$$

where $\widehat{\boldsymbol{R}} = \boldsymbol{D}_{\boldsymbol{S}}^{-1/2} \boldsymbol{S} \boldsymbol{D}_{\boldsymbol{S}}^{-1/2}$, $\boldsymbol{D}_{\boldsymbol{S}} = \mathrm{diag}(s_{11}, ..., s_{pp})$, $\boldsymbol{S} = (s_{ij})$ and

$$c_{p,n} = 1 + \frac{\mathrm{tr}\widehat{\boldsymbol{R}}^2}{p^{3/2}} \xrightarrow{p} 1 \quad \text{as } (n,p) \to \infty.$$

Assuming some conditions, similar to $(\star)$ on the correlation matrix $\boldsymbol{R}$, and under the hypothesis of equality of two mean vectors, T_{SD} has asymptotically standard normal distribution.

Introduction – Multivariate Linear Model

Definition. Let $\boldsymbol{B} : p \times m$ be an unknown parameter matrix, $\boldsymbol{C} : m \times N$ known design matrix such that $r = \mathrm{rank}(\boldsymbol{C})$ and $r + p \leq N$. The Multivariate Linear Model (MLM) is given by

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{C} + \boldsymbol{E},$$

where the columns of $\boldsymbol{E}$ are assumed to be independently $p-$variate normally distributed with mean zero and an unknown positive definite covariance matrix $\boldsymbol{\Sigma}$, i.e.,

$$\boldsymbol{E} \sim N_{p,N}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_N) \quad \Leftrightarrow \quad \mathrm{vec}\, \boldsymbol{E} \sim N_{pN}(\boldsymbol{0}, \boldsymbol{I}_N \otimes \boldsymbol{\Sigma})$$

where $\boldsymbol{E} = (\boldsymbol{e}_1, \dots, \boldsymbol{e}_N) : p \times N$ and $\mathrm{vec}\, \boldsymbol{E} = (\boldsymbol{e}_1', \dots, \boldsymbol{e}_N')' : pN \times 1$.

Multivariate Linear model – MLEs

If $\boldsymbol{C}$ has full rank, the MLEs for the MLM is given by

$$\widehat{\boldsymbol{B}} = \boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1},$$

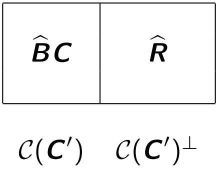
$$N\widehat{\boldsymbol{\Sigma}} = \boldsymbol{X}\left(\boldsymbol{I}_N - \boldsymbol{P}_{\boldsymbol{C}'}\right)\boldsymbol{X}' = \widehat{\boldsymbol{R}}\widehat{\boldsymbol{R}}' = \boldsymbol{V},$$

where $\boldsymbol{P}_{\boldsymbol{C}'} = \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}$, i.e., the projection on the space $\mathcal{C}(\boldsymbol{C}')$ and the estimated mean structure and residual are

$$\widehat{\boldsymbol{B}}\boldsymbol{C} = \boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}'},$$

$$\widehat{\boldsymbol{R}} = \boldsymbol{X}\left(\boldsymbol{I}_N - \boldsymbol{P}_{\boldsymbol{C}'}\right).$$

The whole space can now be decomposed as $\mathcal{C}(\boldsymbol{C}') \boxplus \mathcal{C}(\boldsymbol{C}')^{\perp}$,



The general linear hypothesis

The general linear hypothesis is expressed as $H : \boldsymbol{K}\boldsymbol{B}' = \mathbf{0}$, where $\boldsymbol{K}$ is a known $q \times m$ matrix of rank $q \leq m$.

Let the error sum of squares and products be given by the matrix

$$\boldsymbol{V} = \boldsymbol{X}\left(\boldsymbol{I}_N - \boldsymbol{P}_{\boldsymbol{C}'}\right)\boldsymbol{X}' \quad \text{and} \quad \boldsymbol{S} = \frac{1}{n}\boldsymbol{V}, \quad n = N - m,$$

and the sum of squares and products due to regression under the hypothesis H is

$$\boldsymbol{W} = \widehat{\boldsymbol{B}}\boldsymbol{K}'(\boldsymbol{K}(\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{K}')^{-1}\boldsymbol{K}\widehat{\boldsymbol{B}}'.$$

The LRT of size α of the hypothesis $H : \boldsymbol{KB}' = \mathbf{0}$, rejects H if $\Lambda \leq c_\alpha$, where

$$\Lambda = \frac{|\boldsymbol{V}|}{|\boldsymbol{V} + \boldsymbol{W}|}$$

and c_α is chosen such that the size of the test is α.

The asymptotic expansion for this LRT is given by

$$P\left[-\left(f - \frac{1}{2}(p - m + 1)\right)\ln\Lambda \geq z\right] \approx$$
$$\approx P[\chi_l^2 \geq z] + \frac{\gamma}{f^2}(P[\chi_{l+4}^2 \geq z] - P[\chi_l^2 \geq z]),$$

where $f = n - q$, $l = pm$ and $\gamma = \dfrac{l(p^2 + m^2 - 5)}{48}$.

Test given by Fujikoshi et al.

Fujikoshi et al. (2004) generalize the two-sample test given by Dempster (1958) to the MANOVA problem, under the assumption that

$$(p/n) \to c \in (0, \infty).$$

The statistic given by Fujikoshi et al. is

$$\tilde{T}_D = \sqrt{p}\left(\frac{\mathrm{tr}\boldsymbol{W}}{\mathrm{tr}\boldsymbol{S}} - q\right)$$

and

$$\frac{\tilde{T}_D}{\hat{\sigma}_D} \to N(0,1),$$

where $\hat{\sigma}_D = 2q\dfrac{\hat{a}_2}{\hat{a}_1}$, $\hat{a}_1 = \dfrac{\mathrm{tr}\boldsymbol{S}}{p}$, and $\hat{a}_2 = \dfrac{1}{p}\left(\mathrm{tr}\boldsymbol{S}^2 - \dfrac{1}{n}(\mathrm{tr}\boldsymbol{S})^2\right)$.

Srivastava and Fujikoshi's test

Other tests that do not require the assumption

$$(p/n) \to c \in (0, \infty)$$

have been proposed by Srivastava and Fujikoshi (2006) with the test statistic

$$T_{SF} = \frac{\sqrt{p}(\mathrm{tr}\,\boldsymbol{W} - q\,\mathrm{tr}\,\boldsymbol{S})}{\sqrt{2q\hat{a}_2}} \underset{H}{\sim} N(0,1).$$

Schott (2007) proposed the same test as proposed by Srivastava and Fujikoshi (2006) but required the assumption above to obtain the asymptotic distribution of the test statistic.

Yamada and Srivastava's test

The above tests are, however, not invariant under the transformation by non-singular diagonal matrices. A test that has this property for the MANOVA problem has been recently proposed by Yamada and Srivastava (2012) under normality.

The test statistic given by Yamada and Srivastava (2012) is

$$T_{YS} = \frac{\mathrm{tr}\,\boldsymbol{W}\boldsymbol{D}_S^{-1} - \frac{n}{n-2}pq}{\sqrt{2q\left(\mathrm{tr}\widehat{\boldsymbol{R}}^2 - \frac{p^2}{n}\right)c_{p,n}}},$$

where $\widehat{\boldsymbol{R}} = \boldsymbol{D}_{\boldsymbol{S}}^{-1/2}\boldsymbol{S}\boldsymbol{D}_{\boldsymbol{S}}^{-1/2}$ and $c_{p,n} = 1 + \dfrac{\mathrm{tr}\widehat{\boldsymbol{R}}^2}{p^{3/2}}$

Assuming certain conditions on the correlation matrix $\boldsymbol{R}$ the asymptotic distribution, when $n, p \to \infty$, under the null hypothesis is standard normal.

Growth Curve Model (Potthoff and Roy, 1964)

Definition. Let $\boldsymbol{X} : p \times N$ and $\boldsymbol{B} : q \times m$ be the observation and parameter matrices, respectively, and let $\boldsymbol{A} : p \times q$ and $\boldsymbol{C} : m \times N$ be the within and between individual design matrices, respectively. Suppose that $q \leq p$ and $p \leq N - r = n$, where $r = \mathrm{rank}(\boldsymbol{C})$.

The Growth Curve model (GCM) is given by

$$\boldsymbol{X} = \boldsymbol{ABC} + \boldsymbol{E},$$

where $\boldsymbol{E} \sim N_{p,N}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_N)$.

We will assume that $r = \mathrm{rank}(\boldsymbol{C}) = m$, i.e., $n = N - m$.

For the GCM, the mean parameter space is independent of p and n, whereas the covariance matrix $\boldsymbol{\Sigma}$ increases in size with p.

Growth Curve Model – MLEs

If $\boldsymbol{A}$ and $\boldsymbol{C}$ has full rank, the MLEs for the GCM is given

$$\begin{aligned}
\widehat{\boldsymbol{B}}_{MLE} &= \left(\boldsymbol{A}'\boldsymbol{V}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\boldsymbol{V}^{-1}\boldsymbol{X}\boldsymbol{C}'\left(\boldsymbol{C}\boldsymbol{C}'\right)^{-1}, \text{ i.e.,}\\
\boldsymbol{A}\widehat{\boldsymbol{B}}_{MLE}\boldsymbol{C} &= \boldsymbol{P}_{\boldsymbol{A}}^{\boldsymbol{V}}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}'},\\
N\widehat{\boldsymbol{\Sigma}}_{MLE} &= \left(\boldsymbol{X}-\boldsymbol{A}\widehat{\boldsymbol{B}}_{MLE}\boldsymbol{C}\right)\left(\boldsymbol{X}-\boldsymbol{A}\widehat{\boldsymbol{B}}_{MLE}\boldsymbol{C}\right)' = \underbrace{\widehat{\boldsymbol{R}}\widehat{\boldsymbol{R}}'}_{=\boldsymbol{V}}+\widehat{\boldsymbol{R}}_1\widehat{\boldsymbol{R}}_1',
\end{aligned}$$

where

$$\begin{aligned}
\widehat{\boldsymbol{R}}_1 &= \left(\boldsymbol{I}_p-\boldsymbol{P}_{\boldsymbol{A}}^{\boldsymbol{V}}\right)\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}'},\\
\widehat{\boldsymbol{R}} &= \boldsymbol{X}\left(\boldsymbol{I}_N-\boldsymbol{P}_{\boldsymbol{C}'}\right),\\
\boldsymbol{V} &= \widehat{\boldsymbol{R}}\widehat{\boldsymbol{R}}' = \boldsymbol{X}\left(\boldsymbol{I}_N-\boldsymbol{P}_{\boldsymbol{C}'}\right)\boldsymbol{X}',\\
\boldsymbol{P}_{\boldsymbol{C}'} &= \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C} = \text{ projection on } \mathcal{C}(\boldsymbol{C}'),\\
\boldsymbol{P}_{\boldsymbol{A}}^{\boldsymbol{V}} &= \boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{V}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\boldsymbol{V}^{-1} = \text{ projection on } \mathcal{C}_{\boldsymbol{V}}(\boldsymbol{A}).
\end{aligned}$$

Growth Curve Model – MLEs

If $\boldsymbol{A}$ and $\boldsymbol{C}$ has full rank, the MLEs for the GCM is given

$$\widehat{\boldsymbol{B}}_{MLE} = \left(\boldsymbol{A}'\boldsymbol{V}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\boldsymbol{V}^{-1}\boldsymbol{X}\boldsymbol{C}'\left(\boldsymbol{C}\boldsymbol{C}'\right)^{-1}, \text{ i.e.,}$$

$$\boldsymbol{A}\widehat{\boldsymbol{B}}_{MLE}\boldsymbol{C} = \boldsymbol{P}_{\boldsymbol{A}}^{\boldsymbol{V}}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}'},$$

$$N\widehat{\boldsymbol{\Sigma}}_{MLE} = \left(\boldsymbol{X} - \boldsymbol{A}\widehat{\boldsymbol{B}}_{MLE}\boldsymbol{C}\right)\left(\boldsymbol{X} - \boldsymbol{A}\widehat{\boldsymbol{B}}_{MLE}\boldsymbol{C}\right)' = \underbrace{\widehat{\boldsymbol{R}}\widehat{\boldsymbol{R}}'}_{=\boldsymbol{V}} + \widehat{\boldsymbol{R}}_1\widehat{\boldsymbol{R}}_1',$$

where

$$\widehat{\boldsymbol{R}}_1 = \left(\boldsymbol{I}_p - \boldsymbol{P}_{\boldsymbol{A}}^{\boldsymbol{V}}\right)\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}'},$$

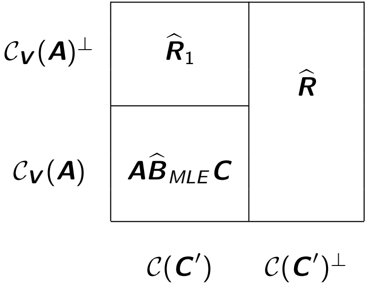
$$\widehat{\boldsymbol{R}} = \boldsymbol{X}\left(\boldsymbol{I}_N - \boldsymbol{P}_{\boldsymbol{C}'}\right),$$

$$\boldsymbol{V} = \widehat{\boldsymbol{R}}\widehat{\boldsymbol{R}}' = \boldsymbol{X}\left(\boldsymbol{I}_N - \boldsymbol{P}_{\boldsymbol{C}'}\right)\boldsymbol{X}',$$

$$\boldsymbol{P}_{\boldsymbol{C}'} = \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C} = \text{ projection on } \mathcal{C}(\boldsymbol{C}'),$$

$$\boldsymbol{P}_{\boldsymbol{A}}^{\boldsymbol{V}} = \boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{V}^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\boldsymbol{V}^{-1} = \text{ projection on } \mathcal{C}_{\boldsymbol{V}}(\boldsymbol{A}).$$

$$
\begin{aligned}
&\mathcal{C}_{\boldsymbol{V}}(\boldsymbol{A}) \otimes \mathcal{C}(\boldsymbol{C}') \boxplus (\mathcal{C}_{\boldsymbol{V}}(\boldsymbol{A}) \otimes \mathcal{C}(\boldsymbol{C}'))^{\perp} \\
&\quad = (\mathcal{C}_{\boldsymbol{V}}(\boldsymbol{A}) \otimes \mathcal{C}(\boldsymbol{C}')) \boxplus hcalC_{\boldsymbol{V}}(\boldsymbol{A})^{\perp} \otimes \mathcal{C}(\boldsymbol{C}') \boxplus \mathcal{V} \otimes \mathcal{C}(\boldsymbol{C}')^{\perp}
\end{aligned}
$$



$$\widehat{\boldsymbol{R}} = \boldsymbol{X}\,(\boldsymbol{I}_N - \boldsymbol{P}_{\boldsymbol{C}'})$$

$$\widehat{\boldsymbol{R}}_1 = \left(\boldsymbol{I}_p - \boldsymbol{P}_{\boldsymbol{A}}^{\boldsymbol{V}}\right)\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}'}$$

$$
\begin{aligned}
\boldsymbol{A}\widehat{\boldsymbol{B}}_{MLE}\boldsymbol{C} &= \boldsymbol{P}_{\boldsymbol{A}}^{\boldsymbol{V}}\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{C}'}, \\
N\widehat{\boldsymbol{\Sigma}}_{MLE} &= \widehat{\boldsymbol{R}}\widehat{\boldsymbol{R}}' + \widehat{\boldsymbol{R}}_1\widehat{\boldsymbol{R}}_1'.
\end{aligned}
$$

Properties of the estimators

Mean and covariance for $\widehat{\boldsymbol{B}}_{MLE}$ are (Kollo and von Rosen, 2005)

$$\mathrm{E}\left(\widehat{\boldsymbol{B}}_{MLE}\right) = \boldsymbol{B}, \qquad \textit{and}$$

$$\mathrm{cov}\left(\widehat{\boldsymbol{B}}_{MLE}\right) = \frac{n-1}{n-1-(p-q)}(\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1},$$

if $n-1-(p-q) > 0$, where $n = N - m$.

Since $q \le p \le n$ we have $\boxed{\dfrac{n-1}{n-1-(p-q)} \ge 1.}$

von Rosen (1991) show that the estimator $\widehat{\boldsymbol{\Sigma}}_{MLE}$ is a biased, as

$$\mathrm{E}\left(\widehat{\boldsymbol{\Sigma}}_{MLE}\right) = \boldsymbol{\Sigma} - \frac{m}{N}\frac{n-1-2(p-q)}{n-1-(p-q)}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'.$$

The bias depends on the design $\boldsymbol{A}$ and thus it could be significant.

Example - p time points and N observations

In a small simulation example we may use the parameters and designs

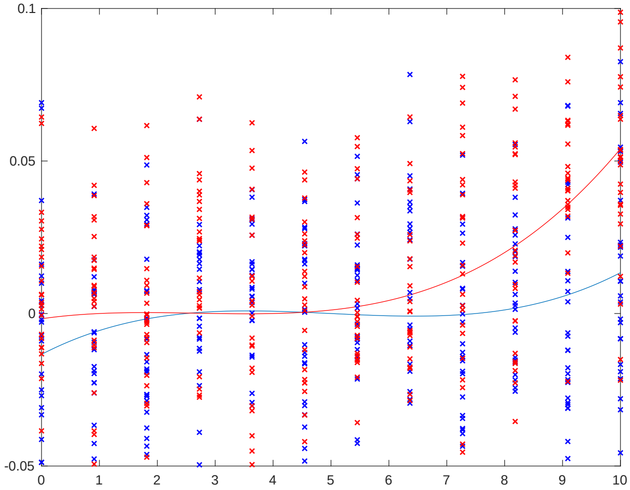
$$\boldsymbol{B} = \begin{pmatrix} b_{01} & b_{02} \\ b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix} = \begin{pmatrix} -0.0134 & -0.0017 \\ 0.0098 & 0.0027 \\ -0.0021 & -0.0011 \\ 0.0001 & 0.0001 \end{pmatrix},$$

$$\boldsymbol{A} = \begin{pmatrix} 1 & t_1 & t_1^2 & t_1^3 \\ 1 & t_2 & t_2^2 & t_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_p & t_p^2 & t_p^3 \end{pmatrix} \quad \text{and} \quad \boldsymbol{C} = \begin{pmatrix} \mathbf{1}'_{N_1} & \mathbf{0}'_{N_2} \\ \mathbf{0}'_{N_1} & \mathbf{1}'_{N_2} \end{pmatrix},$$

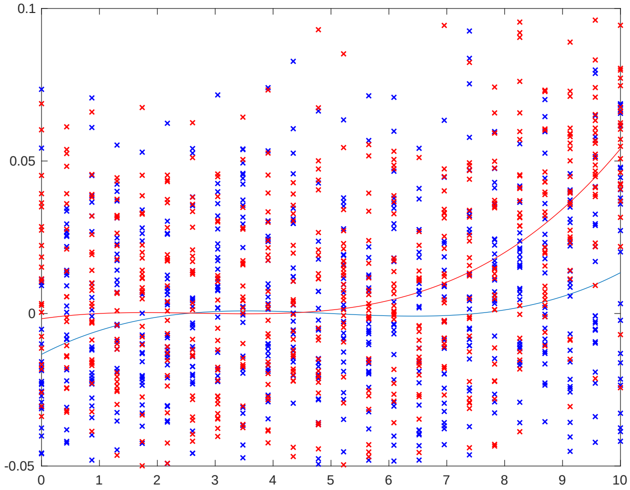
where we have used $q = 4$ (i.e., cubic growth) and $m = 2$ groups for simplicity with $N_1 = N_2 = N/2$. Furthermore, we put $t_1 = 0$ and

$$t_i = i\frac{10}{p-1}, \quad \text{for } i = 1, \ldots, p-1.$$

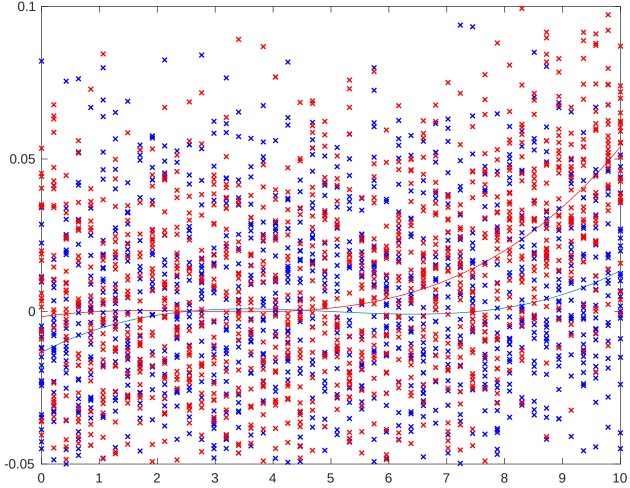
$N_1 = N_2 = 25$ $(n = 48)$ and $p = 12$



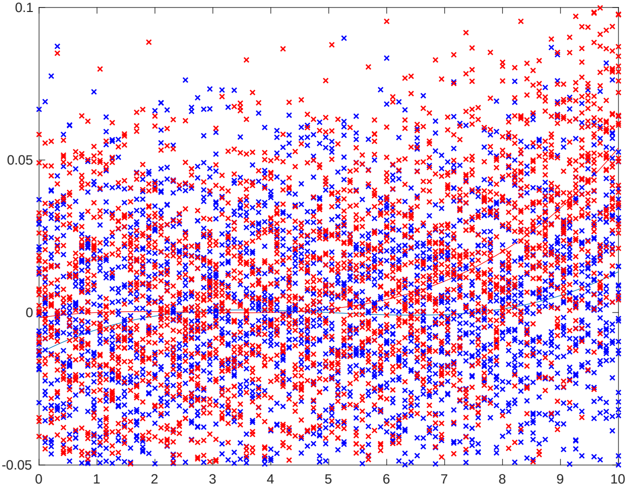
$N_1 = N_2 = 25$ $(n = 48)$ and $p = 24$



$N_1 = N_2 = 25$ $(n = 48)$ and $p = 48$



$N_1 = N_2 = 25$ ($n = 48$) and $p = 96$



Unweighted estimator of $\boldsymbol{B}$

A natural alternative to the MLE would be an unweighted estimator of $\boldsymbol{B}$ given by

$$\boxed{\widehat{\boldsymbol{B}} = (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}.}$$

This estimator is simpler than the MLE, since we do not need to calculate the inverse of the sum of squares matrix $\boldsymbol{V}^{-1}$.

This unweighted estimator is obtained by considering the model

$$\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1} = \boldsymbol{A}\boldsymbol{B} + \widetilde{\boldsymbol{E}},$$

where $\widetilde{\boldsymbol{E}} = \boldsymbol{E}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1} \sim N_{p,m}\left(\mathbf{0}, \boldsymbol{\Sigma}, (\boldsymbol{C}\boldsymbol{C}')^{-1}\right)$.

The distribution of the estimator is given by

$$\widehat{\boldsymbol{B}} \sim N_{q,m}(\boldsymbol{B}, (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}, (\boldsymbol{C}\boldsymbol{C}')^{-1}),$$

i.e., we have

$$\mathrm{E}\left(\widehat{\boldsymbol{B}}\right) = \boldsymbol{B},$$

and

$$\mathrm{cov}\left(\widehat{\boldsymbol{B}}\right) = (\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}.$$

Again the bias for estimating $\boldsymbol{\Sigma}$ could be significant with

$$\mathrm{E}(\widehat{\boldsymbol{\Sigma}}_{UW}) = \boldsymbol{\Sigma} + \frac{m}{N}\left(\boldsymbol{\Sigma} + (\boldsymbol{I}_p - \boldsymbol{P}_{\boldsymbol{A}})\,\boldsymbol{\Sigma}\,(\boldsymbol{I}_p - \boldsymbol{P}_{\boldsymbol{A}})'\right).$$

Estimator based on an asymptotic likelihood function

Due to the normality assumption, i.e., since the distribution is symmetric around the mean, in order to estimate the mean parameters it is natural to consider

$$\frac{1}{p}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{ABC})()'\} =$$
$$= \frac{1}{p}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{XP_{C'}}-\boldsymbol{ABC})()'\} + \frac{1}{p}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\}$$

Kollo et al. (2011) showed that

$$\frac{\frac{1}{p}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\}-n}{\sqrt{(n-r(\boldsymbol{C}))/p}} \underset{asymp.}{\sim} N(0,2) \quad \text{as } p,N\to\infty \text{ and } \frac{p}{N}\to c>0,$$

where $n = N - \mathrm{rank}(\boldsymbol{C}) = N - r$, and

$$\frac{\frac{1}{\sqrt{p}}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{XP_{C'}}-\boldsymbol{ABC})()'\}-r\sqrt{p}}{\sqrt{r}} \underset{asymp.}{\sim} N(0,2) \quad \text{as } p,N\to\infty.$$

Based on these two asymptotic distributions one can find the asymptotic likelihood function for the GCM.

Using an approach similar to the restricted maximum likelihood method Kollo et al. (2011) found an likelihood based estimator for the mean $\boldsymbol{B}$ as

$$\widehat{\boldsymbol{B}}_{AL} = (\boldsymbol{A}'\boldsymbol{V}^{+}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{V}^{+}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-} + (\boldsymbol{A}'\boldsymbol{V}^{+}\boldsymbol{A})^{o}\boldsymbol{Z}_1 + \boldsymbol{A}'\boldsymbol{V}^{+}\boldsymbol{A}\boldsymbol{Z}_2\boldsymbol{C}^{o\prime},$$

where $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are arbitrary matrices.

The estimator $\widehat{\boldsymbol{B}}_{AL}$ is unique and with probability 1 equals

$$\widehat{\boldsymbol{B}}_{AL} = (\boldsymbol{A}'\boldsymbol{V}^{+}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{V}^{+}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1},$$

if and only if $\mathrm{rank}(\boldsymbol{A}) = q < \min\{p, n\}$, where $n = N - r$, $r = \mathrm{rank}(\boldsymbol{C})$ and

$$\mathcal{C}(\boldsymbol{A}) \cap \mathcal{C}(\boldsymbol{V})^{\perp} = \{\mathbf{0}\}.$$

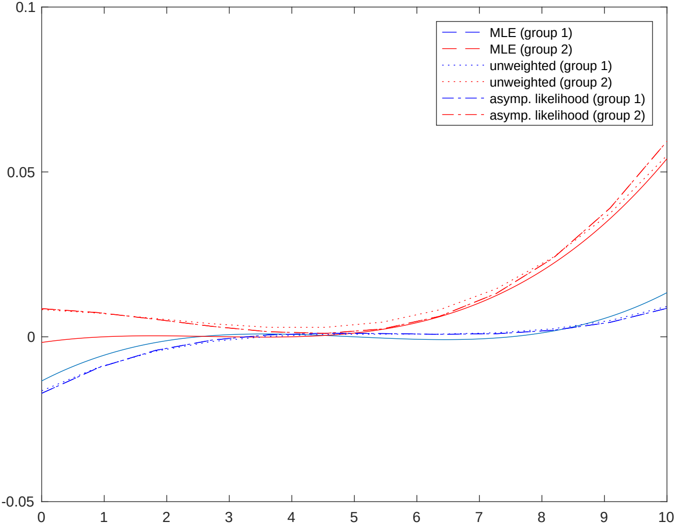
Furthermore, one can show that

$$
\begin{aligned}
\mathrm{E}(\widehat{\boldsymbol{B}}_{AL}) &= \boldsymbol{B}, \\
\mathrm{cov}\left(\widehat{\boldsymbol{B}}_{AL}\right) &\approx \frac{(p-q-1)(p-1)}{(n-q-1)(p-n+q-1)}(\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1},
\end{aligned}
$$

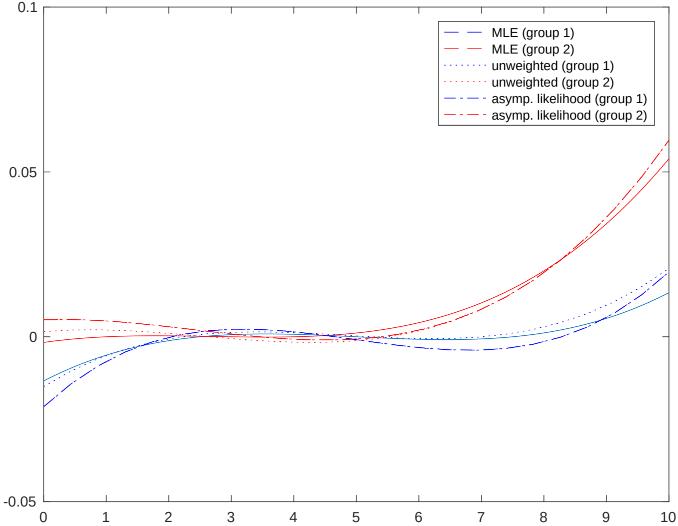
where $n = N - \mathrm{rank}(\boldsymbol{C}) = N - r$.

Example, cont.

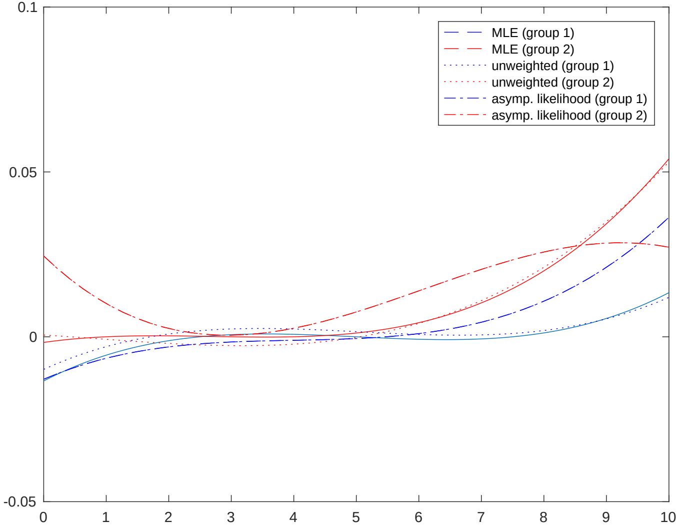
$N_1 = N_2 = 25$ $(n = 48)$ and $p = 12$ $\quad (q \leq p \leq n)$



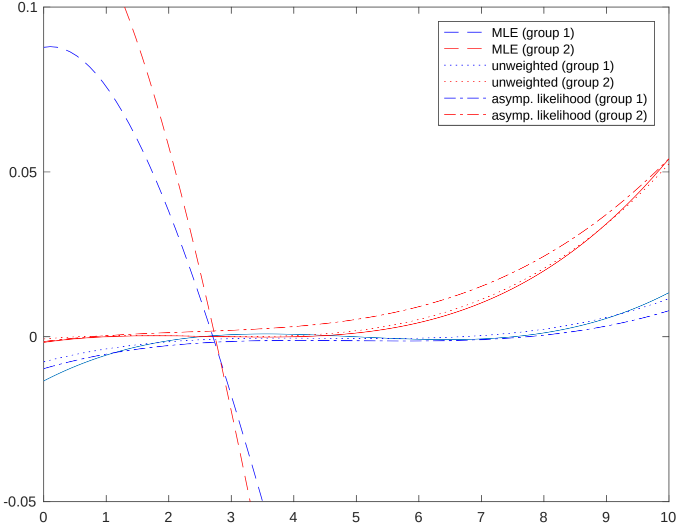
$N_1 = N_2 = 25$ $(n = 48)$ and $p = 24$ $\quad (q \leq p \leq n)$



$N_1 = N_2 = 25$ $(n = 48)$ and $p = 48$ $\quad (q \leq p \leq n)$



$N_1 = N_2 = 25$ $(n = 48)$ and $p = 96$ $\quad (q \leq p \leq n)$



Compare the estimators $\widehat{\boldsymbol{B}}_{MLE}$ and $\widehat{\boldsymbol{B}}$

All three estimators $\widehat{\boldsymbol{B}}_{MLE}$, $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{B}}_{AL}$ are unbiased.

The covariances for $\widehat{\boldsymbol{B}}_{MLE}$, $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{B}}_{AL}$ respectively are given by

$$
\begin{aligned}
\operatorname{cov}\left(\widehat{\boldsymbol{B}}_{MLE}\right) &= \frac{n-1}{n-1-(p-q)}(\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1},\\
\operatorname{cov}\left(\widehat{\boldsymbol{B}}\right) &= (\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}.\\
\operatorname{cov}\left(\widehat{\boldsymbol{B}}_{AL}\right) &\approx \frac{(p-q-1)(p-1)}{(n-q-1)(p-n+q-1)}(\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}.
\end{aligned}
$$

To compare the estimators we must compare their covariances, i.e., we want to compare

$$
(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1} \quad \text{and} \quad (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}.
$$

Following Rao (1967) (Lemma 2.c) or Baksalary and Puntanen (1991) one can show that

$$\boxed{(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1} \leq (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}}$$

with equality if and only if $\mathcal{C}(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}) = \mathcal{C}(\boldsymbol{A})$.

The inequality is with respect to the Loewner partial ordering, i.e., $\boldsymbol{C} \leq \boldsymbol{A}$ if $\boldsymbol{A} - \boldsymbol{C}$ is nonnegative definite.

For large n, the unweighted unbiased estimator of $\boldsymbol{B}$ *has a larger covariance than the weighted one, as expected since the weighted estimator is the MLE.*

Note when $\mathcal{C}(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}) = \mathcal{C}(\boldsymbol{A})$

Under the restriction $\mathcal{C}(\boldsymbol{\Sigma}^{-1}\boldsymbol{A}) = \mathcal{C}(\boldsymbol{A})$, the MLE for the GCM is given by the unweighted estimator.

This condition is fulfilled, for example when

- sphericity $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_p$, or
- intraclass covariance matrix $\boldsymbol{\Sigma} = \sigma^2((1-\rho)\boldsymbol{I}_p + \rho \mathbf{1}\mathbf{1}')$ and $\boldsymbol{A}$ includes a column vector of ones, e.g., $\boldsymbol{A} = [\mathbf{1} \;:\; \boldsymbol{A}_1]$.

Testing the mean

Model: $\boldsymbol{X} = \boldsymbol{ABC} + \boldsymbol{E}$, where $\boldsymbol{E} \sim N_{p,N}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_N)$.

Srivastava and Singull (2017a) tested the hypothesis

$$H : \boldsymbol{B} = \mathbf{0} \quad \text{vs.} \quad A : \boldsymbol{B} \neq \mathbf{0}.$$

Four proposed tests

- $p < N$ - two statistics based on the MLE and the unweighted estimator, respectively (T_1 and T_2).
- $p > N$ - two new statistics based on the trace of the variation matrices due to the hypothesis (between sum of squares) and the error (within sum of squares) (T_3 and T_4).

LRT T_1 based on MLEs

Given the MLEs the LRT is given as

$$\lambda_{MLE}^{2/N} = \frac{|\boldsymbol{V} + \widehat{\boldsymbol{R}}_1 \widehat{\boldsymbol{R}}_1'|}{|\boldsymbol{V} + \boldsymbol{V}_1|},$$

where $\widehat{\boldsymbol{R}}_1$, $\boldsymbol{V}$ are given above, and $\boldsymbol{V}_1 = \boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{X}'$.
Using Box's method for approximate the distribution of

$$T_1 = -r \log \lambda_{MLE}^{2/N},$$

one can show that for large N,

$$P_0\left(T_1 > c\right) = P\left(\chi_f^2 > c\right),$$

where $r = n - p + q - (q - m + 1)/2$ and $f = qm$.

See Srivastava and Khatri (1979) for more details.

Equivalently testing the hypothesis $H : \boldsymbol{B} = \mathbf{0}$ one can test

$$H : \boldsymbol{\eta} = \mathbf{0} \quad \text{vs.} \quad A : \boldsymbol{\eta} \neq \mathbf{0},$$

where $\boldsymbol{\eta} = (\boldsymbol{A}'\boldsymbol{A})^{1/2}\boldsymbol{B}(\boldsymbol{C}\boldsymbol{C}')^{1/2}$.

Using the unweighted estimator $\widehat{\boldsymbol{B}}$ given above we have

$$\widehat{\boldsymbol{\eta}} = (\boldsymbol{A}'\boldsymbol{A})^{1/2}\widehat{\boldsymbol{B}}(\boldsymbol{C}\boldsymbol{C}')^{1/2} \sim N_{q,m}(\boldsymbol{\eta}, \boldsymbol{\Delta}, \boldsymbol{I}_m),$$

where

$$\boldsymbol{\Delta} = (\boldsymbol{A}'\boldsymbol{A})^{-1/2}\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1/2}.$$

Based on the distribution of $\widehat{\boldsymbol{\eta}}$ another LRT is given by

$$\lambda^{2/N} = \frac{|\boldsymbol{V}^*|}{|\boldsymbol{V}^* + \boldsymbol{W}|},$$

where $\boldsymbol{V}^* = n\widehat{\boldsymbol{\Delta}} \sim W_q(\boldsymbol{\Delta}, N - m)$ and $\boldsymbol{W} = \widehat{\boldsymbol{\eta}}\widehat{\boldsymbol{\eta}}' \underset{H}{\sim} W_q(\boldsymbol{\Delta}, m)$.

LRT T_2

Again using Box's method to approximate the distribution, one can shown that for large N the distribution of

$$T_2 = -r \log \lambda^{2/N}$$

is given by

$$P_0\left(T_2 > c\right) = P\left(\chi^2_f > c\right),$$

where $r = n - (q - m + 1)/2$ and $f = qm$.

See Srivastava and Khatri (1979) for more details.

Large p and small N - Test statistic T_3, based on $\mathrm{tr}\boldsymbol{W}$

For high dimensions, when $N - m < p$, then $\boldsymbol{V}$ is singular and none of the tests given in above are applicable.

We will propose two new tests.

Again consider the variabel

$$\widehat{\boldsymbol{\eta}} \sim N_{q,m}(\boldsymbol{\eta}, \boldsymbol{\Delta}, \boldsymbol{I}_m),$$

Under the hypothesis $H : \boldsymbol{B} = \boldsymbol{0}$ (i.e., $\boldsymbol{\eta} = 0$) we have

$$\boldsymbol{W} = \widehat{\boldsymbol{\eta}}\widehat{\boldsymbol{\eta}}' \sim W_q(\boldsymbol{\Delta}, m).$$

We also see that $\boldsymbol{W}$ and $\widehat{\boldsymbol{\Sigma}}$ are independently distributed.

Test statistic based on $\operatorname{tr} \boldsymbol{W}$

The mean and variance of the statistic $\operatorname{tr} \boldsymbol{W}$, under the hypothesis $H : \boldsymbol{B} = \mathbf{0}$, are given by

$$\mathrm{E}(\operatorname{tr} \boldsymbol{W}) = m \operatorname{tr} \boldsymbol{\Delta},$$
$$\operatorname{var}(\operatorname{tr} \boldsymbol{W}) = 2m \operatorname{tr} \boldsymbol{\Delta}^2.$$

Under the assumption of normality, unbiased and consistent estimators of $\operatorname{tr} \boldsymbol{\Delta}$ and $\operatorname{tr} \boldsymbol{\Delta}^2$ are given by

$$\widehat{\operatorname{tr} \boldsymbol{\Delta}} = \frac{1}{n} \operatorname{tr} \boldsymbol{V}^*,$$
$$\widehat{\operatorname{tr} \boldsymbol{\Delta}^2} = \frac{1}{(n-1)(n+2)} \left(\operatorname{tr} \boldsymbol{V}^{*2} - \frac{1}{n} (\operatorname{tr} \boldsymbol{V}^*)^2 \right),$$

respectively. See Srivastava (2005) for details.

Test statistic T_3

One can show

$$\widetilde{T}_3 = \frac{1}{\sqrt{m}} \frac{\mathrm{tr}\,\boldsymbol{W} - m\,\mathrm{tr}\,\boldsymbol{\Delta}}{\sqrt{2\,\mathrm{tr}\,\boldsymbol{\Delta}^2}} \underset{H}{\to} N(0,1).$$

Substituting unbiased and consistent estimators of $\mathrm{tr}\,\boldsymbol{\Delta}$ and $\mathrm{tr}\,\boldsymbol{\Delta}^2$ we get a test statistic, proposed by Srivastava and Fujikoshi (2006) and Srivastava (2007), which is given by

$$T_3 = \frac{\mathrm{tr}\,\boldsymbol{W} - \frac{m}{n}\mathrm{tr}\,\boldsymbol{V}^*}{\sqrt{\frac{2m}{(n-1)(n+2)}\left(\mathrm{tr}\,\boldsymbol{V}^{*2} - \frac{1}{n}(\mathrm{tr}\,\boldsymbol{V}^*)^2\right)}} \underset{H}{\to} N(0,1).$$

The test statistic T_3 is invariant under the group of orthogonal transformations, but not invariant under the units of measurements, which is an undesirable feature.

That is, the test is not invariant under a diagonal transformation and the test statistic T_3 changes.

We will now propose a test that is invariant under diagonal transformation.

We will also show that this new test performs better than the test above.

Test statistic based on $\mathrm{tr}\boldsymbol{W}\boldsymbol{D}_{\widehat{\boldsymbol{\Delta}}}^{-1}$

The test statistic will be based on the quantity $\mathrm{tr}\boldsymbol{W}\boldsymbol{D}_{\widehat{\boldsymbol{\Delta}}}^{-1}$, where $\boldsymbol{D}_{\widehat{\boldsymbol{\Delta}}}$ is the diagonal matrix with the diagonal elements of $\widehat{\boldsymbol{\Delta}}$.

More precise the test statistic is based on $\mathrm{tr}\boldsymbol{W}\boldsymbol{D}_{\boldsymbol{V}^*}^{-1}$ as

$$T_4 = \frac{n\mathrm{tr}\boldsymbol{W}\boldsymbol{D}_{\boldsymbol{V}^*}^{-1} - nqm/(n-2)}{\sqrt{2m(\mathrm{tr}\widehat{\boldsymbol{R}}^2 - q^2/n)c_{q,n}}},$$

where $\widehat{\boldsymbol{R}} = \boldsymbol{D}_{\boldsymbol{V}^*}^{-1/2}\boldsymbol{V}^*\boldsymbol{D}_{\boldsymbol{V}^*}^{-1/2}$ and $c_{q,n} = 1 + \mathrm{tr}\widehat{\boldsymbol{R}}^2/q^{3/2}$ is an adjustment factor converging to 1 in probability as $(n,q) \to \infty$, $n = O(q^\delta)$, $\delta > 1/2$ proposed by Srivastava and Du (2008).

Test statistic T_4

Define the population correlation matrix as

$$\boldsymbol{R} = \boldsymbol{D}_{\boldsymbol{\Delta}}^{-1/2} \boldsymbol{\Delta} \boldsymbol{D}_{\boldsymbol{\Delta}}^{-1/2}.$$

It has been shown by Srivastava and Du (2008) that a consistent estimator of $\mathrm{tr}\boldsymbol{R}^2/q$ is given by

$$\frac{1}{q}\left(\mathrm{tr}\widehat{\boldsymbol{R}}^2 - \frac{q^2}{n}\right).$$

Hence, for large n and q we have

$$T_4 \stackrel{d}{=} \frac{(n\mathrm{tr}(\boldsymbol{W}\boldsymbol{D}_{\boldsymbol{V}^*}^{-1}) - qm)/\sqrt{q}}{\sqrt{2m\mathrm{tr}\boldsymbol{R}^2/q}} \underset{H}{\rightarrow} N(0,1),$$

when $(n,q) \to \infty$ and under the assumptions $n = O(q^{\delta})$ for $\delta > 1/2$, $\frac{1}{q}\mathrm{tr}\boldsymbol{R}^2 = O(1)$ and $\frac{1}{q^2}\mathrm{tr}\boldsymbol{R}^4 = o(1)$ as $q \to \infty$.

Compare the performance – ASL and power

To compare the four tests we can compute the attained significance level (ASL) and the empirical power.

Let c be the critical value from the distribution considered for the test statistics. With 10000 simulated replications under the null hypothesis, the ASL is computed as

$$\widehat{\alpha} = \frac{(\#\text{ of } t_H \geq c)}{(\#\text{ simulated replications})},$$

where t_H is the values of the test statistics derived from the simulated data under the null hypothesis.

We set the nominal significance level to $\alpha = 5\%$.

For the simulations let

$$\boldsymbol{A} = (a_{ij}), \quad a_{ij} \sim U(0,1), \quad i = 1, ..., p, \quad j = 1, ..., q$$

$$\text{and} \quad \boldsymbol{C} = \begin{pmatrix} \mathbf{1}'_{N_1} & \mathbf{0}'_{N_2} \\ \mathbf{0}'_{N_1} & \mathbf{1}'_{N_2} \end{pmatrix},$$

i.e., with $m = 2$.

For simplicity we will put N even and $N_1 = N_2 = N/2$.

Furthermore, N, p and q will vary depending on which asymptotic is considered.

Since the covariances for the estimators depending on $\mathbf{\Sigma}$, we will use three different covariance matrices for the simulation study.

(I): The first one is identity, i.e., $\mathbf{\Sigma}_1 = \boldsymbol{I}_p$.

Furthermore, let $\boldsymbol{D}_j = diag(\sigma_1^{(j)}, \ldots, \sigma_p^{(j)})$, for $j = 2, 3$, be two different diagonal matrices. Define $\sigma_i^{(2)} = 2 + (p - i + 1)/p$, for $i = 1, \ldots, p$, and $\sigma_i^{(3)}$, for $i = 1, \ldots, p$, are independent observ- ations from $\sqrt{U[0, 2]}$, respectively.

Also, let $\boldsymbol{R} = (\rho_{ij})$, where $\rho_{ij} = (-1)^{i+j} r^{|i-j|^f}$.

(II)-(III): The other two covariance matrices that we will use are given as

$$\mathbf{\Sigma}_j = \boldsymbol{D}_j \boldsymbol{R} \boldsymbol{D}_j, \quad \text{with } r = 0.2, f = 0.1, \text{ for } j = 2, 3.$$

To compute the empirical power we can either use the critical value c from the asymptotic distribution, or we can use the estimated critical value $\widehat{c}$ calculated from the simulated data under the null hypothesis, i.e., the critical value calculated from the empirical null distribution.

We will use the estimated critical value since the ASL is greatly affected for some tests.

The empirical power is calculated from 10000 new replications simulated under the alternative hypothesis when $\boldsymbol{B} = (b_{ij})$ and $b_{ij} = 0.1$ if $i + j$ is even and zero otherwise.

Let t_A be the value of the test statistic derived from the simulated data under the alternative hypothesis.

The empirical power are given as

$$\widehat{\beta} = \frac{(\#\text{ of } t_A \geq \widehat{c})}{(\#\text{ simulated replications})}.$$

ASL and empirical power for T_1, T_2, T_3 and T_4

$p = 30$ and $N = 50$

	q	ASL T_1	ASL T_2	Power T_1	Power T_2	ASL T_3	ASL T_4	Power T_3	Power T_4
(I)	4	5.43	5.42	36.83	82.06	8.01	5.69	86.83	85.35
	6	5.20	5.29	61.02	96.70	7.72	6.00	98.59	98.45
	10	5.51	5.37	95.23	100.00	6.86	5.44	100.00	100.00
	14	6.34	5.36	99.84	100.00	6.68	5.58	100.00	100.00
	18	6.86	5.95	100.00	100.00	6.23	4.98	100.00	100.00
	24	9.82	8.21	100.00	100.00	6.54	5.62	100.00	100.00
(II)	4	4.81	4.81	24.91	51.81	7.40	4.92	38.00	39.13
	6	5.17	4.65	39.86	75.92	7.41	5.43	65.52	67.01
	10	5.87	5.03	77.23	98.71	6.89	5.58	98.60	98.48
	14	6.83	5.60	95.66	99.97	6.69	5.17	99.99	100.00
	18	7.42	6.05	99.81	100.00	7.55	5.96	100.00	100.00
	24	9.37	7.77	100.00	100.00	7.61	5.91	100.00	100.00
(III)	4	5.09	5.23	53.61	91.99	7.73	5.39	85.18	87.24
	6	5.15	5.08	77.53	99.39	7.35	5.38	99.30	99.33
	10	5.84	5.15	99.00	100.00	7.22	5.58	100.00	100.00
	14	6.79	5.45	100.00	100.00	6.90	5.22	100.00	100.00
	18	7.61	6.04	100.00	100.00	6.70	5.45	100.00	100.00
	24	9.52	7.77	100.00	100.00	7.70	5.87	100.00	100.00

One can see that for larger q, the significance level of test T_1 and T_2 are greatly affected, while the ASL of T_3 is affected for smaller q and slightly larger than the true level for the rest.

The ASL of test T_4 follows the true level.

From the empirical power we see that T_2 performs much better than T_1, i.e., the unweighted estimator is preferred compared to the weighted estimator given by the MLE.

Comparing the empirical power of all four tests, we see that T_2, T_3 and T_4 are comparable and perform much better than T_1.

Since the test T_4 has both good ASL and power that test should be preferred.

ASL and empirical power of T_1 and T_2 – (I)

$q = 4$	ASL		Power		$q = 10$	ASL		Power	
p, N	T_1	T_2	T_1	T_2	p, N	T_1	T_2	T_1	T_2
5, 8	9.60	7.74	4.96	5.36	10, 12	44.11	44.11	6.79	6.79
5, 12	5.54	5.31	6.58	6.73	10, 16	11.44	11.44	12.67	12.67
5, 16	5.30	5.21	7.59	7.68	10, 20	7.21	7.21	21.22	21.22
5, 32	5.21	5.06	11.03	11.65	10, 30	5.11	5.11	46.89	46.89
10, 12	16.76	5.49	5.67	8.14	20, 22	44.07	6.98	8.23	56.05
10, 16	5.99	5.35	6.89	9.15	20, 26	10.97	5.88	25.21	74.46
10, 20	5.60	5.16	8.36	11.90	20, 32	7.07	5.66	55.95	90.41
10, 30	5.16	4.91	14.09	18.52	20, 40	5.49	5.11	85.42	98.01
20, 22	16.63	5.19	5.12	25.50	30, 32	45.42	5.86	8.63	98.71
20, 26	6.51	4.85	8.99	33.18	30, 36	11.34	5.54	35.46	99.68
20, 32	5.67	5.22	15.65	40.71	30, 40	7.16	5.47	66.10	99.94
20, 40	5.11	5.35	28.55	52.59	30, 50	5.60	5.06	97.19	100.00
50, 52	17.45	5.13	7.28	98.40	50, 52	44.97	5.40	10.96	100.00
50, 56	5.89	5.03	16.34	99.08	50, 56	10.76	5.04	52.40	100.00
50, 70	5.12	5.17	62.91	99.87	50, 70	5.76	5.10	99.65	100.00
50, 100	4.91	5.50	98.30	100.00	50, 100	5.01	5.16	100.00	100.00

ASL and empirical power of T_1 and T_2 – (II)

$q = 4$	ASL		Power		$q = 10$	ASL		Power	
p, N	T_1	T_2	T_1	T_2	p, N	T_1	T_2	T_1	T_2
5, 8	9.53	7.44	5.39	5.37	10, 12	44.66	44.66	6.06	6.06
5, 12	5.60	5.43	5.98	5.94	10, 16	11.22	11.22	9.08	9.08
5, 16	5.56	5.29	5.83	6.21	10, 20	7.40	7.40	12.42	12.42
5, 32	4.82	4.57	8.98	9.25	10, 30	5.92	5.92	22.26	22.26
10, 12	17.36	6.20	5.20	5.62	20, 22	46.44	6.95	6.68	33.22
10, 16	6.31	5.13	5.80	7.26	20, 26	11.38	6.19	15.19	45.01
10, 20	5.76	5.21	6.81	7.67	20, 32	6.56	6.01	36.27	62.33
10, 30	4.86	4.95	9.97	10.78	20, 40	5.97	5.25	60.09	82.00
20, 22	16.80	5.16	5.05	14.19	30, 32	44.73	6.24	7.92	81.63
20, 26	6.34	4.77	7.01	17.73	30, 36	11.07	5.80	24.67	90.41
20, 32	4.88	5.19	12.03	20.39	30, 40	7.50	5.77	43.44	94.57
20, 40	4.80	5.03	18.28	27.11	30, 50	5.64	5.28	82.46	99.33
50, 52	17.24	4.66	6.23	81.24	50, 52	44.86	5.29	9.02	99.99
50, 56	5.70	4.96	12.15	85.12	50, 56	11.25	5.25	33.78	100.00
50, 70	5.27	5.05	41.07	93.37	50, 70	5.67	4.84	95.07	100.00
50, 100	5.28	4.82	88.23	99.22	50, 100	5.23	5.35	100.00	100.00

ASL and empirical power of T_1 and T_2 – (III)

$q = 4$	ASL		Power		$q = 10$	ASL		Power	
p, N	T_1	T_2	T_1	T_2	p, N	T_1	T_2	T_1	T_2
5, 8	9.36	7.25	5.72	5.57	10, 12	43.89	43.89	7.03	7.03
5, 12	5.72	5.70	6.57	6.65	10, 16	10.59	10.59	16.36	16.36
5, 16	5.67	5.40	7.01	7.27	10, 20	7.55	7.55	23.36	23.36
5, 32	4.72	4.71	11.43	11.12	10, 30	5.30	5.30	53.32	53.32
10, 12	16.17	5.63	5.97	8.80	20, 22	44.41	6.57	8.68	68.94
10, 16	6.18	5.54	7.45	10.77	20, 26	11.09	5.95	34.18	86.22
10, 20	5.49	5.33	10.50	14.40	20, 32	6.41	5.35	74.80	97.34
10, 30	5.44	5.46	18.65	22.01	20, 40	5.11	5.12	96.34	99.73
20, 22	16.07	5.04	6.82	27.38	30, 32	44.64	5.82	9.81	99.81
20, 26	6.09	5.09	10.10	33.79	30, 36	11.03	5.17	45.58	99.96
20, 32	5.52	4.96	19.29	43.59	30, 40	7.13	4.83	78.61	100.00
20, 40	5.38	5.02	32.80	56.44	30, 50	6.04	5.19	99.46	100.00
50, 52	16.40	5.01	6.79	99.60	50, 52	44.58	5.41	12.55	100.00
50, 56	5.81	5.10	22.03	99.87	50, 56	11.22	4.78	68.40	100.00
50, 70	5.44	5.39	76.21	99.98	50, 70	5.82	4.93	99.99	100.00
50, 100	4.95	5.44	99.78	100.00	50, 100	5.19	5.09	100.00	100.00

Comparing test T_1 and T_2 for $q = 4$ and $q = 10$ one can see that the ASL of T_1 is greatly affected when p is close to N while the ASL of T_2 is affected just when p is close to N for smaller N and larger q, i.e., when $q = 10$.

One can also see that the empirical power of the test T_2 is similar or better than T_1. The test T_1 has poor power for the cases when p is close to N while T_2 has good power in these cases. It is worth to note that the empirical power of the test T_1 is almost never greater then the empirical power of T_2.

Hence, again the unweighted estimator is preferred compared to the weighted estimator given by the MLE.

ASL and empirical power of T_3 and T_4 – (I)

$q = 4$	ASL		Power		$q = 10$	ASL		Power	
p, N	T_3	T_4	T_3	T_4	p, N	T_3	T_4	T_3	T_4
5, 8	13.30	13.75	6.49	6.66	10, 12	9.84	10.97	22.81	18.86
5, 16	9.92	7.95	9.47	8.88	10, 20	8.32	7.93	41.63	36.91
5, 32	8.06	5.93	16.95	16.00	10, 30	7.54	6.38	65.02	62.63
10, 12	10.88	9.85	8.47	7.86	20, 22	7.76	7.27	93.66	91.20
10, 20	9.05	6.91	11.27	10.93	20, 32	7.35	6.18	99.55	99.34
10, 30	8.56	6.36	16.34	15.24	20, 40	6.94	5.75	99.97	99.96
20, 22	8.42	6.74	28.00	26.19	30, 32	7.28	6.31	99.97	99.94
20, 32	8.05	5.87	41.60	40.64	30, 40	6.83	5.53	100.00	100.00
20, 40	8.04	5.64	52.21	51.33	30, 50	6.84	5.26	100.00	100.00
50, 52	7.32	5.01	98.66	98.39	50, 52	7.05	5.86	100.00	100.00
50, 70	7.50	4.90	99.83	99.84	50, 70	6.71	5.16	100.00	100.00
50, 100	6.92	4.53	100.00	100.00	50, 100	6.92	5.10	100.00	100.00
100, 50	7.84	5.37	100.00	100.00	100, 50	7.43	6.10	100.00	100.00
100, 70	7.70	5.16	100.00	100.00	100, 70	6.99	5.39	100.00	100.00
200, 100	7.13	4.34	100.00	100.00	200, 100	6.99	5.00	100.00	100.00
200, 150	7.29	4.58	100.00	100.00	200, 150	6.86	5.10	100.00	100.00

ASL and empirical power of T_3 and T_4 – (II)

$q = 4$	ASL		Power		$q = 10$	ASL		Power	
p, N	T_3	T_4	T_3	T_4	p, N	T_3	T_4	T_3	T_4
5, 8	14.70	14.09	5.02	4.99	10, 12	10.93	10.68	8.58	8.83
5, 16	10.51	8.59	5.83	5.87	10, 20	8.74	7.25	13.04	14.54
5, 32	8.48	5.64	7.64	8.00	10, 30	7.71	6.08	18.77	23.06
10, 12	10.65	9.80	6.10	5.78	20, 22	8.22	7.12	47.07	47.20
10, 20	9.30	7.11	6.60	7.13	20, 32	7.93	6.46	70.65	71.38
10, 30	8.19	5.66	8.28	8.87	20, 40	7.47	5.82	84.82	84.95
20, 22	9.14	7.13	12.53	12.70	30, 32	6.78	6.03	90.79	89.65
20, 32	7.62	5.62	19.41	19.37	30, 40	6.99	5.75	96.96	96.54
20, 40	7.77	5.38	22.85	23.12	30, 50	7.33	5.76	99.40	99.40
50, 52	7.55	5.41	67.50	66.56	50, 52	7.17	5.74	100.00	100.00
50, 70	7.11	4.50	85.17	85.42	50, 70	7.20	5.47	100.00	100.00
50, 100	7.33	4.71	96.42	96.51	50, 100	6.72	4.88	100.00	100.00
100, 50	7.56	5.12	97.82	97.57	100, 50	7.20	5.34	100.00	100.00
100, 70	7.58	4.94	99.79	99.80	100, 70	6.75	4.94	100.00	100.00
200, 100	6.79	4.29	100.00	100.00	200, 100	6.86	4.94	100.00	100.00
200, 150	6.98	4.59	100.00	100.00	200, 150	6.79	4.72	100.00	100.00

ASL and empirical power of T_3 and T_4 – (III)

$q = 4$	ASL		Power		$q = 10$	ASL		Power	
$p,\ N$	T_3	T_4	T_3	T_4	$p,\ N$	T_3	T_4	T_3	T_4
5, 8	14.26	13.61	6.81	5.88	10, 12	10.22	10.33	19.61	19.59
5, 16	9.98	8.38	8.29	8.41	10, 20	8.65	7.69	35.87	38.91
5, 32	8.40	6.01	13.11	14.44	10, 30	8.14	6.37	60.29	64.69
10, 12	11.40	9.67	7.79	8.02	20, 22	8.41	7.09	92.87	92.49
10, 20	9.36	7.20	9.82	10.33	20, 32	7.86	6.29	99.48	99.55
10, 30	8.74	6.06	13.13	14.62	20, 40	7.14	5.45	99.98	99.99
20, 22	8.76	6.86	28.05	26.93	30, 32	7.83	6.47	99.99	99.98
20, 32	7.96	5.67	46.67	45.95	30, 40	7.30	5.58	100.00	100.00
20, 40	7.66	5.43	58.70	58.21	30, 50	7.00	5.42	100.00	100.00
50, 52	7.94	5.42	99.46	99.35	50, 52	6.96	5.26	100.00	100.00
50, 70	7.30	4.75	99.97	99.96	50, 70	6.99	5.29	100.00	100.00
50, 100	6.75	4.55	100.00	100.00	50, 100	6.34	4.68	100.00	100.00
100, 50	8.31	5.56	100.00	100.00	100, 50	6.78	5.12	100.00	100.00
100, 70	7.10	4.65	100.00	100.00	100, 70	6.30	4.65	100.00	100.00
200, 100	6.88	4.41	100.00	100.00	200, 100	7.01	5.11	100.00	100.00
200, 150	7.05	4.20	100.00	100.00	200, 150	6.72	4.80	100.00	100.00

If we compare T_3 and T_4 we see that the ASL for both the tests are similar meanwhile the ASL for T_4 seems to follow the true level slightly better.

The empirical power of the two test are similar, even if the power for test T_3 is marginally better.

Observe that these two test also work and behave good when the dimension p is larger than N.

Comparing all four tests we see that the test T_2 and T_4 have best ASL. All four test are greatly affected when p is close to N for smaller N and larger q but T_3 and T_4 not as much as T_1 and T_2.

The empirical power of T_2, T_3 and T_4 are similar and all better than T_1.

Concluding Remarks - Mean testing

- The MLE for the mean for a Growth Curve model is a weighted estimator with the inverse of the sample covariance matrix, which is very unstable for p close to N and singular for N less than p. This fact makes the LRT and MLE not suitable for *'large p and small N'*.
- We have modified the MLE to an unbiased and unweighted estimator, just by removing the inverse of the sample covariance matrix.
- We have proposed three new test statistic, which is based on the unweighted estimator.
- We have shown by simulation that these three test statistics, based on the unweighted estimator, are preferred compared to the LRT based on the MLE, i.e., the weighted estimator.

Testing Sphericity

An unbiased and consistent estimator of the covariance matrix $\boldsymbol{\Sigma}$ is given by

$$n\widehat{\boldsymbol{\Sigma}} = \boldsymbol{V} = \boldsymbol{X}\left(\boldsymbol{I}_N - \boldsymbol{P}_{\boldsymbol{C}'}\right)\boldsymbol{X}' \sim W_p\left(\boldsymbol{\Sigma}, n\right),$$

where $n = N - m$, irrespective of which estimator of $\boldsymbol{B}$ is used.

We will now test the hypothesis about sphericity, i.e.,

$$H : \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_p \quad \text{vs.} \quad A : \boldsymbol{\Sigma} > 0.$$

We will give four different test statistics with corresponding asymptotic null distribution.

Standard Likelihood Ratio Test for sphericity

The standard LRT testing the hypothesis H is based on the MLEs.

This test should not be preferred since it is based on a biased estimator of $\boldsymbol{\Sigma}$, and the bias depending on the design matrix $\boldsymbol{A}$.

Anyway, the standard LRT as given by

$$\lambda_1 = \frac{|\boldsymbol{V} + (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}}^{\boldsymbol{V}})\boldsymbol{V}_1(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}}^{\boldsymbol{V}})'|}{\left(\frac{\mathrm{tr}\,\boldsymbol{V} + \mathrm{tr}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}}^{\boldsymbol{V}})\boldsymbol{V}_1}{p}\right)^p},$$

where $\boldsymbol{V}$, $\boldsymbol{V}_1$ and $\boldsymbol{P}_{\boldsymbol{A}}^{\boldsymbol{V}}$ are given above.

One can show that for large N,

$$P_H\left(-\kappa_1 \log \lambda_1 > c\right) = P\left(\chi^2_{f_1} > c\right),$$

where

$$\kappa_1 = n - 2\left(\frac{2p^2+p+2}{12p} - \frac{m(p-q)}{2p} + \frac{qm(p-q)(p+m)}{2p(p^2+p-2)}\right)$$

and $f_1 = \dfrac{p(p+1)}{2} - 1$.

LRT for sphericity based on the unweighted est. of $\boldsymbol{B}$

Using the estimator for $\boldsymbol{\Sigma}$ based on the unweighted estimator for $\boldsymbol{B}$, i.e., $\widehat{\boldsymbol{\Sigma}}_{UW}$, we can propose another LR test statistic.

This will be a modified LRT and not exact. The test statistic is based on

$$\left(\boldsymbol{X} - \boldsymbol{A}\widehat{\boldsymbol{B}}\boldsymbol{C}\right)\left(\boldsymbol{X} - \boldsymbol{A}\widehat{\boldsymbol{B}}\boldsymbol{C}\right) = \boldsymbol{V} + (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}})\boldsymbol{V}_1(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}}).$$

where $\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{A}\right)^{-1}\boldsymbol{A}'$, and is given by

$$\lambda_2 = \frac{|\boldsymbol{V} + (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}})\boldsymbol{V}_1(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}})'|}{\left(\frac{\mathrm{tr}\boldsymbol{V} + \mathrm{tr}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}})\boldsymbol{V}_1}{p}\right)^p}.$$

Given this modified LRT we will assume the same asymptotic distribution and correction factor κ_1 as for λ_1.

LRT for sphericity based only on the matrix $\boldsymbol{V}$

We can also derive a LRT using the fact that the sum of squares matrix is Wishart distributed as $\boldsymbol{V} \sim W_p(\boldsymbol{\Sigma}, n)$.

This test statistic will be the same as under a linear model assumption and is given by

$$\lambda_3 = \frac{|\boldsymbol{V}|}{\left(\frac{\operatorname{tr}\boldsymbol{V}}{p}\right)^p}.$$

One can show that for large N,

$$P_H\left(-\kappa_3 \log \lambda_3 > c\right) = P\left(\chi^2_{f_1} > c\right),$$

where $\kappa_3 = n - \dfrac{2p^2 + p + 2}{6p}$.

Test for sphericity based on $\mathrm{tr}\boldsymbol{V}$ and $\mathrm{tr}\boldsymbol{V}^2$

Following Srivastava (2005) we will give a test statistic based on $\mathrm{tr}\boldsymbol{V}$ and $\mathrm{tr}\boldsymbol{V}^2$. Let a_1 and a_2 be defined as

$$a_1 = \frac{\mathrm{tr}\boldsymbol{\Sigma}}{p} \quad \text{and} \quad a_2 = \frac{\mathrm{tr}\boldsymbol{\Sigma}^2}{p},$$

and let σ_i, $i = 1, \ldots, p$ be the eigenvalues of $\boldsymbol{\Sigma}$.

From the Cauchy-Schwarz inequality, it follows that

$$a_1^2 = \frac{(\sum_{i=1}^{p} \sigma_i)^2}{p} \leq \frac{\sum_{i=1}^{p} \sigma_i^2}{p} = a_2,$$

with equality if and only if $\sigma_1 = \ldots = \sigma_p = \sigma^2$, i.e., if and only if $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_p$.

Srivastava (2005) defined a measure of sphericity given by

$$\lambda_4 = \frac{a_2}{a_1^2} - 1,$$

which is ≥ 0 and takes the value 0 if and only if $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_p$. Srivastava (2005) proposed a test based on unbiased and consistent estimators of a_1 and a_2, given as

$$\hat{a}_1 = \frac{\operatorname{tr} \boldsymbol{V}}{np},$$

$$\hat{a}_2 = \frac{1}{p(n-1)(n+2)} \left(\operatorname{tr} \boldsymbol{V}^2 - \frac{1}{n} (\operatorname{tr} \boldsymbol{V})^2 \right),$$

respectively. Furthermore, the asymptotic distribution for $\hat{a}_1$ is given by Srivastava (2005) and using this we can get an unbiased estimator of a_1^2 as

$$\hat{a}_1^2 - 2\frac{\hat{a}_2}{np},$$

since $\mathrm{E}(\hat{a}_1^2) = \mathrm{var}(\hat{a}_1) + (\mathrm{E}(\hat{a}_1))^2 = 2\frac{a_2}{np} + a_1^2$.

The test statistic that we propose for sphericity is

$$\hat{\lambda}_4 = \frac{\hat{a}_2}{\hat{a}_1^2 - c_f \frac{\hat{a}_2}{np},} - 1,$$

where c_f is a correction factor.

The correction factor should be $c_f = 2$ to make the denominator unbiased, but this will not with certainty make the statistic $\hat{\lambda}_4$ unbiased.

We have observed through simulation that $c_f = 1$ is a better choice.

Under the null hypothesis that $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_p$, and $n = \mathcal{O}(p^\delta)$, $\delta > 1/2$, asymptotically as $(n, p) \to \infty$

$$\frac{n}{2}\hat{\lambda}_4 \sim N(0, 1).$$

Note that it is a one-sided test for testing the hypothesis that $\lambda_4 = 0$ vs. $\lambda_4 > 0$.

Also, note that the test statistic based on $\hat{\lambda}_4$ can be performed for all values of n and p as opposed to the test based on the likelihood ratio test which requires that $n \geq p$.

Compare the performance

To compare the performance for the different tests we can compute the attained significance level (ASL) and the empirical power.

Let c be the critical value from the distribution considered for the test statistics.

With $10,000$ simulated replications under the null hypothesis, the ASL is computed as

$$\widehat{\alpha} = \frac{(\# \text{ of } t_H \geq c)}{(\# \text{ simulated replications})},$$

where t_H is the values of the test statistics derived from the simulated data under the null hypothesis. We set the nominal significance level to $\alpha = 5\%$.

For the simulations let

$$\boldsymbol{A}_1 = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & p \end{pmatrix}'$$

or $\boldsymbol{A}_2 = (a_{ij})$, where $a_{ij} \sim U(0,2)$, $i = 1, ..., p$, $j = 1, 2$ and

$$\boldsymbol{C} = \begin{pmatrix} \mathbf{1}'_{N_1} & \mathbf{0}'_{N_2} \\ \mathbf{0}'_{N_1} & \mathbf{1}'_{N_2} \end{pmatrix},$$

i.e., with $q = 2$ and $m = 2$. For simplicity we will choose N even and $N_1 = N_2 = N/2$.

For the power simulations let $\boldsymbol{D} = diag(\sigma_1, \ldots, \sigma_p)$ and define $\sigma_i = 2 + (p - i + 1)/p$, for $i = 1, \ldots, p$, Also, let $\boldsymbol{R} = (\rho_{ij})$, where $\rho_{ij} = (-1)^{i+j} r^{|i-j|^f}$. Under the alternatives we will assume the covariance matrix

$$\boldsymbol{\Sigma}_A = \boldsymbol{DRD}, \quad \text{with } r = 0.2 \text{ and } f = 0.1.$$

The empirical power is calculated from $10,000$ new replications simulated under the alternative hypothesis when $\mathbf{\Sigma} = \mathbf{\Sigma}_A$.

Let t_A be the value of the test statistic derived from the simulated data under the alternative hypothesis. The empirical power is given as

$$\widehat{\beta} = \frac{(\# \text{ of } t_A \geq \widehat{c})}{(\# \text{ simulated replications})},$$

where $\widehat{c}$ is the estimated critical value calculated from the simulated data under the null hypothesis, i.e., the critical value calculated from the empirical null distribution.

		$\boldsymbol{A}_1$							
		ASL				Power			
N	p	λ_1	λ_2	λ_3	λ_4	λ_1	λ_2	λ_3	λ_4
8	4	8.17	2.01	6.90	3.99	10.67	12.90	9.92	14.11
8	6	33.15	2.45	25.18	4.55	7.22	13.32	6.67	17.24
8	16	-	-	-	4.90	-	-	-	33.03
12	3	4.96	3.24	4.91	4.20	19.76	20.86	16.89	17.86
12	6	7.80	2.92	6.63	4.66	23.16	26.86	20.68	29.73
12	9	27.43	5.34	22.52	5.19	17.21	25.97	15.90	38.13
12	24	-	-	-	4.98	-	-	-	63.47
20	5	4.90	3.44	4.88	5.22	49.04	50.25	43.69	47.77
20	10	8.30	4.14	7.37	5.23	56.61	60.42	51.24	70.82
20	15	27.21	11.71	26.10	5.62	48.19	55.88	43.05	80.65
20	40	-	-	-	5.19	-	-	-	96.49
60	15	5.71	4.93	5.59	5.88	99.99	99.99	99.98	99.99
60	45	57.03	49.96	61.70	5.83	100.00	100.00	99.99	100.00
60	120	-	-	-	5.03	-	-	-	100.00
100	25	5.89	5.45	5.84	5.31	100.00	100.00	100.00	100.00
100	50	15.77	14.45	16.03	5.50	100.00	100.00	100.00	100.00
100	75	88.51	86.06	90.46	5.50	100.00	100.00	100.00	100.00
100	200	-	-	-	5.15	-	-	-	100.00
200	50	6.23	5.87	6.31	5.02	100.00	100.00	100.00	100.00
200	100	31.76	30.42	32.32	4.96	100.00	100.00	100.00	100.00
200	400	-	-	-	4.80	-	-	-	100.00

		$\boldsymbol{A}_2$							
		ASL				Power			
N	p	λ_1	λ_2	λ_3	λ_4	λ_1	λ_2	λ_3	λ_4
8	4	8.02	1.85	6.52	4.01	13.15	16.95	10.61	13.20
8	6	33.25	2.64	25.12	4.59	7.06	12.09	6.95	18.11
8	16	-	-	-	5.10	-	-	-	32.13
12	3	5.09	3.43	5.32	4.23	16.27	15.18	15.88	17.68
12	6	8.16	2.77	6.96	4.73	22.06	26.52	19.32	29.77
12	9	27.77	5.58	22.86	5.36	17.44	24.67	15.43	37.57
12	24	-	-	-	4.94	-	-	-	63.77
20	5	5.48	3.63	5.29	5.34	45.67	46.68	42.25	47.94
20	10	8.33	4.43	7.34	5.65	57.70	60.23	52.83	69.89
20	15	26.94	11.52	25.42	5.94	48.73	55.91	44.15	80.95
20	40	-	-	-	5.21	-	-	-	96.31
60	15	5.83	5.08	5.77	6.29	99.99	99.99	99.99	100.00
60	45	56.71	49.43	60.46	5.82	99.99	100.00	99.97	100.00
60	120	-	-	-	5.71	-	-	-	100.00
100	25	5.51	5.04	5.68	5.16	100.00	100.00	100.00	100.00
100	50	15.29	13.90	15.55	5.19	100.00	100.00	100.00	100.00
100	75	88.02	85.74	90.23	5.26	100.00	100.00	100.00	100.00
100	200	-	-	-	5.08	-	-	-	100.00
200	50	6.62	6.34	6.49	5.58	100.00	100.00	100.00	100.00
200	100	32.20	31.06	32.90	5.23	100.00	100.00	100.00	100.00
200	400	-	-	-	5.51	-	-	-	100.00

Concluding Remarks – Sphericity

We see that when testing sphericity, λ_4 seems to be the best.

For small p compared to N there is no really difference, but when p is larger, λ_4 is definitely better with controlled ASL and good power.

Observe also that the test statistic λ_4 works fine for high dimensions, i.e., when $p > N$.

Testing intraclass covariance structure

Now we consider the hypothesis testing intraclass (IC) covariance structure, i.e.,

$$H : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{IC} \equiv \sigma^2((1-\rho)\boldsymbol{I}_p + \rho\mathbf{1}\mathbf{1}') \quad \text{vs.} \quad A : \boldsymbol{\Sigma} > 0,$$

with $-\dfrac{1}{p-1} < \rho < 1$.

We will give four different test statistics with corresponding asymptotic null distribution.

Standard LRT for IC covariance structure

The standard LRT, testing the hypothesis H is based on the MLEs, given above. This test should not be preferred, since it is not based on a unbiased estimator of $\boldsymbol{\Sigma}$.

Anyway, the standard LRT is given by Khatri (1973) as

$$\gamma_1 = \frac{|\boldsymbol{V} + (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}}^{\boldsymbol{V}})\boldsymbol{V}_1(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}}^{\boldsymbol{V}})'|}{\left(\frac{\mathbf{1}'\boldsymbol{V}\mathbf{1}}{p}\right)\left(\frac{\mathrm{tr}\boldsymbol{P}_{\mathbf{1}'}\boldsymbol{V} + \mathrm{tr}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}})\boldsymbol{V}_1}{p-1}\right)^{p-1}}.$$

As usual, one can show that for large N that

$$P_H\left(-n\log\gamma_1 > c\right) = P\left(\chi^2_{f_2} > c\right),$$

where $f_2 = \dfrac{p(p+1)}{2} - 2$.

LRT for IC covariance structure based on the unweighted estimator of $\boldsymbol{B}$

Again we can use the estimator of $\boldsymbol{\Sigma}$ based on the unweighted estimator for $\boldsymbol{B}$, i.e., the estimator $\widehat{\boldsymbol{\Sigma}}_{UW}$.

The modified LRT statistic is given by

$$\gamma_2 = \frac{|\boldsymbol{V} + (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}})\boldsymbol{V}_1(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}})'|}{\left(\frac{\boldsymbol{1}'\boldsymbol{V}\boldsymbol{1}}{p}\right)\left(\frac{\mathrm{tr}\boldsymbol{P}_{\boldsymbol{1}'}\boldsymbol{V} + \mathrm{tr}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A}})\boldsymbol{V}_1}{p-1}\right)^{p-1}}.$$

We assume the same asymptotic distribution for γ_2 as for γ_1 above.

LRT for IC structure based only on the matrix $\boldsymbol{V}$

We can also derive a LRT using the fact that $\boldsymbol{V} \sim W_p(\boldsymbol{\Sigma}, n)$.

This test statistic will be the same as under a linear model assumption and is given by

$$\gamma_3 = \frac{|\boldsymbol{V}|}{\left(\frac{\mathbf{1}'\boldsymbol{V}\mathbf{1}}{p}\right)\left(\frac{p\operatorname{tr}\boldsymbol{V} - \mathbf{1}'\boldsymbol{V}\mathbf{1}}{p(p-1)}\right)^{p-1}}.$$

For large N, one can shown that

$$P_H\left(-\nu_3 \log \gamma_3 > c\right) = P\left(\chi^2_{f_2} > c\right),$$

where $\nu_3 = n - \frac{p(p+1)^2(2p-3)}{6(p-1)(p^2+p-4)}$.

Test for intraclass covariance structure based on a measure of sphericity

Let the matrix $\boldsymbol{Q}$ be an orthogonal matrix of order p and let the first column be a normalized column of ones, i.e.,

$$\boldsymbol{Q} = \begin{pmatrix} p^{-1/2}\mathbf{1}_p & \boldsymbol{Q}_2 \end{pmatrix} : p \times p.$$

Given the transformation $\boldsymbol{X}^* = \boldsymbol{Q}'\boldsymbol{X}$ we have the following model

$$\boldsymbol{X}^* \sim N_{p,N}\left(\boldsymbol{A}^*\boldsymbol{B}\boldsymbol{C}, \boldsymbol{\Sigma}^*, \boldsymbol{I}_N\right),$$

where

$$\begin{aligned} \boldsymbol{A}^* &= \boldsymbol{Q}'\boldsymbol{A}, \\ \boldsymbol{\Sigma}^* &= \boldsymbol{Q}'\boldsymbol{\Sigma}\boldsymbol{Q} = \begin{pmatrix} \sigma_{11}^* & \boldsymbol{\sigma}_{12}^* \\ \boldsymbol{\sigma}_{21}^* & \boldsymbol{\Sigma}_{22}^* \end{pmatrix}, \end{aligned}$$

and $\boldsymbol{\sigma}_{12}^* = (\boldsymbol{\sigma}_{21}^*)' : 1 \times (p-1)$, $\boldsymbol{\Sigma}_{22}^* : (p-1) \times (p-1)$.

However, under the null hypothesis H_2 we have

$$\boldsymbol{\Sigma}^* = \boldsymbol{Q}'\boldsymbol{\Sigma}_{IC}\boldsymbol{Q} = \begin{pmatrix} \sigma^2(1+(p-1)\rho) & \mathbf{0}' \\ \mathbf{0} & \sigma^2(1-\rho)\boldsymbol{I}_{p-1} \end{pmatrix}.$$

Thus, instead of direct test H we will test the hypothesis

$$H : \boldsymbol{\Sigma}_{22}^* = \tilde{\sigma}^2\boldsymbol{I}_{p-1} \quad \text{vs.} \quad A : \boldsymbol{\Sigma}_{22}^* > 0,$$

for some $\tilde{\sigma}^2$.

Hypothesis H is tested using the same procedure as for testing sphericity above.

Compare the performance

Again, for the simulations let

$$\boldsymbol{A}_1 = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & p \end{pmatrix}'$$

or $\boldsymbol{A}_2 = (a_{ij})$, where $a_{ij} \sim U(0,2)$, $i = 1, ..., p$, $j = 1, 2$ and

$$\boldsymbol{C} = \begin{pmatrix} \mathbf{1}'_{N_1} & \mathbf{0}'_{N_2} \\ \mathbf{0}'_{N_1} & \mathbf{1}'_{N_2} \end{pmatrix},$$

i.e., with $q = 2$ and $m = 2$. For simplicity we will choose N even and $N_1 = N_2 = N/2$.

For the power simulations let $\boldsymbol{D} = diag(\sigma_1, \ldots, \sigma_p)$ and define $\sigma_i = 2 + (p - i + 1)/p$, for $i = 1, \ldots, p$, Also, let $\boldsymbol{R} = (\rho_{ij})$, where $\rho_{ij} = (-1)^{i+j} r^{|i-j|^f}$. Under the alternatives we will assume the covariance matrix

$$\boldsymbol{\Sigma}_A = \boldsymbol{DRD}, \quad \text{with } r = 0.2 \text{ and } f = 0.1.$$

		$\boldsymbol{A}_1$							
		ASL				Power			
N	p	γ_1	γ_2	γ_3	γ_4	γ_1	γ_2	γ_3	γ_4
12	3	9.13	6.26	5.03	3.12	18.38	28.13	16.10	11.35
12	4	10.66	5.93	5.17	3.93	19.93	29.44	18.98	20.40
12	6	19.13	9.01	6.91	4.83	20.52	18.60	19.19	28.11
12	24	-	-	-	5.04	-	-	-	63.87
24	6	8.36	6.18	5.23	5.27	60.53	53.58	57.87	61.42
24	8	12.36	8.63	5.50	5.36	67.69	46.71	63.59	73.10
24	12	30.56	22.41	7.98	5.46	71.88	33.11	67.12	84.05
24	48	-	-	-	4.90	-	-	-	99.25
48	12	11.92	10.01	5.26	5.72	99.59	82.08	99.43	99.71
48	16	22.64	19.64	6.13	5.99	99.83	79.91	99.74	99.95
48	24	71.34	66.74	9.69	5.38	99.90	80.45	99.85	99.99
48	96	-	-	-	5.12	-	-	-	100.00
96	24	24.46	23.14	5.55	5.26	100.00	99.84	100.00	100.00
96	32	57.37	55.46	7.15	5.67	100.00	99.99	100.00	100.00
96	48	99.84	99.78	15.03	5.30	100.00	100.00	100.00	100.00
96	192	-	-	-	5.56	-	-	-	100.00
192	48	64.78	63.99	6.22	5.37	100.00	100.00	100.00	100.00
192	64	98.47	98.35	8.42	5.49	100.00	100.00	100.00	100.00
192	96	100.00	100.00	31.05	5.36	100.00	100.00	100.00	100.00
192	384	-	-	-	5.09	-	-	-	100.00

		$\boldsymbol{A}_2$							
		ASL				Power			
N	p	γ_1	γ_2	γ_3	γ_4	γ_1	γ_2	γ_3	γ_4
12	3	11.40	6.53	5.19	3.19	16.72	23.28	16.46	11.25
12	4	12.25	6.29	5.15	4.01	14.82	22.95	18.36	19.69
12	6	22.35	9.06	7.02	4.48	18.76	31.58	19.29	28.40
12	24	-	-	-	5.12	-	-	-	63.79
24	6	10.39	6.80	5.62	5.65	56.51	60.33	56.47	59.62
24	8	15.75	8.64	5.35	5.15	61.58	68.80	63.60	72.86
24	12	33.49	22.85	7.91	5.77	69.61	71.31	68.42	84.62
24	48	-	-	-	5.62	-	-	-	99.15
48	12	15.32	10.66	5.42	5.79	99.39	99.24	99.41	99.62
48	16	28.67	20.44	5.88	5.73	99.73	99.56	99.74	99.90
48	24	77.86	68.51	9.77	5.15	99.77	99.75	99.80	99.99
48	96	-	-	-	5.53	-	-	-	100.00
96	24	36.44	25.51	5.73	5.45	100.00	100.00	100.00	100.00
96	32	64.05	56.28	7.05	5.60	100.00	100.00	100.00	100.00
96	48	99.87	99.80	14.97	5.21	100.00	100.00	100.00	100.00
96	192	-	-	-	4.90	-	-	-	100.00
192	48	76.61	68.33	6.58	5.23	100.00	100.00	100.00	100.00
192	64	99.05	98.43	8.56	5.28	100.00	100.00	100.00	100.00
192	96	100.00	100.00	31.90	5.42	100.00	100.00	100.00	100.00
192	384	-	-	-	5.18	-	-	-	100.00

Concluding Remarks – IC covariance structure

For testing the intraclass covariance structure we see that test statistic γ_4 controls the ASL better then the others.

Also the power is better with γ_4 for most of the cases.

Note also that γ_3 seems to behave pretty well with controlled ASL and good power.

Observe again that test statistic γ_4 works fine for high dimensions.

Linköping University - Research that makes a difference

LINKÖPING UNIVERSITY