Estimation and testing in multivariate linear models
Part I

Daniel Klein

Institute of Mathematics
Faculty of Science
P. J. Šafárik University in Košice

Będlewo
November 29, 2021

Introduction

Two main problems to solve in data analysis

determine the values of unknown parameters

- data contain unexplained fluctuations or noise, usually some degree of measurement error is present, therefore only some estimates of unknown population parameters can be obtained,

test the hypotheses about the values of unknown parameters

- means to decide whether data are consistent at some level of agreement with a particular population parameter.

Introduction

While collecting data one or more quantities on each sample unit are measured.

Sometimes it could be better to isolate each quantity in the system to study them separately - thus leading to the so called univariate model.

However, the quantities may be influenced by each other to such an extent that the separate analysis would offer poor information about the whole system - the multivariate models, which study all quantities simultaneously, come to foreground.

Comparing to univariate setting more mathematics is required to derive multivariate statistical techniques for making inferences.

Introduction

Linear models

- still the main tool of the applied statistics even if many modern innovative statistical techniques (which often need computer assistance) are at hand
- intensively studied ever since the times of Gauss (1777-1855) and Legendre (1752-1833)
- a huge amount of statistical literature has been published concerning various aspects of linear models and statistical methods associated with them
- acquired their popularity because of nice properties and at the same time they appeared to be versatile and robust enough

Linear (univariate) model

Standard texts on univariate linear models

Arnold, S. (1981). *The theory of linear models and multivariate analysis*, Wiley, New York.

Christensen, R. (1996). *Plane answers to complex questions: The theory of linear models*, Springer-Verlag, New York.

Graybill, F. A. (1976). *Theory and applications of the linear model*, Duxbury Press, North Scituate, MA.

Linear (univariate) model

Linear model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e},$$

$\mathbf{y} = (y_1, \cdots, y_n)'$	vector of observations
$\mathbf{X}_{n \times k}$	known design matrix (of full rank)
$\beta_{k \times 1}$	vector of unknown parameters (of first order), $k < n$
$\mathbf{e}_{n \times 1}$	vector of random errors

Linear (univariate) model

Linear model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e},$$

$\mathbf{y} = (y_1, \cdots, y_n)'$	vector of observations
$\mathbf{X}_{n \times k}$	known design matrix (of full rank)
$\beta_{k \times 1}$	vector of unknown parameters (of first order), $k < n$
$\mathbf{e}_{n \times 1}$	vector of random errors

Usual assumption:

$$\mathbf{e} \sim N_n(\mathbf{0}, \mathbf{\Omega}),$$

$\mathbf{\Omega}_{n \times n}$ matrix of unknown parameters (of second order)

Linear (univariate) model

Linear model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e},$$

$\mathbf{y} = (y_1, \cdots, y_n)'$	vector of observations
$\mathbf{X}_{n\times k}$	known design matrix (of full rank)
$\beta_{k\times 1}$	vector of unknown parameters (of first order), $k < n$
$\mathbf{e}_{n\times 1}$	vector of random errors

Usual assumption:

$$\mathbf{e} \sim N_n(\mathbf{0}, \mathbf{\Omega}),$$

$\mathbf{\Omega}_{n\times n}$	matrix of unknown parameters (of second order)

Basic problem:

inference about the unknown parameters of first order

Linear (univariate) model

Number of unknown parameters:

$$
\begin{array}{ll}
\boldsymbol{\beta}_{k\times 1} & k \\
\boldsymbol{\Omega}_{n\times n} & \frac{n(n+1)}{2}
\end{array}
$$

Linear (univariate) model

Number of unknown parameters:

$$\left.\begin{array}{ll} \beta_{k\times 1} & k \\ \mathbf{\Omega}_{n\times n} & \frac{n(n+1)}{2} \end{array}\right\}\text{overparametrized}$$

Linear (univariate) model

Number of unknown parameters:

$\beta_{k\times 1}$	k
$\mathbf{\Omega}_{n\times n}$	1

Structure of covariance matrix:

$$\mathbf{\Omega} = \sigma^2 \mathbf{I}_n, \quad \text{or} \quad \mathbf{\Omega} = \sigma^2 \mathbf{V}, \text{ with } \mathbf{V} \text{ known}$$

σ^2 unknown variance parameter

Linear (univariate) model

On each independent sampling unit usually one response scalar variable is measured.

Generalization:

> assume univariate model for p response variables, i.e. more than one response variable is measured on each sampling unit - one needs to take into account the dependence of measurements on the same sampling unit

Linear (univariate) model

Examples:

- p air pollutants (CO, NO, etc.) measured on n widely-separated days
- p test scores (e.g. different subjects) for n different students

Linear (univariate) model

Examples:

- p air pollutants (CO, NO, etc.) measured on n widely-separated days
- p test scores (e.g. different subjects) for n different students

Requirement - the same design matrix applies to every response and every independent sampling unit has the same set of response variables

Linear (univariate) model

For j-th response variable we have

$$\begin{array}{l} \mathbf{y}_j = \mathbf{X}\beta_j + \varepsilon_j, \\ \mathrm{E}[\varepsilon_j] = \mathbf{0}, \quad \mathrm{Var}[\varepsilon_j] = \omega_{jj}^2 \mathbf{I}_n, \end{array} \qquad j = 1, \dots, p.$$

$\mathbf{y}_j = (y_{1j}, \cdots, y_{nj})'$	vector of observations of j-th variable
$\mathbf{X}_{n\times k}$	known design matrix (of full rank)
β_j	k-dimensional vector of unknown parameters (of first order) for j-th variable, $k < n$
ε_j	n-dimensional vector of random errors

These models are related by the $p(p-1)/2$ covariances

$$\mathrm{Cov}[\mathbf{y}_i, \mathbf{y}_j] = \omega_{ij}^2 \mathbf{I}_n$$

Multivariate linear model

The classical multivariate model

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E},$$
$$\mathrm{E}\left[\mathbf{E}\right] = 0, \text{and } \mathrm{Var}[\mathbf{E}] = \mathrm{Var}[\mathrm{vec}\,\mathbf{E}] = \boldsymbol{\Omega} \otimes \mathbf{I}_n,$$

$\mathbf{Y}_{n\times p} = (\mathbf{y}_1, \cdots, \mathbf{y}_p)$	matrix of observations
$\mathbf{X}_{n\times k}$	known design matrix (of full rank)
$\mathbf{B}_{k\times p} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)$	matrix of unknown parameters (of first order), $k < n$
$\mathbf{E}_{n\times p} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_p)$	matrix of random errors
$\boldsymbol{\Omega}_{p\times p} = (\omega_{ij}^2)_{ij}$	matrix of unknown (second order) parameters

Usual assumptions:

- $\mathbf{E}$ follows a matrix normal distribution $N_{n,p}(\mathbf{0}, \mathbf{I}_n, \boldsymbol{\Omega})$
- $n > p + r(\mathbf{X})$

Multivariate linear model

This model can be viewed as a sample of n independent p-variate observations

- denoting $\mathbf{y}_i'$ and $\mathbf{x}_i'$ as i-th row of respectively $\mathbf{Y}$ and $\mathbf{X}$

$$\mathbf{y}_i \sim N_p(\mathbf{B}'\mathbf{x}_i, \mathbf{\Omega}), \quad i = 1, \ldots, n$$

Multivariate linear model

This model can be viewed as a sample of n independent p-variate observations

- denoting $\mathbf{y}_i'$ and $\mathbf{x}_i'$ as i-th row of respectively $\mathbf{Y}$ and $\mathbf{X}$

$$\mathbf{y}_i \sim N_p(\mathbf{B}'\mathbf{x}_i, \mathbf{\Omega}), \quad i = 1, \ldots, n$$

Different arrangement of data

- horizontal ("classical") - independent p-variate observations are arranged in horizontal position one below another as rows, thus forming an $n \times p$ observation matrix $\mathbf{Y}$ - preserves the direction of stacking observations in the univariate linear model

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1' \\ \mathbf{y}_2' \\ \vdots \\ \mathbf{y}_n' \end{pmatrix}, \quad \text{thus} \quad \underset{n \times p}{\mathbf{Y}} = \underset{n \times k}{\mathbf{X}} \underset{k \times p}{\mathbf{B}} + \underset{n \times p}{\mathbf{E}}$$

Multivariate linear model

This model can be viewed as a sample of n independent p-variate observations

- denoting $\mathbf{y}_i'$ and $\mathbf{x}_i'$ as i-th row of respectively $\mathbf{Y}$ and $\mathbf{X}$

$$\mathbf{y}_i \sim N_p(\mathbf{B}'\mathbf{x}_i, \boldsymbol{\Omega}), \quad i = 1, \ldots, n$$

Different arrangement of data

- vertical - independent p-variate observations are arranged as a columns one next to another forming a $p \times n$ observation matrix $\mathbf{Y}$ - it is the transposition of the previous arrangement

$$\mathbf{Y}_* = \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_n \end{pmatrix}, \quad \text{thus} \quad \underset{p\times n}{\mathbf{Y}_*} = \underset{p\times k}{\mathbf{B}_*}\,\underset{k\times n}{\mathbf{X}_*} + \underset{p\times n}{\mathbf{E}_*},$$

Matrix normal distribution

For $\mathbf{Y}_* = \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_n \end{pmatrix}$

- the mean is $\mathrm{E}[\mathbf{Y}_*] = \mathbf{B}'\mathbf{X}'$, thus $\mathbf{B}_* = \mathbf{B}'$ and $\mathbf{X}_* = \mathbf{X}'$
- the distribution of $\operatorname{vec} \mathbf{Y}_*$ is

$$\operatorname{vec} \mathbf{Y}_* = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix} \sim N_{pn}(\operatorname{vec}(\mathbf{B}'\mathbf{X}'), \mathbf{I}_n \otimes \boldsymbol{\Omega}),$$

  where $\otimes$ denotes the Kronecker product.

We write this as

$$\mathbf{Y}_* \sim N_{p,n}(\mathbf{B}'\mathbf{X}', \boldsymbol{\Omega}, \mathbf{I}_n)$$

Matrix normal distribution

Since $\mathbf{Y} = \mathbf{Y}_*'$, for $\mathbf{Y}$

- the mean is $\mathrm{E}\left[\mathbf{Y}\right] = \mathrm{E}\left[\mathbf{Y}_*'\right] = \mathbf{XB}$
- the distribution of $\operatorname{vec}\mathbf{Y}$ is

$$\operatorname{vec}\mathbf{Y} = \mathbf{K}_{p,n}\operatorname{vec}\mathbf{Y}' \sim N_{pn}\Big(\mathbf{K}_{p,n}\operatorname{vec}(\mathbf{B}'\mathbf{X}'), \mathbf{K}_{p,n}(\mathbf{I}_n \otimes \boldsymbol{\Omega})\mathbf{K}_{n,p}\Big),$$

where $\mathbf{K}_{p,n}$ is vector a commutation matrix.

Matrix normal distribution

Since $\mathbf{Y} = \mathbf{Y}_*'$, for $\mathbf{Y}$

- the mean is $\mathrm{E}[\mathbf{Y}] = \mathrm{E}[\mathbf{Y}_*'] = \mathbf{XB}$
- the distribution of $\operatorname{vec}\mathbf{Y}$ is

$$\operatorname{vec}\mathbf{Y} = \mathbf{K}_{p,n}\operatorname{vec}\mathbf{Y}' \sim N_{pn}\Big(\underbrace{\mathbf{K}_{p,n}\operatorname{vec}(\mathbf{B}'\mathbf{X}')}_{=\operatorname{vec}\mathbf{XB}}, \underbrace{\mathbf{K}_{p,n}(\mathbf{I}_n \otimes \mathbf{\Omega})\mathbf{K}_{n,p}}_{=\mathbf{\Omega}\otimes\mathbf{I}_n}\Big),$$

where $\mathbf{K}_{p,n}$ is vector a commutation matrix.

Matrix normal distribution

Since $\mathbf{Y} = \mathbf{Y}'_*$, for $\mathbf{Y}$

- the mean is $\mathrm{E}[\mathbf{Y}] = \mathrm{E}[\mathbf{Y}'_*] = \mathbf{XB}$
- the distribution of $\operatorname{vec}\mathbf{Y}$ is

$$\operatorname{vec}\mathbf{Y} = \mathbf{K}_{p,n}\operatorname{vec}\mathbf{Y}' \sim N_{pn}\Big(\operatorname{vec}(\mathbf{XB}), \boldsymbol{\Omega} \otimes \mathbf{I}_n\Big),$$

where $\mathbf{K}_{p,n}$ is vector a commutation matrix.

We write this as

$$\mathbf{Y} \sim N_{n,p}(\mathbf{XB}, \mathbf{I}_n, \boldsymbol{\Omega})$$

Multivariate linear model

The model

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E},$$
$$\mathbf{E} \sim N_{n,p}(\mathbf{0}, \mathbf{I}_n, \mathbf{\Omega}),$$

is very general one. Covers:

- Multivariate regression model

  $\mathbf{X}$ is a matrix of regression constants, usually $\mathbf{1}_n \in \mathcal{C}(\mathbf{X})$

- MANOVA model

  $\mathbf{X}$ is a 0-1 design matrix

- General mean model

  $\mathbf{X} = \mathbf{1}_n$, i.e. $\mathbf{B}' = \boldsymbol{\mu}$ is p-dimensional vector

Multivariate linear model

The model

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E},$$
$$\mathbf{E} \sim N_{n,p}(\mathbf{0}, \mathbf{I}_n, \boldsymbol{\Omega}),$$

is very general one. Covers:

- Generalized MANOVA model (so called growth curve model) introduced by Potthoff and Roy (1964)

  considering a model for i-th row β_i' of unknown matrix $\mathbf{B}$

  $$\underset{p\times 1}{\beta_i} = \underset{p\times r}{\mathbf{Z}}\ \underset{r\times 1}{\mathbf{b}_i}, \quad i = 1, \cdots, k.$$

  - Growth curve model (GCM):

  $$\mathbf{Y} = \mathbf{XBZ}' + \mathbf{E}, \quad \text{where} \quad \mathbf{B} = \begin{pmatrix} \mathbf{b}_1' \\ \vdots \\ \mathbf{b}_k' \end{pmatrix}.$$

  - This model is reduced to general form if $\mathbf{Z} = \mathbf{I}_p$.

Two-level multivariate model

Introduction

Classical multivariate linear model - sample of vector valued (p dimensional, say) observation vectors

Variety of areas of application require the extension to matrix valued, or even more complex multivariate data

- based on subvectors of correlated components which follow from differences across characteristics, locations, time or depths, i.e., several characteristics can be observed on more than one response variable at different locations, repeatedly over time, at different depths, etc.

Introduction

Multi-level (k-level, say) multivariate data - can be presented in the form of a multi-index matrix $\mathcal{Y}$ (i.e. tensor of order k)

One can matricize the tensor $\mathcal{Y}$ - classical multivariate model.

Subvectors may have variances and covariances that differ across locations, time and depths - the covariance matrix $\mathbf{\Omega}$ is subject to some restrictions, which may impose some structure $\mathbf{\Omega}$

Introduction

One can ignore this structure treating $\boldsymbol{\Omega}$ as unstructured (UN) - this may cause overparametrization problems

> Crowder and Hand (1990) - in case of small samples unstructered covariance matrix can result in rather weak inference, in the sense that too many degrees of freedom are used up in estimating the covariance parameters, leaving too few for the parameters of interest.

Several sources of variability (characteristics, time, location, etc.) in tensor $\boldsymbol{\mathcal{Y}}$ - naturally multi-separable covariance structure

- $\boldsymbol{\Omega}$ can be written as the Kronecker product of several variance-covariance matrices

Two-level multivariate data

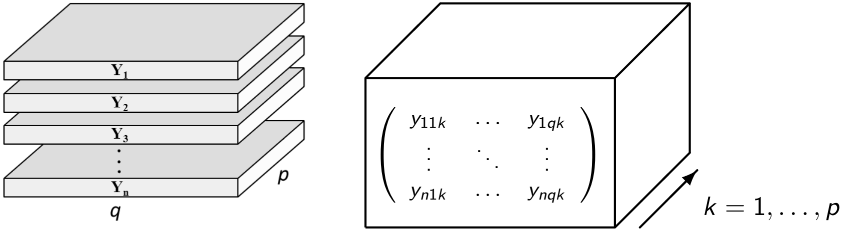
Matrix-valued random variable $\mathbf{Y}_i$

$$q \text{ characteristics} \quad \overset{p \text{ time points}}{\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{q1} & y_{q2} & \cdots & y_{qp} \end{pmatrix}} \quad \text{, with } \mathrm{E}\left[\mathbf{Y}_i\right] = \mathbf{M}_{q\times p} \text{ and } \mathrm{Var}[\operatorname{vec}\mathbf{Y}_i] = \mathbf{\Omega}$$

Two-level multivariate data

Sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ may by presented in a tensor of order three $\mathcal{Y} = (y_{ijk})$ - we have three possible directions of arrangement

- matrices one behind another - forming $q \times p \times n$ tensor
- matrices as vertical slices one next to another - forming $q \times n \times p$ tensor
- matrices as horizontal slices one below another - forming $n \times q \times p$ tensor



Two-level multivariate data

Vectorization of a tensor $\mathcal{Y}$ according to Kolda and Bader (2009) or Singull et al. (2012)

$$\operatorname{vec}\mathcal{Y} = \sum_{i=1}^{n}\sum_{j=1}^{q}\sum_{k=1}^{p} y_{ijk}\,\mathbf{e}_{k:p}\otimes\mathbf{e}_{j:q}\otimes\mathbf{e}_{i:n},$$

where $\mathbf{e}_{i:n}$ is i-th column of $\mathbf{I}_n$.

We assume normality. i.e.

$$\operatorname{vec}\mathcal{Y} \sim N_{nqp}(\operatorname{vec}\mathcal{M}, \boldsymbol{\Omega}\otimes\mathbf{I}_n), \qquad \text{with } \mathcal{M}\in\mathbb{R}^{n\times q\times p}.$$

Two sources of variability - the dispersion of observations naturally separated for rows and columns

Two-level multivariate data

Connection with the classical multivariate model - rearrange the data to transform tensor $\mathcal{Y}$ into a matrix form

referred to as *matricization* (*unfolding* or *flattening*)

Three possibilities of matricization for a three-dimensional tensor $\mathcal{Y}_{n\times q\times p}$

- matricization of $\mathcal{Y}$ with respect to *nq*-mode slices, $\mathbf{Y}_{\cdot\cdot k}$, given side by side - algebraically

$$\mathbf{Y} = \sum_{i=1}^{n}\sum_{j=1}^{q}\sum_{k=1}^{p} y_{ijk}\, \mathbf{e}_{i:n}(\mathbf{e}'_{k:p} \otimes \mathbf{e}'_{j:q}).$$

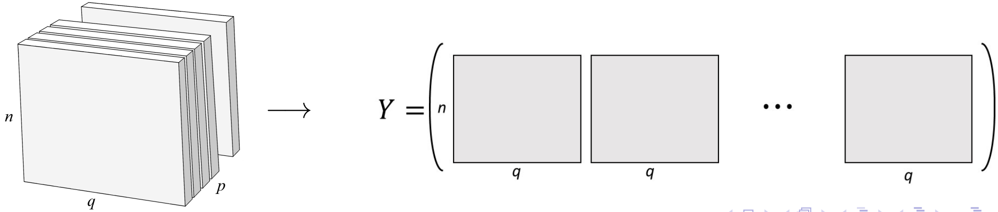
Two-level multivariate data

Connection with the classical multivariate model - rearrange the data to transform tensor $\mathcal{Y}$ into a matrix form

referred to as *matricization* (*unfolding* or *flattening*)

Three possibilities of matricization for a three-dimensional tensor $\mathcal{Y}_{n\times q\times p}$

- vectorize and transpose each $\mathbf{Y}_i$ and write the result as rows underneath forming an $n \times qp$ matrix $\mathbf{Y}$.



Two-level multivariate data

Connection with the classical multivariate model - rearrange the data to transform tensor $\mathcal{Y}$ into a matrix form

referred to as *matricization* (*unfolding* or *flattening*)

Three possibilities of matricization for a three-dimensional tensor $\mathcal{Y}_{n\times q\times p}$

- after matricization $\mathrm{E}(\mathbf{Y}) = \mathbf{M}$, with $\mathbf{M}$ being matricized form of tensor $\mathcal{M}_{n\times q\times p}$.
- we come to classical multivariate linear model with $\mathbf{Y} = \mathbf{M} + \mathbf{E}$, where

$$\mathbf{E} \sim N_{n,qp}(\mathbf{0}, \mathbf{I}_n, \boldsymbol{\Omega}).$$

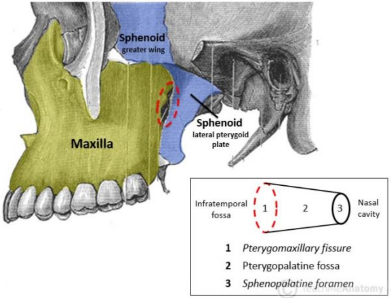
- structure of $\boldsymbol{\Omega}$ in $\mathrm{Var}[\mathrm{vec}\,\mathbf{Y}] = \boldsymbol{\Omega} \otimes \mathbf{I}_n$?

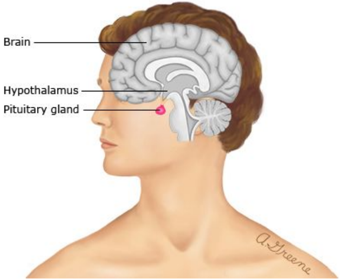
Growth curve model and its extensions

Estimation

Growth curve model

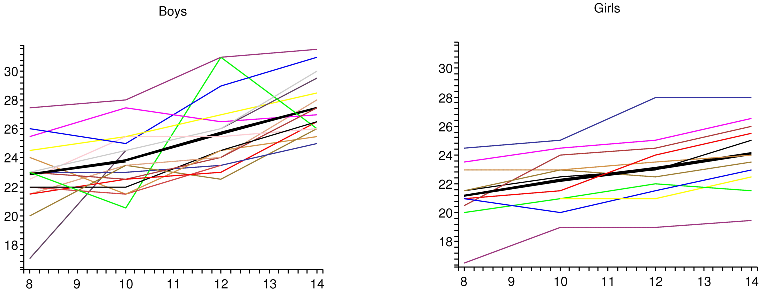
Dental data: the distance between the centre of the pituitary to the pterygomaxillary fissure.





Growth curve model

Dental data: the distance between the centre of the pituitary to the pterygomaxillary fissure. The observations were collected from 11 girls and 16 boys at 4 different ages, 8, 10, 12, and 14 years.



Growth curve model

Assume that the growth is linear in both groups - girls (k=1) and boys (k=2)

$$\beta_{jk} = b_{0k} + b_{1k} t_j, \quad j = 1, 2, 3, 4, \quad k = 1, 2.$$

So

$$\beta_k = \begin{pmatrix} b_{0k} + 8\, b_{1k} \\ b_{0k} + 10\, b_{1k} \\ b_{0k} + 12\, b_{1k} \\ b_{0k} + 14\, b_{1k} \end{pmatrix}, \quad k = 1, 2.$$

Growth curve model

Two groups of regression models with repeated measurements appeared:

$$\mathbf{y}_i^1 = \begin{pmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{pmatrix} \begin{pmatrix} b_{01} \\ b_{11} \end{pmatrix} + \mathbf{e}_i^1 = \mathbf{Z}\mathbf{b}_1 + \mathbf{e}_i^1, \quad i = 1, \cdots, 11,$$

$$\mathbf{y}_j^2 = \begin{pmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{pmatrix} \begin{pmatrix} b_{02} \\ b_{12} \end{pmatrix} + \mathbf{e}_j^2 = \mathbf{Z}\mathbf{b}_2 + \mathbf{e}_j^2, \quad j = 1, \cdots, 16.$$

Growth curve model

Potthoff and Roy realized that both models are connected through the same variance matrix assumption

$$\mathrm{Var}[\mathbf{e}_i^1] = \mathrm{Var}[\mathbf{e}_j^2] = \boldsymbol{\Omega}$$

They joint both models into one model (growth curve model)

$$\begin{pmatrix} \mathbf{y}_1^{1\prime} \\ \vdots \\ \mathbf{y}_{11}^{1}{}' \\ \mathbf{y}_1^{2\prime} \\ \vdots \\ \mathbf{y}_{16}^{2}{}' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \vdots & \\ 1 & 0 \\ 0 & 1 \\ \vdots & \\ 0 & 1 \end{pmatrix} \begin{pmatrix} b_{01} & b_{11} \\ b_{02} & b_{12} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 8 & 10 & 12 & 14 \end{pmatrix} + \begin{pmatrix} \mathbf{e}_1^{1\prime} \\ \vdots \\ \mathbf{e}_{11}^{1}{}' \\ \mathbf{e}_1^{2\prime} \\ \vdots \\ \mathbf{e}_{16}^{2}{}' \end{pmatrix}.$$

$$\mathbf{Y} = \mathbf{X}\mathbf{B}\mathbf{Z}' + \mathbf{E},$$

where $\mathrm{E}\,[\mathbf{E}] = \mathbf{0}$ and $\mathrm{Var}[\mathrm{vec}\,\mathbf{E}] = \boldsymbol{\Omega} \otimes \mathbf{I}_n$

Growth curve model

Definition

Let $\mathbf{Y}_{n\times p}$ and $\mathbf{B}_{k\times r}$ be respectively the observation and parameter matrices, and let $\mathbf{X}_{n\times k}$ and $\mathbf{Z}_{p\times r}$ $(r \leq p)$ be respectively the between- and within-individuals design matrices. Suppose that $n \geq p + r(\mathbf{X})$. The Growth curve model is defined as

$$\mathbf{Y} = \mathbf{XBZ}' + \mathbf{E},$$

where $\mathbf{E} = \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{pmatrix}'$, and $\mathbf{e}_i \sim N_p(\mathbf{0}, \mathbf{\Omega})$, with $\mathbf{e}_i$ and $\mathbf{e}_j$ being independent.

In contrary to ordinary MANOVA model the mean structure in GCM is bilinear.

Due to this structure the model (also called GMANOVA) belongs to the curved exponential family - problems relating to such issues as estimability and non-explicit maximum likelihood estimators (MLEs) may occur.

Growth curve model

Estimation of parameters, hypotheses testing and prediction of future values have been studied by many authors, thus generating a substantial literature - some general reviews

- von Rosen, D. (1991). The growth curve model: A rewiev. *Comm. Statist. Theory Methods 20*, 2791–2822.
- Srivastava, M.S., von Rosen, D. (1999). *Growth curve models*. In: Multivariate Analysis, Design of Experiments, and Survey Sampling, Ed. S. Ghosh. Marcel Dekker, New York, 547–578.
- Žežula, I., Klein, D. (2011). Overview of recent results in growth-curve-type multivariate linear models. *Acta Univ. Palacki. Olomuc., Fac. rer. nat., Mathematica 50(2)*, 137–146.

Many results can be found in textbooks

- Kollo T., von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices.* Dordrecht: Springer.
- Kshirsagar, A.M., Smith, W.B. (1995). *Growth curves*. Dekker, New York.
- Pan, J-X, Fang, K-T (2002). *Growth curve models and statistical diagnostics*. New York: Springer-Verlag.

Multivariate linear model - MLEs

Let $\mathbf{Y} = \mathbf{XB} + \mathbf{E}$, where $\mathbf{B}$ is $k \times p$ unknown parameter matrix, $\mathbf{X}$ is $n \times k$ known design matrix such that $n \geq rank(\mathbf{X}) + p$ and

$$\mathbf{E} \sim N_{n,p}(\mathbf{0}, \mathbf{I}_n, \mathbf{\Omega})$$

The log-likelihood function is given by

$$\ln L(\mathbf{B}, \mathbf{\Omega}|\mathbf{Y}) = -\frac{np}{2} \ln(2\pi) - \frac{n}{2} \ln|\mathbf{\Omega}| - \frac{1}{2} \operatorname{Tr}\left[\mathbf{\Omega}^{-1}(\mathbf{Y} - \mathbf{XB})'(\mathbf{Y} - \mathbf{XB})\right]$$

Multivariate linear model - MLEs

The MLEs for the multivariate linear model

$$\begin{aligned}\widehat{\mathbf{B}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \\ n\widehat{\boldsymbol{\Omega}} &= \mathbf{Y}'\mathbf{Q}_X\mathbf{Y} = \widehat{\mathbf{R}}'\widehat{\mathbf{R}} = \mathbf{S}\end{aligned}$$

where $\widehat{\mathbf{R}} = \mathbf{Q}_X\mathbf{Y}$ is a residual matrix and

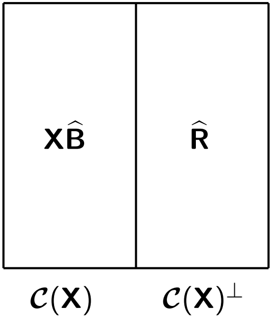
$$\begin{aligned}&\mathbf{Q}_X = \mathbf{I}_n - \mathbf{P}_X && \text{is a projection on the space } \mathcal{C}(\mathbf{X})^{\perp} \\ &\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' && \text{is a projection on the space } \mathcal{C}(\mathbf{X})\end{aligned}$$

The estimated mean structure and residuals are

$$\begin{aligned}\mathbf{X}\widehat{\mathbf{B}} &= \mathbf{P}_X\mathbf{Y}, \\ \widehat{\mathbf{R}} &= \mathbf{Q}_X\mathbf{Y}.\end{aligned}$$

Multivariate linear model - MLEs

$$\mathcal{C}(\mathbf{X}) \boxplus \mathcal{C}(\mathbf{X})^{\perp}$$



Growth curve model - MLEs

Let $\mathbf{Y} = \mathbf{XBZ}' + \mathbf{E}$, where $\mathbf{B}$ is $k \times r$ unknown parameter matrix, $\mathbf{X}_{n \times k}$ and $\mathbf{Z}_{p \times r}$ are known between- and within-individuals design matrices, respectively, such that $n \geq rank(\mathbf{X}) + p$ and

$$\mathbf{E} \sim N_{n,p}(\mathbf{0}, \mathbf{I}_n, \mathbf{\Omega})$$

The log-likelihood function is given by

$$\ln L(\mathbf{B}, \mathbf{\Omega}|\mathbf{Y}) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\mathbf{\Omega}| - \frac{1}{2}\operatorname{Tr}\left[\mathbf{\Omega}^{-1}(\mathbf{Y} - \mathbf{XBZ}')'(\mathbf{Y} - \mathbf{XBZ}')\right]$$

Growth curve model - MLEs

The MLEs for the growth curve model (**X** and **Z** of full rank)

$$\widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{S}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Z})^{-1},$$

$$n\widehat{\boldsymbol{\Omega}} = (\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}}\mathbf{Z}')'(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}}\mathbf{Z}') = \mathbf{S} + \widehat{\mathbf{R}}_1'\widehat{\mathbf{R}_1},$$

where $\mathbf{S} = \widehat{\mathbf{R}}'\widehat{\mathbf{R}} = \mathbf{Y}'\mathbf{Q}_X\mathbf{Y}$.

The estimated mean structure and residuals are

$$\mathbf{X}\widehat{\mathbf{B}}\mathbf{Z}' = \mathbf{P}_X\mathbf{Y}(\mathbf{P}_{Z;S^{-1}})',$$

$$\widehat{\mathbf{R}} = \mathbf{Q}_X\mathbf{Y},$$

$$\widehat{\mathbf{R}}_1 = \mathbf{P}_X\mathbf{Y}(\mathbf{Q}_{Z;S^{-1}})'$$

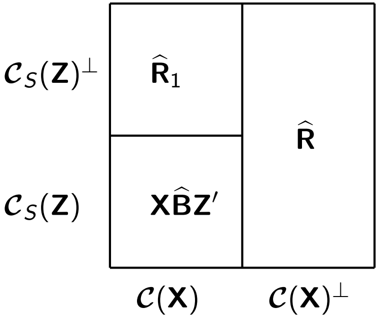
where

$\mathbf{Q}_{Z;S^{-1}} = \mathbf{I}_n - \mathbf{P}_{Z;S^{-1}}$ is a projection on the space $\mathcal{C}_S(\mathbf{Z})^\perp$

$\mathbf{P}_{Z;S^{-1}} = \mathbf{Z}(\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Z})^{-}\mathbf{Z}'\mathbf{S}^{-1}$ is a projection on the space $\mathcal{C}_S(\mathbf{Z})$

Growth curve model - MLEs

$$\mathcal{C}_S(\mathbf{Z}) \otimes \mathcal{C}(\mathbf{X}) \boxplus \left(\mathcal{C}_S(\mathbf{Z}) \otimes \mathcal{C}(\mathbf{X})\right)^{\perp} =$$
$$= \mathcal{C}_S(\mathbf{Z}) \otimes \mathcal{C}(\mathbf{X}) \boxplus \mathcal{C}_S(\mathbf{Z})^{\perp} \otimes \mathcal{C}(\mathbf{X}) \boxplus \boldsymbol{\mathcal{V}} \otimes \mathcal{C}(\mathbf{X})^{\perp}$$



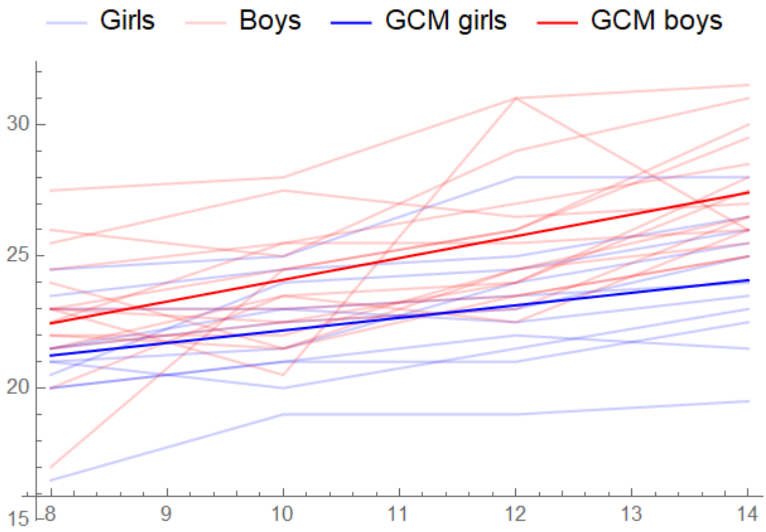
Growth curve model - MLEs in dental example

The MLEs for dental example

$$\widehat{\mathbf{B}} = \begin{pmatrix} 17.43 & 0.48 \\ 15.84 & 0.83 \end{pmatrix} \quad \begin{matrix} \text{Girls} \\ \text{Boys} \end{matrix}$$

$$\widehat{\boldsymbol{\Omega}} = \begin{pmatrix} 5.12 & 2.44 & 3.61 & 2.52 \\ 2.44 & 3.93 & 2.72 & 3.06 \\ 3.61 & 2.72 & 5.98 & 3.82 \\ 2.52 & 3.06 & 3.82 & 4.62 \end{pmatrix}$$

Growth curve model - MLEs in dental example



Extended GCM

Basic GCM: contains a single profile for all groups

Example

- three treatment groups of animals
- each group being subject to a different treatment
- response variable - weight of animals in all groups measured at the same p time points
- measurements on a single animal are assumed to have the same variance matrix $\boldsymbol{\Omega}$

Expected growth curves: polynomial in time, groups differ by order (linear, quadratic, cubic)

$$b_{1j} + b_{2j} t_s + b_{3j} t_s^2 + b_{4j} t_s^3$$

$$\text{for } j = 1, 2, 3, \quad \text{and} \quad s = 1, \ldots, p$$

Extended GCM

Not possible to model the mean structure with a single-profile GCM $\rightarrow$ *extended GCM*

$$\mathbf{Y} = \mathbf{X}_1\mathbf{B}_1\mathbf{Z}_1' + \mathbf{X}_2\mathbf{B}_2\mathbf{Z}_2' + \mathbf{X}_3\mathbf{B}_3\mathbf{Z}_3' + \boldsymbol{E},$$

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \mathbf{0}_{n_2} \\ \mathbf{0}_{n_3} & \mathbf{0}_{n_3} & \mathbf{1}_{n_3} \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} \\ \mathbf{0}_{n_3} & \mathbf{1}_{n_3} \end{pmatrix}, \quad \mathbf{X}_3 = \begin{pmatrix} \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} \\ \mathbf{1}_{n_3} \end{pmatrix},$$

$$\mathbf{B}_1 = \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \\ b_{13} & b_{23} \end{pmatrix}, \quad \mathbf{B}_2 = \begin{pmatrix} b_{32} \\ b_{33} \end{pmatrix}, \quad \mathbf{B}_3 = \begin{pmatrix} b_{43} \end{pmatrix},$$

$$\mathbf{Z}_1 = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_p \end{pmatrix}, \quad \mathbf{Z}_2 = \begin{pmatrix} t_1^2 \\ t_2^2 \\ \vdots \\ t_p^2 \end{pmatrix}, \quad \mathbf{Z}_3 = \begin{pmatrix} t_1^3 \\ t_2^3 \\ \vdots \\ t_p^3 \end{pmatrix}.$$

Extended GCM

Definition

Let $\mathbf{Y}$ and $\mathbf{B}_i$ be respectively $n \times p$ observation and $k_i \times r_i$ parameter matrices, and let $\mathbf{X}_i$ and $\mathbf{Z}_i$ be respectively $n \times k_i$ between- and $p \times r_i$ within-individuals design matrices. Suppose that $n \geq p + r(\mathbf{X}_1)$. The Extended growth curve model (EGCM) with fixed effects, called also sum-of-profiles model is defined as

$$\mathbf{Y} = \sum_{i=1}^{m} \mathbf{X}_i \mathbf{B}_i \mathbf{Z}_i' + \mathbf{E}\,,$$

where $\mathbf{E} \sim N_{n,p}\left(\mathbf{0}, \mathbf{I}_n, \mathbf{\Omega}\right)$.

Introduced by Von Rosen (1984) and independently by Verbyla and Venables (1988) - they presented several examples to illustrate how this model can arise

Extended GCM - MLEs

Usual assumption - nested-subspace condition of between-individual design matrices

$$\mathcal{C}(\mathbf{X}_m) \subseteq \cdots \subseteq \mathcal{C}(\mathbf{X}_1)$$

- necessary condition for the existence of MLE
- Von Rosen (1989) derived the MLEs of the unknown parameters
- nothing is usually said about different $\mathbf{Z}_i$'s
- this model separates different profiles

Extended GCM - MLEs

For simplicity all design matrices assumed to be of full rank. The MLEs for the multivariate linear model (for $m = 2$)

$$\begin{aligned}
\widehat{\mathbf{B}}_2 &= (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{Y}\mathbf{S}_2^{-1}\mathbf{P}_2\mathbf{Z}_2(\mathbf{Z}_2'\mathbf{P}_2'\mathbf{S}_2^{-1}\mathbf{P}_2\mathbf{Z}_2)^{-1},\\
\widehat{\mathbf{B}}_1 &= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{Y} - \mathbf{X}_2\widehat{\mathbf{B}}_2\mathbf{Z}_2')\mathbf{S}_1^{-1}\mathbf{Z}_1(\mathbf{Z}_1'\mathbf{S}_2^{-1}\mathbf{Z}_1)^{-1},\\
n\widehat{\boldsymbol{\Omega}} &= (\mathbf{Y} - \sum_{j=1}^{2}\mathbf{X}_j\widehat{\mathbf{B}}_j\mathbf{Z}_j')'(\mathbf{Y} - \sum_{j=1}^{2}\mathbf{X}_j\widehat{\mathbf{B}}_j\mathbf{Z}_j') = \mathbf{S}_1 + \widehat{\mathbf{R}}_2'\widehat{\mathbf{R}_2} + \widehat{\mathbf{R}}_3'\widehat{\mathbf{R}_3},
\end{aligned}$$

where $\mathbf{P}_2 = \mathbf{Q}_{Z_1;S_1}$ and

$$\begin{aligned}
\mathbf{S}_1 &= \widehat{\mathbf{R}}_1'\widehat{\mathbf{R}}_1 = \mathbf{Y}'\mathbf{Q}_{X_1}\mathbf{Y},\\
\mathbf{S}_2 &= \mathbf{S}_1 + \mathbf{P}_2\mathbf{Y}'\mathbf{P}_{X_1}\mathbf{Q}_{X_2}\mathbf{P}_{X_1}\mathbf{Y}\mathbf{P}_2'
\end{aligned}$$

Extended GCM - MLEs

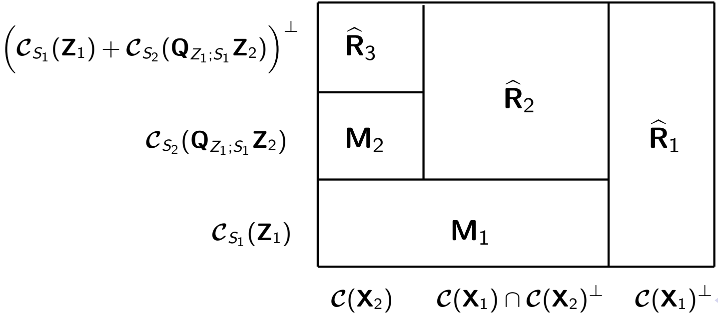
The estimated mean structure and residuals are

$$\begin{aligned}
\sum_{j=1}^{m} \mathbf{X}_j \widehat{\mathbf{B}}_j \mathbf{Z}_j' &= \mathbf{P}_{X_1}\mathbf{Y}(\mathbf{P}_{Z_1;S_1})' + \mathbf{P}_{X_2}\mathbf{Y}(\mathbf{P}_{Q_{Z_1;S_1}Z_2;S_2})' = \mathbf{M}_1 + \mathbf{M}_2, \\
\widehat{\mathbf{R}}_1 &= \mathbf{Q}_{X_1}\mathbf{Y}, \\
\widehat{\mathbf{R}}_2 &= (\mathbf{P}_{X_1} - \mathbf{P}_{X_2})\mathbf{Y}(\mathbf{Q}_{Z_1;S_1})', \\
\widehat{\mathbf{R}}_1 &= \mathbf{P}_{X_2}\mathbf{Y}(\mathbf{Q}_{Z_1;S_1} - \mathbf{P}_{Q_{Z_1;S_1}Z_2;S_2})',
\end{aligned}$$

Extended GCM - MLEs

Decomposition of the whole space according to the between- and within-individuals design matrices

$$\mathcal{C}_{S_1}(\mathbf{Z}_1)\otimes\mathcal{C}(\mathbf{X}_1)\boxplus\mathcal{C}_{S_2}(\mathbf{Q}_{Z_1;S_1}\mathbf{Z}_2)\otimes\mathcal{C}(\mathbf{X}_2)\boxplus$$
$$\boxplus\left(\mathcal{C}_{S_1}(\mathbf{Z}_1)+\mathcal{C}_{S_2}(\mathbf{Q}_{Z_1;S_1}\mathbf{Z}_2)\right)^{\perp}\otimes\mathcal{C}(\mathbf{X}_2)\boxplus\mathcal{C}_{S_1}(\mathbf{Z}_1)^{\perp}\otimes\left(\mathcal{C}(\mathbf{X}_1)\cap\mathcal{C}(\mathbf{X}_2)^{\perp}\right)\boxplus\mathcal{V}\otimes\mathcal{C}(\mathbf{X}_1)^{\perp}$$



Extended GCM

Filipiak and von Rosen (2011) - discussed the model with the nested-subspace condition of within-individual design matrices

$$\mathcal{C}(\mathbf{Z}_m) \subseteq \cdots \subseteq \mathcal{C}(\mathbf{Z}_1)\,.$$

- they showed that the two models are equivalent via reparametrization
- because of non-linearity the properties of estimators cannot be transmitted directly
- they gave MLEs of unknown parameters for the three component model

Extended GCM

Hu (2009) came with the idea of orthogonal decomposition in EGCM - the idea is to separate groups rather than models

$$\mathbf{X}_i'\mathbf{X}_j = \mathbf{0} \quad \forall i \neq j$$

while no assumption about $\mathcal{C}(\mathbf{Z}_i)$'s

Extended GCM

- Weight of animals in three groups measured at the same p time points - separating different profiles

$$\mathbf{Y} = \mathbf{X}_1\mathbf{B}_1\mathbf{Z}_1' + \mathbf{X}_2\mathbf{B}_2\mathbf{Z}_2' + \mathbf{X}_3\mathbf{B}_3\mathbf{Z}_3' + \boldsymbol{E},$$

$$\begin{aligned}
\mathbf{X}_1 &= \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \mathbf{0}_{n_2} \\ \mathbf{0}_{n_3} & \mathbf{0}_{n_3} & \mathbf{1}_{n_3} \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} \\ \mathbf{0}_{n_3} & \mathbf{1}_{n_3} \end{pmatrix}, \quad \mathbf{X}_3 = \begin{pmatrix} \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} \\ \mathbf{1}_{n_3} \end{pmatrix}, \\
\mathbf{B}_1 &= \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \\ b_{13} & b_{23} \end{pmatrix}, \quad \mathbf{B}_2 = \begin{pmatrix} b_{32} \\ b_{33} \end{pmatrix}, \quad \mathbf{B}_3 = \begin{pmatrix} b_{43} \end{pmatrix}, \\
\mathbf{Z}_1 &= \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_p \end{pmatrix}, \quad \mathbf{Z}_2 = \begin{pmatrix} t_1^2 \\ t_2^2 \\ \vdots \\ t_p^2 \end{pmatrix}, \quad \mathbf{Z}_3 = \begin{pmatrix} t_1^3 \\ t_2^3 \\ \vdots \\ t_p^3 \end{pmatrix}.
\end{aligned}$$

Extended GCM

Weight of animals in three groups measured at the same p time points - separating different groups (treatments)

$$\mathbf{Y} = \mathbf{X}_1\mathbf{B}_1\mathbf{Z}_1' + \mathbf{X}_2\mathbf{B}_2\mathbf{Z}_2' + \mathbf{X}_3\mathbf{B}_3\mathbf{Z}_3' + \boldsymbol{E},$$

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \mathbf{0}_{n_2} \\ \mathbf{0}_{n_3} & \mathbf{0}_{n_3} & \mathbf{1}_{n_3} \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} \\ \mathbf{0}_{n_3} & \mathbf{1}_{n_3} \end{pmatrix}, \quad \mathbf{X}_3 = \begin{pmatrix} \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} \\ \mathbf{1}_{n_3} \end{pmatrix},$$

$$\mathbf{B}_1 = \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \\ b_{13} & b_{23} \end{pmatrix}, \quad \mathbf{B}_2 = \begin{pmatrix} b_{32} \\ b_{33} \end{pmatrix}, \quad \mathbf{B}_3 = \begin{pmatrix} b_{43} \end{pmatrix},$$

$$\mathbf{Z}_1 = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_p \end{pmatrix}, \quad \mathbf{Z}_2 = \begin{pmatrix} t_1^2 \\ t_2^2 \\ \vdots \\ t_p^2 \end{pmatrix}, \quad \mathbf{Z}_3 = \begin{pmatrix} t_1^3 \\ t_2^3 \\ \vdots \\ t_p^3 \end{pmatrix}.$$

group 1

Extended GCM

Weight of animals in three groups measured at the same p time points - separating different groups (treatments)

$$\mathbf{Y} = \mathbf{X}_1\mathbf{B}_1\mathbf{Z}_1' + \mathbf{X}_2\mathbf{B}_2\mathbf{Z}_2' + \mathbf{X}_3\mathbf{B}_3\mathbf{Z}_3' + \boldsymbol{E},$$

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \mathbf{0}_{n_2} \\ \mathbf{0}_{n_3} & \mathbf{0}_{n_3} & \mathbf{1}_{n_3} \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} \\ \mathbf{0}_{n_3} & \mathbf{1}_{n_3} \end{pmatrix}, \quad \mathbf{X}_3 = \begin{pmatrix} \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} \\ \mathbf{1}_{n_3} \end{pmatrix},$$

$$\mathbf{B}_1 = \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \\ b_{13} & b_{23} \end{pmatrix}, \quad \mathbf{B}_2 = \begin{pmatrix} b_{32} \\ b_{33} \end{pmatrix}, \quad \mathbf{B}_3 = \begin{pmatrix} b_{43} \end{pmatrix},$$

$$\mathbf{Z}_1 = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_p \end{pmatrix}, \quad \mathbf{Z}_2 = \begin{pmatrix} t_1^2 \\ t_2^2 \\ \vdots \\ t_p^2 \end{pmatrix}, \quad \mathbf{Z}_3 = \begin{pmatrix} t_1^3 \\ t_2^3 \\ \vdots \\ t_p^3 \end{pmatrix}.$$

group 2

Extended GCM

Weight of animals in three groups measured at the same p time points - separating different groups (treatments)

$$\mathbf{Y} = \mathbf{X}_1\mathbf{B}_1\mathbf{Z}_1' + \mathbf{X}_2\mathbf{B}_2\mathbf{Z}_2' + \mathbf{X}_3\mathbf{B}_3\mathbf{Z}_3' + \boldsymbol{E},$$

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \mathbf{0}_{n_2} \\ \mathbf{0}_{n_3} & \mathbf{0}_{n_3} & \mathbf{1}_{n_3} \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} \\ \mathbf{0}_{n_3} & \mathbf{1}_{n_3} \end{pmatrix}, \quad \mathbf{X}_3 = \begin{pmatrix} \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} \\ \mathbf{1}_{n_3} \end{pmatrix},$$

$$\mathbf{B}_1 = \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \\ b_{13} & b_{23} \end{pmatrix}, \quad \mathbf{B}_2 = \begin{pmatrix} b_{32} \\ b_{33} \end{pmatrix}, \quad \mathbf{B}_3 = \begin{pmatrix} b_{43} \end{pmatrix},$$

$$\mathbf{Z}_1 = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_p \end{pmatrix}, \quad \mathbf{Z}_2 = \begin{pmatrix} t_1^2 \\ t_2^2 \\ \vdots \\ t_p^2 \end{pmatrix}, \quad \mathbf{Z}_3 = \begin{pmatrix} t_1^3 \\ t_2^3 \\ \vdots \\ t_p^3 \end{pmatrix}.$$

group 3

Extended GCM

Weight of animals in three groups measured at the same p time points - separating different groups (treatments)

$$\mathbf{Y} = \mathbf{X}_1\mathbf{B}_1\mathbf{Z}_1' + \mathbf{X}_2\mathbf{B}_2\mathbf{Z}_2' + \mathbf{X}_3\mathbf{B}_3\mathbf{Z}_3' + \boldsymbol{E},$$

$$\begin{aligned}
\mathbf{X}_1 &= \begin{pmatrix} \mathbf{1}_{n_1} \\ \mathbf{0}_{n_2} \\ \mathbf{0}_{n_3} \end{pmatrix}, \quad \mathbf{X}_2 = \begin{pmatrix} \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} \\ \mathbf{0}_{n_3} \end{pmatrix}, \quad \mathbf{X}_3 = \begin{pmatrix} \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} \\ \mathbf{1}_{n_3} \end{pmatrix}, \\
\mathbf{B}_1 &= \begin{pmatrix} b_{11} & b_{21} \end{pmatrix}, \quad \mathbf{B}_2 = \begin{pmatrix} b_{12} & b_{22} & b_{32} \end{pmatrix}, \quad \mathbf{B}_3 = \begin{pmatrix} b_{13} & b_{23} & b_{33} & b_{43} \end{pmatrix}, \\
\mathbf{Z}_1 &= \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_p \end{pmatrix}, \quad \mathbf{Z}_2 = \begin{pmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & & \\ 1 & t_p & t_p^2 \end{pmatrix}, \quad \mathbf{Z}_3 = \begin{pmatrix} 1 & t_1 & t_1^2 & t_1^3 \\ 1 & t_2 & t_2^2 & t_2^3 \\ \vdots & & & \\ 1 & t_p & t_p^2 & t_p^3 \end{pmatrix}.
\end{aligned}$$

Orthogonal extended GCM

$$\mathbf{Y} = \sum_{i=1}^{m} \mathbf{X}_i \mathbf{B}_i \mathbf{Z}_i' + \mathbf{E}, \quad \mathbf{E} \sim N_{n,p}(\mathbf{0}, \mathbf{I}_n, \mathbf{\Omega}),$$

- $\sum_{i=1}^{m} r(\mathbf{X}_i) + p \leq n$
- design matrices $\mathbf{X}_i$ satisfy the condition $\mathbf{X}_i'\mathbf{X}_j = 0 \ \ \forall i \neq j$
- nested column spaces of all $\mathbf{Z}_i$'s

$$\mathcal{C}(\mathbf{Z}_1) \subseteq \cdots \subseteq \mathcal{C}(\mathbf{Z}_m)$$

This naturally arises in situations when different groups use polynomial regression functions of different order.

Orthogonal extended GCM

Many tasks, which are difficult or impossible to handle in basic models, can be done with ease in models consisting of mutually orthogonal components.

- it allows to determine explicit forms of estimators, simplifies proving their properties, and sometimes also enables to find their distributions.

Orthogonal extended GCM

Many tasks, which are difficult or impossible to handle in basic models, can be done with ease in models consisting of mutually orthogonal components.

- it allows to determine explicit forms of estimators, simplifies proving their properties, and sometimes also enables to find their distributions.

Klein and Žežula (2015) show that via a simple transformation von Rosen model can be transformed into an equivalent one with orthogonal column spaces of $\mathbf{X}_i$

- established the maximum likelihood estimators of unknown parameters
- derived the moments of the estimators

Orthogonal extended GCM -MLEs

All $\mathbf{X}_i$'s and $\mathbf{Z}_i$'s are of full rank

- estimators in von Rosen model are rather complicated, calculated recursively
- in orthogonal model the estimators have nice closed form

$$\hat{\mathbf{B}}_i = \left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{Y}\mathbf{S}_i^{-1}\mathbf{Z}_i\left(\mathbf{Z}_i'\mathbf{S}_i^{-1}\mathbf{Z}_i\right)^{-1},$$

$$n\widehat{\mathbf{\Omega}} = \mathbf{S}_1 + \sum_{i=1}^{m}\mathbf{Q}_{Z_i;S_i}\mathbf{Y}'\mathbf{P}_{X_i}\mathbf{Y}\left(\mathbf{Q}_{Z_i;S_i}\right)',$$

where

$$\mathbf{S}_1 = \mathbf{Y}'\mathbf{Q}_{(X_1\cdots X_m)}\mathbf{Y},$$

$$\mathbf{S}_i = \mathbf{S}_{i-1} + \mathbf{Q}_{Z_{i-1};S_{i-1}}\mathbf{Y}'\mathbf{P}_{X_{i-1}}\mathbf{Y}\left(\mathbf{Q}_{Z_{i-1};S_{i-1}}\right)', \quad \text{for } i = 2,\ldots,m,$$

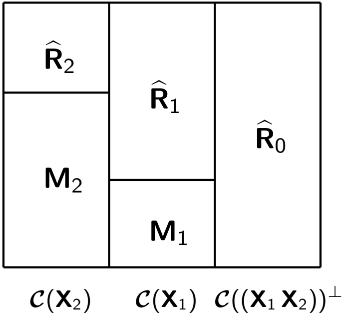
Orthogonal extended GCM -MLEs

The estimated mean structure and residuals are

$$\begin{aligned}
\sum_{j=1}^{m} \mathbf{X}_j \widehat{\mathbf{B}}_j \mathbf{Z}_j' &= \sum_{j=1}^{m} \mathbf{P}_{X_i} \mathbf{Y} (\mathbf{P}_{Z_i;S_i})' = \sum_{j=1}^{m} \mathbf{M}_i, \\
\widehat{\mathbf{R}}_0 &= \mathbf{Q}_{(X_1 \cdots X_m)} \mathbf{Y}, \\
\widehat{\mathbf{R}}_i &= \mathbf{P}_{X_i} \mathbf{Y} (\mathbf{Q}_{Z_i;S_i})',
\end{aligned}$$

Orthogonal extended GCM -MLEs

$$\mathcal{C}_{S_1}(\mathbf{Z}_1) \otimes \mathcal{C}(\mathbf{X}_1) \boxplus \mathcal{C}_{S_2}(\mathbf{Z}_2) \otimes \mathcal{C}(\mathbf{X}_2) \boxplus$$
$$\boxplus\, \mathcal{C}_{S_2}(\mathbf{Z}_2)^{\perp} \otimes \mathcal{C}(\mathbf{X}_2) \boxplus \mathcal{C}_{S_1}(\mathbf{Z}_1)^{\perp} \otimes \mathcal{C}(\mathbf{X}_1) \boxplus \mathcal{V} \otimes \mathcal{C}((\mathbf{X}_1\, \mathbf{X}_2))^{\perp}$$



Orthogonal extended GCM - moments of estimators

Moments of estimators in von Rosen model are derived only for three component model (calculations are very tedious in general k component model)

In orthogonal model the moments for all unknown parameters can be given

Orthogonal extended GCM - moments of estimators

$$\begin{aligned}
\mathrm{E}\,[\widehat{\mathbf{B}}_i] =& \mathbf{B}_i, \\
\mathrm{Var}\left[\widehat{\mathbf{B}}_i\right] =& \Bigg((1+c_i)\left(\mathbf{Z}_i'\boldsymbol{\Omega}^{-1}\mathbf{Z}_i\right)^{-1} + \sum_{j=1}^{i-1}(d_j-1)d_{j+1}\dots d_{i-1}c_i \times \\
& \times \left(\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\mathbf{Z}_i'\mathbf{Z}_j\left(\mathbf{Z}_j'\boldsymbol{\Omega}^{-1}\mathbf{Z}_j\right)^{-1}\mathbf{Z}_j'\mathbf{Z}_i\left(\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\Bigg) \otimes \left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}, \\
\mathrm{E}\left[\widehat{\boldsymbol{\Omega}}\right] =& \left(1+\sum_{i=1}^{k}\frac{r\left(\mathbf{X}_i\right)}{n}\left[(c_i-1)\mathbf{P}_{Z_i;\Omega} + \sum_{j=1}^{i-1}(d_j-1)d_{j+1}\dots d_{i-1}c_i\mathbf{P}_{Z_j;\Omega}\right]\right)\boldsymbol{\Omega},
\end{aligned}$$

where $c_i = \frac{p-r(\mathbf{Z}_i)}{n-\sum_{l=i}^{m} r(\mathbf{X}_l)-p+r(\mathbf{Z}_i)-1}$ and $d_j = \frac{n-\sum_{l=j+1}^{m} r(\mathbf{X}_l)-p+r(\mathbf{Z}_j)-1}{n-\sum_{l=j}^{m} r(\mathbf{X}_l)-p+r(\mathbf{Z}_j)-1}$ for $1 \le j < i$.

Thank you for your attention!

Estimation and testing in multivariate linear models
Part II

Daniel Klein

Institute of Mathematics
Faculty of Science
P. J. Šafárik University in Košice

Będlewo
November 29, 2021

Special covariance structures

Introduction

Large datasets routinely collected nowadays due to rapid advances in computer-based or web-based commerce and data-collection technology - research has become very active in response to an increasingly important need for analysis of massive and large-dimensional data

Dempster (1958, 1960) - a non-exact test for the two-sample significance test for larger dimension of data than the degrees of freedom - he raised the question: what statisticians should do if traditional multivariate statistical theory does not apply when the dimension of data is too large.

Bai and Saranadasa (1996) - even when traditional approaches can be applied, they are much less powerful than the non-exact test when the dimension of data is large - another question: how classical multivariate statistical procedures could be adapted and improved when the data dimension is large.

Introduction

Two directions to solve these problems:

- to propose special statistical procedures to solve ad hoc large-dimensional statistical problems where traditional multivariate statistical procedures are inapplicable or perform poorly. The family of various non-exact tests follows this approach.
- to make systematic corrections to the classical multivariate statistical procedures so that the effect of large dimension is overcome - this is achieved by employing new and powerful asymptotic tools borrowed from the theory of random matrices.

Special (patterned) covariance structures

- While estimating 2nd order parameters:

  $\mathbf{\Omega}$ can be treated as completely unknown, i.e. the matrix $\mathbf{\Omega}$ ranges over

$$\mathcal{V} = \{\mathbf{\Omega} : \ \mathbf{\Omega} \ \text{ is symmetric positive definite}\}\,.$$

- Number of unknown parameters in $\mathbf{\Omega}$: $\frac{p(p+1)}{2}$
- Problem when p is increasing:
  - Number of 2nd order parameters (elements of $\mathbf{\Omega}$) grows quickly.
  - Estimability and stability of the estimators requires a lot of observations.
  - Crowder and Hand (1990) - in case of small samples unstructured covariance matrix can result in rather weak inference, in the sense that too many degrees of freedom are used up in estimating the covariance parameters, leaving too few for the parameters of interest.

Special (patterned) covariance structures

- To avoid over-parametrization and to allow parsimonious modelling: considering a simpler covariance structure - keeps the number of unknown parameters reasonable.
- Various structures are studied in the literature
  - linearly structured (compound symmetry, generalized compound symmetry, Toeplitz, circulant Toeplitz)
  - banded
  - short auto-regression time series - serial correlation structure
- Several authors have assumed the structure of the form

$$\boldsymbol{\Sigma} = \sigma^2 R(\varrho)$$

where σ^2 is the scale parameter and the patterned correlation matrix $R(\varrho)$ is a function of the correlation scalar/vector parameter.

Toeplitz structure

$$\boldsymbol{\Omega} = \sigma^2 \left[\mathbf{I}_p + \sum_{i=1}^{p-1} \varrho_i (\mathbf{C}_p^i + \mathbf{C}_p'^i) \right],$$

where $\mathbf{C}_p$: $p \times p$ circulant matrix with ones on the first supra-diagonal and zeros elsewhere.

Entries on the i-th diagonal are equal to $\sigma^2 \varrho_{i-1}$, $i = 1, \ldots, p$ with $\varrho_0 = 1$.

Often used in the context of time series or longitudinal models.

MLEs have no closed form and their finite sample properties are not known.

Special Toeplitz type structures - the parameters ϱ_i are subject to some constraint, e.g., $\varrho_i = \varrho$ for all i leads to *compound symmetry* structure

Compound symmetry structure

- Wilks (1946): considered the compound symmetry (also known as intraclass or uniform) covariance structure when dealing with measurements on k equivalent psychological tests

$$R(\varrho) = (1-\varrho)\mathbf{I}_p + \varrho \mathbf{1}_p \mathbf{1}_p',$$

where $-(p-1)^{-1} < \varrho < 1$.

  - developed statistical test criteria for testing equality in means, equality in variances and equality in covariances.

Compound symmetry structure

- Wilks (1946): considered the compound symmetry (also known as intraclass or uniform) covariance structure when dealing with measurements on k equivalent psychological tests

$$R(\varrho) = (1-\varrho)\mathbf{I}_p + \varrho \mathbf{1}_p \mathbf{1}_p',$$

where $-(p-1)^{-1} < \varrho < 1$.

  - developed statistical test criteria for testing equality in means, equality in variances and equality in covariances.

- CS structure $\boldsymbol{\Omega}_{CS} = \sigma^2 R(\varrho)$- called sometimes equicorrelation or exchangeable structure

$$\text{for } \mathbf{P} \text{ being permutation matrix and } \mathrm{Var}[\mathbf{x}] = \boldsymbol{\Omega}_{CS}$$

$$\mathrm{Var}[\mathbf{Px}] = \mathrm{Var}[\mathbf{x}]$$

Compound symmetry structure

The eigenvectors of $\boldsymbol{\Omega}_{CS}$ do not depend on unknown parameters with eigenvalues $\sigma^2(1-(p-1)\varrho)$ and $\sigma^2(1-\varrho)$

$$\mathbf{H}'\begin{pmatrix} \sigma^2\varrho & \varrho & \cdots & \varrho \\ \varrho & \sigma^2\varrho & \cdots & \varrho \\ \vdots & \vdots & \ddots & \vdots \\ \varrho & \varrho & \cdots & \sigma^2\varrho \end{pmatrix}\mathbf{H} = \begin{pmatrix} \sigma^2(1-(p-1)\varrho) & 0 & \cdots & 0 \\ 0 & \sigma^2(1-\varrho) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2(1-\varrho) \end{pmatrix},$$

where $\mathbf{H}$ is any orthogonal matrix with first column $\frac{1}{\sqrt{p}}\mathbf{1}_p$ (e.g. Helmert matrix)

Compound symmetry structure

The eigenvectors of $\boldsymbol{\Omega}_{\mathrm{CS}}$ do not depend on unknown parameters with eigenvalues $\sigma^2(1-(p-1)\varrho)$ and $\sigma^2(1-\varrho)$

$$\mathbf{H}'\begin{pmatrix} \sigma^2\varrho & \varrho & \cdots & \varrho \\ \varrho & \sigma^2\varrho & \cdots & \varrho \\ \vdots & \vdots & \ddots & \vdots \\ \varrho & \varrho & \cdots & \sigma^2\varrho \end{pmatrix}\mathbf{H} = \begin{pmatrix} \sigma^2(1-(p-1)\varrho) & 0 & \cdots & 0 \\ 0 & \sigma^2(1-\varrho) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2(1-\varrho) \end{pmatrix},$$

where $\mathbf{H}$ is any orthogonal matrix with first column $\frac{1}{\sqrt{p}}\mathbf{1}_p$ (e.g. Helmert matrix)

Possible reparametrization

$$\boldsymbol{\Omega}_{\mathrm{CS}} = \sigma^2(1+(p-1)\varrho)\mathbf{P}_p + \sigma^2(1-\varrho)\mathbf{Q}_p,$$

where $\mathbf{P}_p = \mathbf{P}_{1_p}$ and $\mathbf{Q}_p = \mathbf{Q}_{1_p}$ are projectors onto $\mathcal{C}(\mathbf{1}_p)$ and $\mathcal{C}(\mathbf{1}_p)^\perp$, respectively

Generalized compound symmetry structure

Khatri (1973) considered generalized CS structure for the hypothesis testing of covariance structure

Generalized CS structure

$$\mathbf{\Omega}_{GCS} = \theta_1 \mathbf{G} + \theta_2 \mathbf{w}\mathbf{w}',$$

where $\mathbf{G}_{p\times p}$ is known positive definite matrix, $\mathbf{w}$ is a given p-dimensional vector and $\theta_i,\ i = 1, 2,$ are unknown scalar parameters

Reduces to CS for $\mathbf{G} = \mathbf{I}_p$ and $\mathbf{w} = \mathbf{1}_p$

Group symmetry models

CS belongs to the class of group symmetry models

- introduces by Andersson (1975), discussed and summarized in Perlman (1987)

The covariance matrix is assumed to satisfy symmetry restrictions

$$\text{Var}[\mathbf{y}] = Var[\mathbf{L}\mathbf{y}] = \mathbf{L}\,\text{Var}[\mathbf{y}]\mathbf{L}'$$

for all matrices $\mathbf{L}$ belonging to a finite group of orthogonal matrices.

A family $\mathbf{\Omega}$ of $p \times p$ covariance matrices is a group symmetry model if and only if there exist positive integers $t, p_1, \dots, p_t, r_1, \dots, r_t$, such that $\sum_{i=1}^{t} p_i r_i = p$, and a fixed orthogonal matrix $\mathbf{\Gamma}_{p\times p}$ such that

$$\mathbf{\Gamma}\mathbf{\Omega}\mathbf{\Gamma}' = \text{diag}(\underbrace{\mathbf{\Omega}_1, \dots, \mathbf{\Omega}_1}_{r_1}, \dots, \underbrace{\mathbf{\Omega}_t, \dots, \mathbf{\Omega}_t}_{r_t}),$$

where $\mathbf{\Omega}_i$ is a $p_i \times p_i$ covariance matrix.

Circular Toeplitz structure

Special Toeplitz type structure assuming $\varrho_i = \varrho_{p-i}$, i.e.

$$\boldsymbol{\Omega} = \sigma^2 \begin{pmatrix} 1 & \varrho_1 & \varrho_2 & \cdots & \varrho_2 & \varrho_1 \\ \varrho_1 & 1 & \varrho_1 & \cdots & \varrho_3 & \varrho_2 \\ \varrho_2 & \varrho_1 & 1 & \cdots & \varrho_4 & \varrho_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \varrho_2 & \varrho_3 & \varrho_4 & \cdots & 1 & \varrho_1 \\ \varrho_1 & \varrho_2 & \varrho_3 & \cdots & \varrho_1 & 1 \end{pmatrix}.$$

Belongs to the group symmetry models.

The covariance does not change under circular shift of observations

$$\mathrm{Var}[\mathbf{y}] = \mathrm{Var}[\mathbf{P}^r\mathbf{y}], \quad r = 0, 1, \ldots, p-1$$

where

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

Circular Toeplitz structure

Discussed by Olkin and Press (1969) and Olkin (1973)

Olkin and Press (1969) derived MLEs for the parameters and considered likelihood ratio tests (LRT) for testing various hypotheses connected with circular Toeplitz structure.

Olkin (1973) extended the structure to the case of block circular Toeplitz structure (CT structure appeared in blocks, blocks unstructured). Various LRTs were obtained.

Properties of some patterned covariance matrices arising under different symmetry restrictions in the context of linear mixed models studied by Nahtman (2006), Nahtman and von Rosen (2008) or Liang et al. (2020).

Linear structure

Anderson (1969) and Rogers and Young (1977) studied the case

$$\boldsymbol{\Omega} = \theta_1 \mathbf{G}_1 + \cdots + \theta_n \mathbf{G}_n,$$

where θ's are unknown parameters and $\mathbf{G}_1, \ldots, \mathbf{G}_n$ is a set of symmetric linearly independent matrices.

CS and circular Toeplitz structures fall into this class.

When the matrices $\mathbf{G}_1, \ldots, \mathbf{G}_n$ are simultaneously diagonalizable (by an orthogonal matrix independent of θ's), explicit maximum likelihood estimates are readily obtainable.

Linear structure

Ohlson and Rosen (2010) studied the linearly structured covariance matrix in the GCM, i.e. for

$$\mathbf{\Omega} = (\omega_{ij})$$

the only linear structure between the elements is given by $|\omega_{ij}| = |\omega_{kl}|$ and there exists at least one $(i,j) \neq (k,l)$ so that $|\omega_{ij}| = |\omega_{kl}|$.

Derived the least squares estimator and showed its properties (unbiasedness and consistency).

The idea generalized for the Extendend GCM - Nzabanita (2021) used the idea of decomposition of the space generated by design matrices to derive explicit and consistent estimators of the mean and linearly structured covariance matrix.

Linear structure

Diaconis (1989) provides a discussion of how such patterns can arise. Patterns that are exhibited: circular Toeplitz, or circulant indexed by groups, e.g.

$$\mathbf{\Omega} = \begin{pmatrix} a & b & b & b & b & c \\ b & a & b & c & b & b \\ b & b & a & b & c & b \\ b & c & b & a & b & b \\ b & b & c & b & a & b \\ c & b & b & b & b & a \end{pmatrix},$$

where from first to second the switch operation (1,2)(3,4)(5,6) is applied (same holds for row 4 from 3 and 6 from 5).

As pointed by Viana (2003) - a simple determinant of linear structure is to show that $\mathbf{\Omega}_i$ and $\mathbf{\Omega}_j$ commute, or equivalently, that $\mathbf{\Omega}_i\mathbf{\Omega}_j$ is symmetric.

Autoregression covariance structure

Autoregression correlation structure AR(1) (called also serial structure)

$$\boldsymbol{\Omega}_{\mathrm{AR}} = \sigma^2 \left\{ \varrho^{|i-j|} \right\}_{ij} = \sigma^2 \left[\mathbf{I}_p + \sum_{i=1}^{p-1} \varrho^i (\mathbf{C}_p^i + \mathbf{C}_p'^i) \right],$$

where $-1 < \varrho < 1$.

Toeplitz type structure with the restriction $\varrho_i = \varrho^i$ for all i.

Obtained if we assume the first-order autoregressive model of errors. The correlation decline exponentially with distance between observations. Useful for modeling short auto-regression time series.

Autoregression covariance structure

In the context of linear models it was discussed by Pantula and Pollock (1985), Ware (1985), or Jennrich and Schluchter (1986)

Seems to be very natural covariance structure for growth curve data, since such data are usually repeated measurements of very short time series.

In context of GCM the structure was discussed by Hudson (1983), Lee (1988) or Fujikoshi et al. (1990)

Growth curve model with special covariance structures

GCM with CS structure

The MLEs under CS (more precisely under generalized CS) structure were addressed by Khatri (1973) - discussed three hypotheses concerning the covariance structure

- testing the independence of sets of observations
- testing the sphericity,
- testing the generalized CS structure (called intraclass by Khatri)

The likelihood ratio test statistic for testing the generalized CS structure

$$\lambda = \frac{(p-1)^{p-1}|\mathbf{S}||\mathbf{I}_p + \mathbf{S}^{-1}\mathbf{Q}_{Z;S}\mathbf{S}_1|}{|\mathbf{G}|\,\mathrm{Tr}[\mathbf{G}^{-1}\mathbf{P}_{w;G}\mathbf{S}]\Big(\mathrm{Tr}[\mathbf{G}^{-1}\mathbf{Q}_{w;G}\mathbf{S}] + \mathrm{Tr}[\mathbf{G}^{-1}\mathbf{Q}_{Z;G}\mathbf{S}_1]\Big)^{p-1}},$$

where $\mathbf{S} = \mathbf{Y}\mathbf{Q}_X\mathbf{Y}$ and $\mathbf{S}_1 = \mathbf{Y}'\mathbf{P}_X\mathbf{Y}$.

GCM with CS structure

The distribution of λ was determined to be the same as the product of independent beta random variables - no parameters of this beta variables were given

Given all the moments of λ, i.e. $\mathrm{E}[\lambda^h]$ for $h \in \mathbb{N}$ - expressed by means of Gamma and multivariate Gamma functions.

As was pointed in the paper:

> From this, we can obtain the exact distribution of λ by using the inverse Mellin's transform, and the approximate distribution can be obtained by using the results given by Anderson (1958).

GCM with CS structure

Jurková et al. (2020) revised the LRT for testing the CS structure - the LRT statistic

$$\lambda = \frac{(p-1)^{p-1}|\mathbf{S}||\mathbf{I}_p + \mathbf{S}^{-1}\mathbf{Q}_{Z;S}\mathbf{S}_1|}{\mathrm{Tr}[\mathbf{P}_p\mathbf{S}]\Big(\mathrm{Tr}[\mathbf{Q}_p\mathbf{S}] + \mathrm{Tr}[\mathbf{Q}_Z\mathbf{S}_1]\Big)^{p-1}}.$$

Asymptotic distribution of $\Lambda = -n \ln \lambda \overset{asympt.}{\sim} \chi^2_\nu$, with ν being the difference between the number of unknown parameters under alternative and null hypotheses

For large n the approximation works well, however, in practice this test procedure is very often used also for small samples

? Exact distribution ?

GCM with CS structure

Jurková et al. (2020) revised the LRT for testing the CS structure - the LRT statistic

$$\lambda = \frac{(p-1)^{p-1}|\mathbf{S}||\mathbf{I}_p + \mathbf{S}^{-1}\mathbf{Q}_{Z;S}\mathbf{S}_1|}{\mathrm{Tr}[\mathbf{P}_p\mathbf{S}]\Big(\mathrm{Tr}[\mathbf{Q}_p\mathbf{S}] + \mathrm{Tr}[\mathbf{Q}_Z\mathbf{S}_1]\Big)^{p-1}}.$$

Asymptotic distribution of $\Lambda = -n\ln\lambda \overset{asympt.}{\sim} \chi^2_\nu$, with ν being the difference between the number of unknown parameters under alternative and null hypotheses

For large n the approximation works well, however, in practice this test procedure is very often used also for small samples

? Exact distribution ?

in many cases it appeared to be the same as the product of independent beta random variables

GCM with CS structure

Determined all the moments of λ

$$\mathrm{E}\,\lambda^h = (p-1)^{(p-1)h} \cdot \frac{\Gamma_k\left(\frac{n-r-p+k}{2}+h\right)}{\Gamma_k\left(\frac{n-r-p+k}{2}\right)} \cdot \frac{\Gamma_{p-k}\left(\frac{n}{2}+h\right)}{\Gamma_{p-k}\left(\frac{n}{2}\right)} \times$$
$$\times \frac{\Gamma\left(\frac{n-r}{2}\right)}{\Gamma\left(\frac{n-r}{2}+h\right)} \cdot \frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d}{2}+(p-1)h\right)},$$

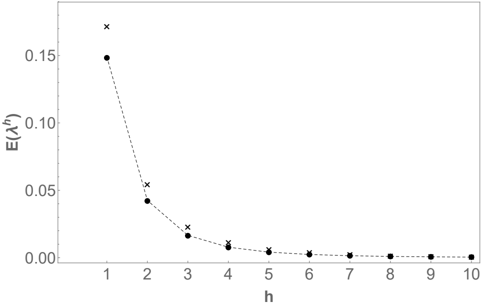
where $d = n(p-1) - r(k-1)$, and r and k being ranks of design matrices $\mathbf{X}$ and $\mathbf{Z}$, respectively.

This formula corrects some errors contained in Khatri's original solution.

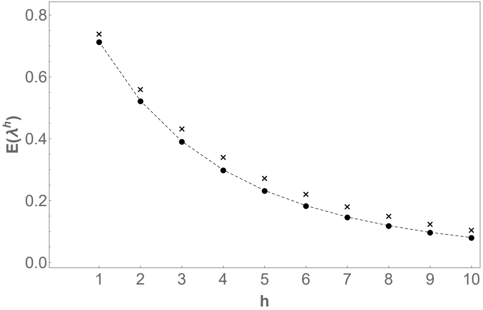
GCM with CS structure

The moments of λ comparison with $p = 4,\ r = 2,\ k = 2$;

circle - corrected, cross - Khatri, dashed line - empirical calculated from 10,000 simulated test statistic.



n=7



n=26

GCM with CS structure

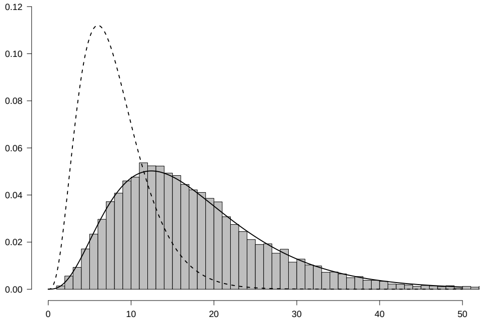
The characteristic function of $\Lambda = -n \ln \lambda$ can be derived

$$\varphi_\Lambda(t) = (p-1)^{-(p-1)itn} \cdot \frac{\Gamma_k\left(\frac{n-r-p+k}{2} - itn\right)}{\Gamma_k\left(\frac{n-r-p+k}{2}\right)} \cdot \frac{\Gamma_{p-k}\left(\frac{n}{2} - itn\right)}{\Gamma_{p-k}\left(\frac{n}{2}\right)} \times$$
$$\times \frac{\Gamma\left(\frac{n-r}{2}\right)}{\Gamma\left(\frac{n-r}{2} - itn\right)} \cdot \frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d}{2} - (p-1)itn\right)}.$$

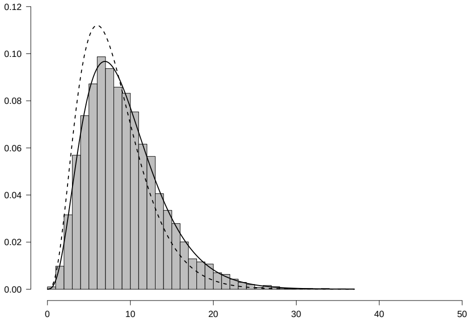
PDF/CDF can be numericaly calculated by the inversion formula using software developed by Witkovský (2018) (Matlab package CharFunTool) or its R location due to Gajdoš (2018) (R package CharFunToolR).

GCM with CS structure

Comparison of the asymptotic χ^2-approximation and exact PDF for $p = 4$, $r = 2$, $k = 2$. The histogram of the empirical distribution is based on 50 000 simulated test statistics.



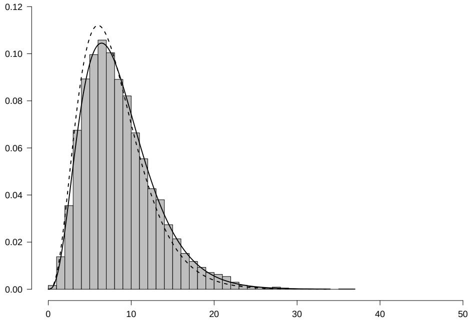
n=7



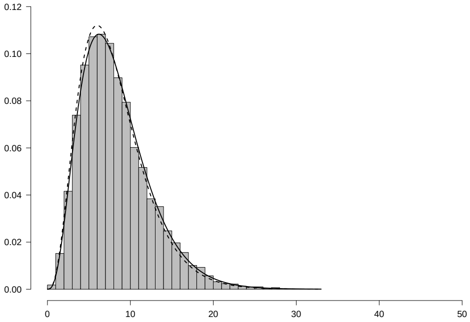
n=25

GCM with CS structure

Comparison of the asymptotic χ^2-approximation and exact PDF for $p = 4$, $r = 2$, $k = 2$. The histogram of the empirical distribution is based on 50 000 simulated test statistics.



n=50



n=100

GCM with CS structure

Other tests of CS structure:

- Rao's score test
$$\lambda = \frac{n}{2} \operatorname{Tr} \left(\widehat{\boldsymbol{\Sigma}}_0^{-1} \widehat{\boldsymbol{\Sigma}} - \mathbf{I} \right)^2$$
where $\widehat{\boldsymbol{\Sigma}}_0$ and $\widehat{\boldsymbol{\Sigma}}$ are MLEs under H_0 and H_1.

GCM with CS structure

Other tests of CS structure:

- Srivastava and Singull (2017) discussed three other possible tests
  1. LRT for CS structure based on the unweighted estimator of **B**

$$\lambda = \frac{(p-1)^{p-1}|\mathbf{S} + \mathbf{Q}_Z \mathbf{S}_1 \mathbf{Q}_Z|}{\operatorname{Tr} \mathbf{P}_p \mathbf{S} (\operatorname{Tr} \mathbf{Q}_p \mathbf{S} + \operatorname{Tr} \mathbf{Q}_Z \mathbf{S}_1)^{p-1}}$$

  2. LRT for CS structure based only on the matrix **S**

$$\lambda = \frac{(p-1)^{p-1}|\mathbf{S}|}{\operatorname{Tr} \mathbf{P}_p \mathbf{S} (\operatorname{Tr} \mathbf{Q}_p \mathbf{S} + \operatorname{Tr} \mathbf{Q}_Z \mathbf{S}_1)^{p-1}}$$

GCM with CS structure

Other tests of CS structure:

- Srivastava and Singull (2017) discussed three other possible tests
  3. Let the matrix $\mathbf{H}_{p\times p}$ be an orthogonal matrix with first column being a normalized column of ones, then transforming the data to $\mathbf{Y}^\star = \mathbf{H}'\mathbf{Y}$ we have

$$\mathbf{Y}^\star \sim N_{n,p}(\mathbf{Z}^\star\mathbf{B}\mathbf{X}', \boldsymbol{\Omega}^\star, \mathbf{I}_n),$$

where $\mathbf{Z}^\star = \mathbf{H}'\mathbf{Z}$ and

$$\boldsymbol{\Omega}^\star = \mathbf{H}'\boldsymbol{\Sigma}\mathbf{H} = \begin{pmatrix} \omega_{11}^\star & \omega_{12}^\star \\ \omega_{21}^\star & \boldsymbol{\Omega}_{22}^\star \end{pmatrix} \overset{H_0}{=} \begin{pmatrix} \omega_2 & \mathbf{0}' \\ \mathbf{0} & \omega_1\mathbf{I}_{p-1} \end{pmatrix}$$

To test the CS structure it is tested the hypothesis

$$H_0 : \ \boldsymbol{\Omega}_{22}^\star = \tilde{\omega}\mathbf{I}_{p-1} \quad \text{vs.} \quad H_1 : \ \boldsymbol{\Omega}_{22}^\star > 0$$

GCM with CS structure

Other tests of CS structure:

- Srivastava and Singull (2017) discussed three other possible tests
  3. Based on unbiased and consistent estimators of $a_1 = \frac{1}{p}\operatorname{Tr}\mathbf{\Omega}$ and $a_2 = \frac{1}{p}\operatorname{Tr}\mathbf{\Omega}^2$, given as

$$\widehat{a}_1 = \frac{1}{np}\operatorname{Tr}\mathbf{S}, \quad \text{and} \quad \widehat{a}_2 = \frac{1}{np(n-1)(n+2)}\left(n\operatorname{Tr}\mathbf{S}^2 - \operatorname{Tr}\mathbf{S}^2\right)$$

The test statistic

$$\lambda = \frac{n}{2}\left(\frac{\widehat{a}_2}{\widehat{a}_1^2 - \frac{2}{np}\widehat{a}_1} - 1\right) \overset{(n,p)\to\infty}{\sim} N(0,1)$$

GCM with CS structure

Power comparison - small simulation study

Setting: $r(\mathbf{X}) = r(\mathbf{Z}) = 2$

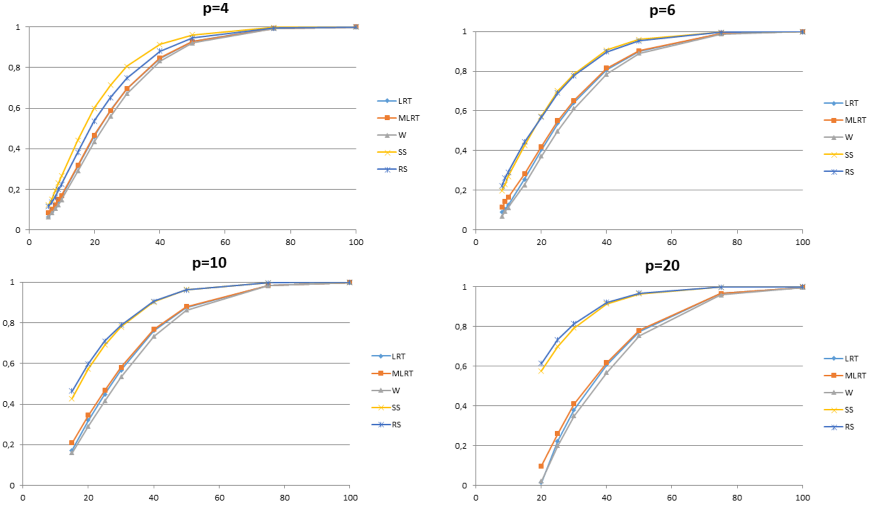
$$\mathbf{Z} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & p \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 20 & 30 \\ 10 & 50 \end{pmatrix} \quad \text{and} \quad \mathbf{X}' = \begin{pmatrix} \mathbf{1}'_{\lfloor n/2 \rfloor} & \mathbf{0}'_{n-\lfloor n/2 \rfloor} \\ \mathbf{0}'_{\lfloor n/2 \rfloor} & \mathbf{1}'_{n-\lfloor n/2 \rfloor} \end{pmatrix}$$

Different true $\mathbf{\Omega}$ as alternative:

1. $\mathbf{\Omega} = \mathbf{DWD}$, where $\mathbf{D} = diag(\sigma_i)$ with $\sigma_i = \sqrt{R(0.9, 1.1)}$ and $\mathbf{W} = (w_{ij})$ with $w_{ij} = (-1)^{i+j} (0.5/\ln p)^{|i-j|^{0.1}}$
2. $\mathbf{\Omega}$ such that in $\mathbf{\Omega}^\star$ there is $\mathbf{\Omega}^\star_{22}$ close to sphericity, but $\omega^\star_{21} \neq \mathbf{0}$
3. Almost CS, just ω_{11} slightly modified
4. $\mathbf{\Omega}$ such that $\mathbf{\Omega}^\star = AR(0.1)$

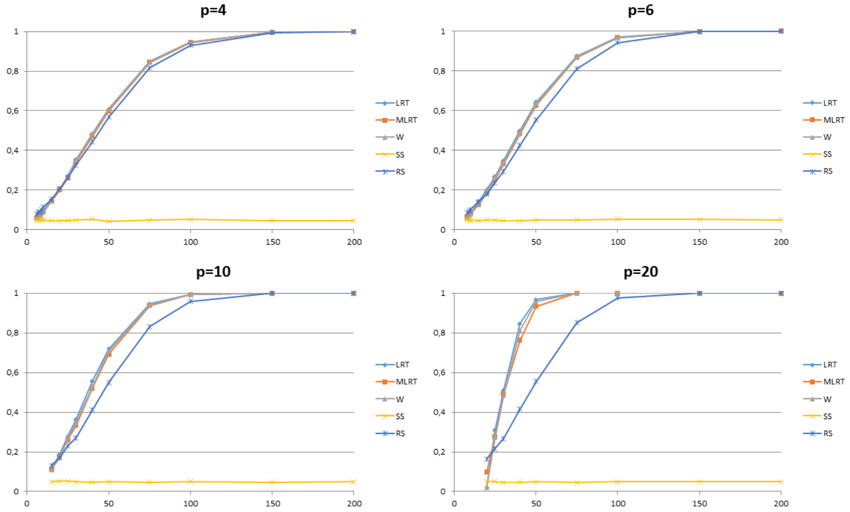
GCM with CS structure

Alternative 1



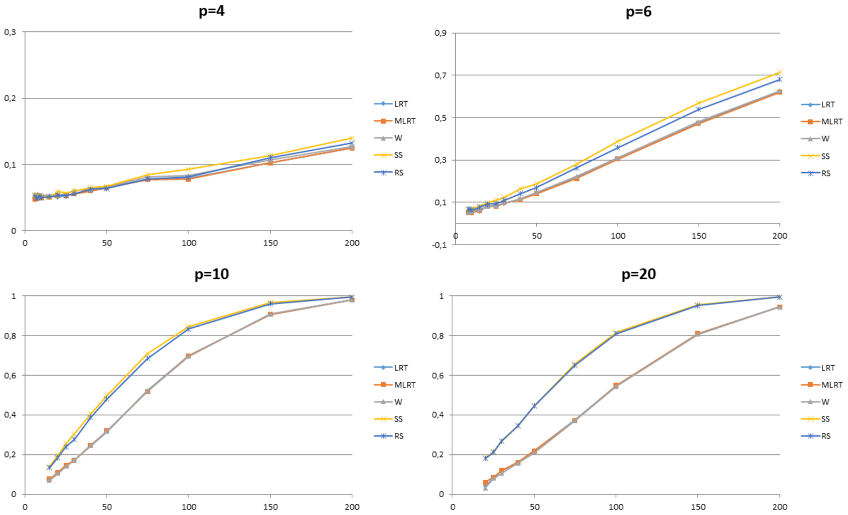
GCM with CS structure

Alternative 2



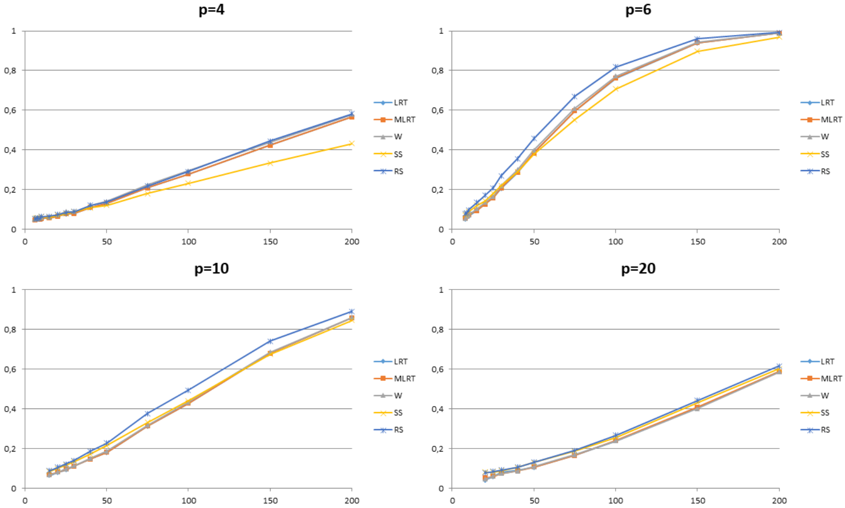
GCM with CS structure

Alternative 3



GCM with CS structure

Alternative 4



GCM with CS structure

Lee (1988) - considered prediction and estimation of growth curves with CS and AR(1) structure

For CS structure determined the MLEs of unknown parameters when $\mathbf{1}_p \in \mathcal{C}(\mathbf{Z})$

$$\begin{aligned}
\widehat{\mathbf{B}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}, \\
\widehat{\sigma}^2 &= \frac{\mathrm{Tr}[\mathbf{S}^*]}{np}, \\
\widehat{\varrho} &= \frac{1}{p-1}\left(\frac{\mathrm{Tr}[\mathbf{J}_p\mathbf{S}^*]}{\mathrm{Tr}[\mathbf{S}^*]} - 1\right),
\end{aligned}$$

where $\mathbf{J}_p = \mathbf{1}_p\mathbf{1}_p'$ and $\mathbf{S}^* = \mathbf{Y}'\mathbf{Q}_X\mathbf{Y} + \mathbf{Q}_Z\mathbf{Y}'\mathbf{P}_X\mathbf{Y}\mathbf{Q}_Z$.

GCM with CS structure

Žežula (2006) and Ye and Wang (2009) proposed different estimators based on moment method

the only difference was the sum of squares matrix used in the covariance parameter estimators

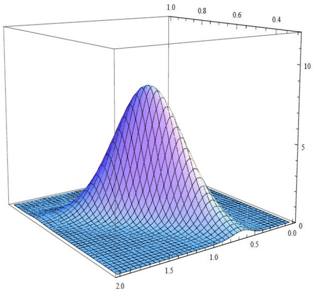
$$\widehat{\sigma}^2 = \frac{\mathrm{Tr}[\mathbf{S}]}{np}, \quad \widehat{\varrho} = \frac{1}{p-1}\left(\frac{\mathrm{Tr}[\mathbf{J}_p\mathbf{S}]}{\mathrm{Tr}[\mathbf{S}]} - 1\right),$$

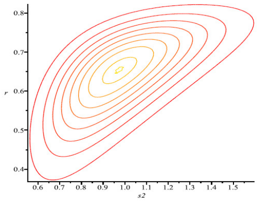
where $\mathbf{S} = \mathbf{Y}'\mathbf{Q}_X\mathbf{Y}$.

Žežula and Klein (2010) showed that the estimators coincide and derived their distributions.

GCM with CS structure

Klein and Žežula (2013) derived the joint density of 2nd order parameters and construct the confidence regions for unknown parameters





Jurková et. al (2020)- discussed an unbiased estimator of correlation coefficient ϱ.

GCM with AR(1) structure

AR(1) structure

$$\mathbf{\Omega}_{\mathrm{AR}} = \sigma^2 \mathbf{R}(\varrho) = \sigma^2 \begin{pmatrix} 1 & \varrho & \varrho^2 & \cdots & \varrho^{p-1} \\ \varrho & 1 & \varrho & \cdots & \varrho^{p-2} \\ \varrho^2 & \varrho & 1 & \cdots & \varrho^{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \varrho^{p-1} & \varrho^{p-2} & \varrho^{p-3} & \cdots & 1 \end{pmatrix},$$

where $-1 < \varrho < 1$.

Among other structures one of the useful - a natural structure for time series and repeated measurements.

A nonlinear structure - there is no analytically closed form of MLEs

GCM with AR(1) structure

Has nice forms of determinant and inverse

$$|\boldsymbol{\Omega}_{\mathrm{AR}}| = (\sigma^2)^p(1-\varrho)^{p-1},$$
$$\boldsymbol{\Omega}_{\mathrm{AR}}^{-1} = \frac{1}{\sigma^2(1-\varrho^2)}\left(\varrho^2\mathbf{C}_1 - \varrho\mathbf{C}_2 + \mathbf{I}_p\right),$$

where

$$\mathbf{C}_1 = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \quad \mathbf{C}_2 = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

GCM with AR(1) structure

Lee (1988) - determined the MLEs of unknown parameters for AR(1) structure

$$\widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{R}(\widehat{\varrho})^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}(\widehat{\varrho})^{-1}\mathbf{Z})^{-1},$$

$$\widehat{\sigma}^2 = \frac{1}{np}\,\mathrm{Tr}\left[\mathbf{R}(\widehat{\varrho})^{-1}\mathbf{S}^*\right],$$

$\widehat{\varrho}$ is found by maximizing profile likelihood $L(\varrho) = |\widehat{\sigma}^2\mathbf{R}(\varrho)|^{-n/2}$,

where $\mathbf{S}^* = \mathbf{Y}'\mathbf{Q}_X\mathbf{Y} + \mathbf{Q}_Z\mathbf{Y}'\mathbf{P}_X\mathbf{Y}\mathbf{Q}_Z$.

MLE of ϱ can be obtained by a one-dimensional search.

GCM with AR(1) structure

Wu (1998) derived the necessary and sufficient conditions for existence of uniformly minimum risk unbiased estimators of the unknown parameters in EGCM - three different structures of covariance matrix were considered

completely UN, CS and AR(1)

Under these conditions

- Wu (2000) derived explicit formulas for MLEs of unknown parameters assuming CS structure,
- Klein and Žežula (2009) derived half-explicit MLEs of unknown parameters assuming AR(1) structure.

GCM with AR(1) structure

For the GCM even with assumptions

$$\mathbf{Q}_Z\mathbf{C}_1\mathbf{Z} = \mathbf{0}, \quad \mathbf{Q}_Z\mathbf{C}_2\mathbf{Z} = \mathbf{0},$$

MLE of σ^2 is expressed in explicit form (as a function of $\widehat{\varrho}$), however, MLE of ϱ is the solution of third order polynomial

$$2(p-1)\,\mathrm{Tr}[\mathbf{C}_1\mathbf{S}^*]\varrho^3 + (2-p)\,\mathrm{Tr}[\mathbf{C}_2\mathbf{S}^*]\varrho^2 - 2\,\mathrm{Tr}[(\mathbf{I}_p + p\mathbf{C}_1)\mathbf{S}^*]\varrho + p\,\mathrm{Tr}[\mathbf{C}_2\mathbf{S}^*] = 0.$$

There is always a unique solution of in the interval $(-1; 1)$ - guarantees positive definitenes of the covariance matrix estimator.

GCM with AR(1) structure

Žežula (2006) proposed estimators based on moment method

the estimator of ϱ is the solution of polynomial of the $(p-1)$-th order.

Fang, Wang and Rosen (2006) - proposed restricted expected multivariate least squares (REMLS) principle for estimation of unknown parameters in in multivariate linear models

- based on the fitting function

$$F(\mathrm{E}\,\mathbf{Y}, \mathbf{\Omega}) = \frac{1}{np}\,\mathrm{Tr}[\mathbf{\Omega}^{-1}(\mathbf{Y} - \mathrm{E}\,\mathbf{Y})'(\mathbf{Y} - \mathrm{E}\,\mathbf{Y})]$$

- the principle is based on minimization of $|F(\mathrm{E}\,\mathbf{Y}, \mathbf{\Omega}) - \mathrm{E}\,F|$ by finding functions $h_1(\mathbf{\Omega})$ and $h_2(\mathrm{E}\,\mathbf{Y}, \mathbf{\Omega})$ such that $F(\mathrm{E}\,\mathbf{Y}, \mathbf{\Omega}) = h_1(\mathbf{\Omega}) + h_2(\mathrm{E}\,\mathbf{Y}, \mathbf{\Omega}))$, and estimators of parameters in $\mathbf{\Omega}$ and $\mathrm{E}\,\mathbf{Y}$ are based on $h_1(\mathbf{\Omega})$ and $h_2(\mathrm{E}\,\mathbf{Y}, \widehat{\mathbf{\Omega}})$, respectively
- this leads to MLEs in the case of normally distributed GCM

GCM with AR(1) structure

This idea can be used for estimating the unknown parameters in multivariate linear model with serial structure - yields explicit estimators

$$\begin{aligned}
\widehat{\mathbf{B}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \\
\widehat{\sigma}^2 &= \frac{\mathrm{Tr}[\mathbf{S}]}{np}, \\
\widehat{\varrho} &= \frac{\mathrm{Tr}[\mathbf{C}_2\mathbf{S}]}{\mathrm{Tr}[(\mathbf{I}_p + \mathbf{C}_1)\mathbf{S}]},
\end{aligned}$$

where $\mathbf{S} = \mathbf{Y}'\mathbf{Q}_X\mathbf{Y}$.

We derived their properties - $\widehat{\sigma}^2$ is unbiased, while $\widehat{\varrho}$ is biased but consistent, we also derived their joint asymptotic distribution

Still not clear how to use the information of within-individual matrix Z in GCM.

Thank you for your attention!

Estimation and testing in multivariate linear models
Part III

Daniel Klein

Institute of Mathematics
Faculty of Science
P. J. Šafárik University in Košice

Będlewo
November 30, 2021

Two-level multivariate data

For third order tensor of observation $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n \times q \times p}$

$$\operatorname{vec} \boldsymbol{\mathcal{Y}} \sim N_{nqp}(\operatorname{vec} \boldsymbol{\mathcal{M}}, \boldsymbol{\Omega} \otimes \mathbf{I}_n),$$

$$\text{where } \boldsymbol{\mathcal{M}} \in \mathbb{R}^{n \times q \times p} \text{ and } \boldsymbol{\Omega}_{qp \times qp} > 0.$$

After matricization - multivariate linear model

$$\mathbf{Y} \sim N_{n,qp}(\mathbf{M}, \mathbf{I}_n, \boldsymbol{\Omega}), \quad \text{with} \quad \mathbf{M} \in \mathbb{R}^{n \times qp}.$$

? Structure of $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\Omega}$?

Structure of $\boldsymbol{\Omega}$

Reasonable structure

- separable structure: $\boldsymbol{\Omega} = \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}$

$\boldsymbol{\Sigma}_{q \times q}$	the covariance matrix of q characteristics at any given time point; assumed to be the same for all time points
$\boldsymbol{\Psi}_{p \times p}$	the covariance matrix of p repeated measurements on a given characteristic; assumed to be the same for all characteristics

  - since for any $c > 0$

    $$\boldsymbol{\Psi} \otimes \boldsymbol{\Sigma} = c\boldsymbol{\Psi} \otimes \frac{1}{c}\boldsymbol{\Sigma}$$

    to circumvent the identifiability problem Srivastava et al. (2008) proposed to fix e.g. $\psi_{11} = 1$
  - number of unknown parameters: $\frac{p(p+1)}{2} + \frac{q(q+1)}{2} - 1$

Structure of $\boldsymbol{\Omega}$

Reasonable structure

- separable structure with one component structured as CS or AR(1):

  $\boldsymbol{\Omega} = \boldsymbol{\Psi} \otimes \boldsymbol{\Sigma}$ with, say, $\boldsymbol{\Psi}$ being additionally structured, i.e. $\boldsymbol{\Psi}_{\text{CS}}$ or $\boldsymbol{\Psi}_{\text{AR}}$

  - both $\boldsymbol{\Psi}_{\text{CS}}$ or $\boldsymbol{\Psi}_{\text{AR}}$ are taken as correlation matrices
  - number of unknown parameters: $\frac{q(q+1)}{2} + 1$

Structure of $\mathbf{\Omega}$

Reasonable structure

- block compound symmetry (BCS) structure:

$$\mathbf{\Omega}_{\text{BCS}} = \mathbf{I}_p \otimes (\mathbf{\Gamma}_0 - \mathbf{\Gamma}_1) + \mathbf{J}_p \otimes \mathbf{\Gamma}_1 = \begin{pmatrix} \mathbf{\Gamma}_0 & \mathbf{\Gamma}_1 & \cdots & \mathbf{\Gamma}_1 \\ \mathbf{\Gamma}_1 & \mathbf{\Gamma}_0 & \cdots & \mathbf{\Gamma}_1 \\ \vdots & & \ddots & \vdots \\ \mathbf{\Gamma}_1 & \mathbf{\Gamma}_1 & \cdots & \mathbf{\Gamma}_0 \end{pmatrix},$$

$\mathbf{\Gamma}_0$	$q \times q$ covariance matrix of q characteristics at any repeated measurement; assumed to be the same for repeated measurements
$\mathbf{\Gamma}_1$	$q \times q$ covariance matrix of p repeated measurements on a given characteristic; assumed to be the same for all characteristics

- called also exchangeable, as the columns of the data matrix $\mathbf{Y}_i$ may be exchanged without changing the covariance matrix

Structure of $\boldsymbol{\Omega}$

Reasonable structure

- block compound symmetry (BCS) structure:

$$\boldsymbol{\Omega}_{\text{BCS}} = \mathbf{I}_p \otimes (\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1) + \mathbf{J}_p \otimes \boldsymbol{\Gamma}_1 = \begin{pmatrix} \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\Gamma}_1 \\ \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_0 & \cdots & \boldsymbol{\Gamma}_1 \\ \vdots & & \ddots & \vdots \\ \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\Gamma}_0 \end{pmatrix},$$

  - reparametrization

$$\boldsymbol{\Omega}_{\text{BCS}} = \mathbf{P}_p \otimes \boldsymbol{\Delta}_1 + \mathbf{Q}_p \otimes \boldsymbol{\Delta}_2,$$

    where $\boldsymbol{\Delta}_1 = \boldsymbol{\Gamma}_0 + (p-1)\boldsymbol{\Gamma}_1 > 0$, $\boldsymbol{\Delta}_2 = \boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1 > 0$
  - $\boldsymbol{\Psi}_{\text{CS}} \otimes \boldsymbol{\Sigma}$ is a special case of BCS with parameter space restriction $\boldsymbol{\Gamma}_1 = \varrho\boldsymbol{\Gamma}_0$
  - number of unknown parameters: $q(q+1)$

Structure of mean tensor $\mathcal{M}$

Structure of the mean tensor $\mathcal{M} = \mathrm{E}\,[\mathcal{Y}]$:

- generalized growth curve model (Filipiak and Klein (2017)):

$$\mathrm{E}\,[\mathrm{vec}\,\mathcal{Y}] = (\mathbf{U} \otimes \mathbf{Z} \otimes \mathbf{X})\,\mathrm{vec}\,\mathcal{B},$$

$\mathcal{B}_{n_1 \times q_1 \times p_1}$	the third order tensor of unknown constants
$\mathbf{X}_{n \times n_1}$, $\mathbf{Z}_{q \times q_1}$, $\mathbf{U}_{p \times p_1}$	known design matrices

  - alternatively (without vec operator)

$$\mathrm{E}\,[\mathcal{Y}] = [\![\mathcal{B}; \mathbf{X}, \mathbf{Z}, \mathbf{U}]\!]$$

where $[\![\mathcal{B}; \mathbf{X}, \mathbf{Z}, \mathbf{U}]\!]$ denotes the Tucker operator defined as (Kolda (2006) or Savas and Lim (2008))

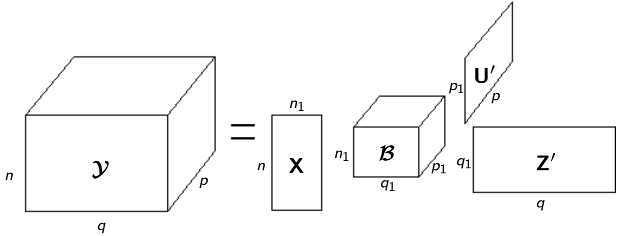
$$[\![\mathcal{B}; \mathbf{X}, \mathbf{Z}, \mathbf{U}]\!]_{ijk} = \sum_{\alpha=1}^{n_1} \sum_{\beta=1}^{q_1} \sum_{\gamma=1}^{p_1} x_{i\alpha} z_{j\beta} u_{k\gamma} b_{\alpha\beta\gamma}.$$

Structure of mean tensor $\mathcal{M}$

Structure of the mean tensor $\mathcal{M} = \mathrm{E}[\mathcal{Y}]$:

- generalized growth curve model:

$$\mathrm{E}[\mathcal{Y}] = [\![\mathcal{B}; \mathbf{X}, \mathbf{Z}, \mathbf{U}]\!]$$



Structure of mean tensor $\mathcal{M}$

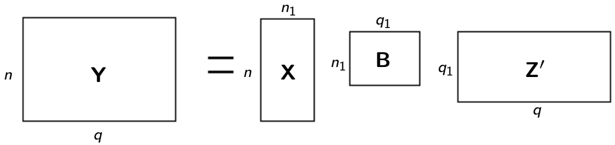
Structure of the mean tensor $\mathcal{M} = \mathrm{E}[\mathcal{Y}]$:

- generalized growth curve model:

$$\mathrm{E}[\mathcal{Y}] = [\![\mathcal{B}; \mathbf{X}, \mathbf{Z}, \mathbf{U}]\!]$$

  - referred to as the trilinear structure - an extension of the GCM bilinear structure

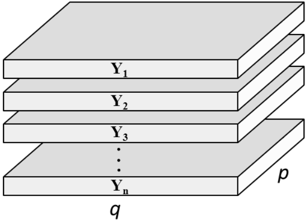
$$\mathbf{XBZ}' = [\![\mathbf{B}; \mathbf{X}, \mathbf{Z}]\!]$$



Structure of mean tensor $\mathcal{M}$

Example: roles of matrices **X**, **Z** and **U**

- The temperature of n lakes measured at q depth levels repeatedly over p time points.
- The measurements for i-th lake, $i \in \{1, \ldots, n\}$, can be arranged into a $q \times p$ matrix Y_i representing spatio-temporal measurements.



Structure of mean tensor $\mathcal{M}$

Example: roles of matrices $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{U}$

- Assuming a polynomial trend of order $q_1 - 1$ in depth and a polynomial trend of order $p_1 - 1$ in time

$$\mathrm{E}\left[\mathbf{Y}_i\right] = \mathbf{Z}\mathbf{B}_i\mathbf{U}' = [\![\mathbf{B}_i; \mathbf{Z}, \mathbf{U}]\!]\,,$$

where

$$\mathbf{Z} = \begin{pmatrix} 1 & z_1 & \cdots & z_1^{q_1-1} \\ 1 & z_2 & \cdots & z_2^{q_1-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_q & \cdots & z_q^{q_1-1} \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} 1 & u_1 & \cdots & u_1^{p_1-1} \\ 1 & u_2 & \cdots & u_2^{p_1-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & u_p & \cdots & u_p^{p_1-1} \end{pmatrix},$$

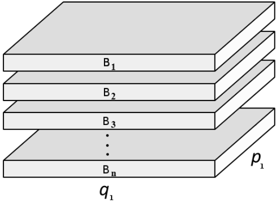
and

$\mathbf{B}_i$ $\quad q_1 \times p_1$ matrix of unknown parameters for the i-th lake

Structure of mean tensor $\mathcal{M}$

Example: roles of matrices $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{U}$

- In the univariate/multivariate model - the trend between the independent observations is modeled via the design matrix $\mathbf{X}$.
- Similarly we may want to model the trend between the lakes e.g. from different regions, which may have different mean.
- For tensor of unknown parameters $\mathcal{B}$



$$\mathbf{B}_1 = \cdots = \mathbf{B}_{a_1}$$
$$\vdots$$
$$\mathbf{B}_{n-a_{n_1}} = \cdots = \mathbf{B}_n$$

$$\sum_{i=1}^{n_1} a_i = n$$

Structure of mean tensor $\mathcal{M}$

Example: roles of matrices $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{U}$

- Tensor $\mathcal{B}$ will be multiplied by the design matrix $\mathbf{X}_{n\times n_1}$ from its n_1-mode.
- The model

$$\mathrm{E}\left[\mathcal{Y}\right] = [\![\mathcal{B}; \mathbf{X}, \mathbf{Z}, \mathbf{U}]\!] \,,$$

$\mathbf{X}_{n\times n_1}$ block diagonal matrix with blocks $\mathbf{1}_{a_i}$

- **Remark**: the matrices $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{U}$ do not need to be Vandermonde or binary matrices; their forms depend on the experiment under consideration.

- The generalized growth curve model in matricized form

$$\mathbf{Y}_{n,qp} \sim N_{n,qp}\Big(\mathbf{XB}(\mathbf{U}' \otimes \mathbf{Z}'), \mathbf{I}_n, \boldsymbol{\Omega}\Big),$$

where $\mathbf{B}_{n_1\times q_1 p_1}$ is matricized $\mathcal{B}$ in the same way as $\mathcal{Y}$

Structure of mean tensor $\mathcal{M}$

Example: roles of matrices $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{U}$

- Assuming the same unknown parameter matrices of all the lakes, i.e. $\mathbf{B}_1 = \cdots = \mathbf{B}_n$ - the model considered by Srivastava et al. (2009)

$$\mathrm{E}(\mathcal{Y}) = [\![\mathcal{B}; \mathbf{1}_n, \mathbf{Z}, \mathbf{U}]\!],$$

  where $\mathcal{B}_{1\times q_1 \times p_1}$ consists of only one horizontal slice.
- model in matricized form

$$\mathbf{Y}_{n,qp} \sim N_{n,qp}\Big(\mathbf{1}_n\mathbf{B}(\mathbf{U}' \otimes \mathbf{Z}'), \mathbf{I}_n, \mathbf{\Omega}\Big),$$

  where $\mathbf{\Omega}$ was assumed to have separable structure

Structure of mean tensor $\mathcal{M}$

Example: roles of matrices $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{U}$

- We may relax on assumption of modeling the growth in the q and p mode - there will be the restrictions on the parameters only through independent units

$$\mathrm{E}\left[\mathcal{Y}\right] = [\![\mathcal{B}; \mathbf{X}, \mathbf{I}_q, \mathbf{I}_p]\!]$$

$\mathcal{B}$	$n_1 \times q \times p$ tensor of unknown parameters
$\mathbf{X}$	$n \times n_1$ known design matrix

- two-level multivariate regression model, which, in contrast to classical multivariate regression model, have matrix valued observations instead of vector valued

$$\mathbf{Y}_{n,qp} \sim N_{n,qp}\Big(\mathbf{XB}, \mathbf{I}_n, \mathbf{\Omega}\Big),$$

$\mathbf{\Omega}$	assumed to have e.g. separable structure

Structure of mean tensor $\mathcal{M}$

Example: roles of matrices $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{U}$

- two-level multivariate MANOVA model - $\mathbf{X}$ is just a binary matrix
- two-level multivariate general mean model - $\mathbf{X} = \mathbf{1}_n$

$$\mathbf{Y}_{n,qp} \sim N_{n,qp}\left(\mathbf{1}_n\boldsymbol{\mu}', \mathbf{I}_n, \boldsymbol{\Omega}\right),$$

$\boldsymbol{\Omega}$	assumed to have e.g. separable structure
$\boldsymbol{\mu}$	qp-dimensional vector of unknown parameters

Two-level multivariate model with separable covariance structure

Two-level model with separable structure

$$\mathbf{Y}_{n\times qp} \sim N_{n,qp}(\mathbf{M}, \mathbf{I}_n, \mathbf{\Psi} \otimes \mathbf{\Sigma}) \quad \to \quad \mathcal{Y} \sim N_{n,q,p}(\mathcal{M}, \mathbf{I}_n, \mathbf{\Sigma}, \mathbf{\Psi}),$$

- rows of $\mathbf{Y}$ are vectorized matrices $\mathbf{Y}_i : q \times p$, $i \in \{1, \ldots, n\}$

$$\mathrm{Var}[\mathrm{vec}\,\mathbf{Y}_i] = \mathbf{\Psi} \otimes \mathbf{\Sigma}$$

- studied by several authors: Naik and Rao (2001); Roy and Khattree (2003); Lu and Zimmerman (2005); Mitchell et al. (2005, 2006); Srivastava et al. (2008)

Two-level model with separable structure -MLEs

General mean model: $\mathbf{M} = \mathbf{1}_n \boldsymbol{\mu}'$

- MLEs (presented in Lu and Zimmermann (2005))

$$\begin{aligned}
\widehat{\boldsymbol{\mu}} &= \bar{\mathbf{Y}} = \tfrac{1}{n}\mathbf{Y}'\mathbf{1}_n, \\
\widehat{\boldsymbol{\Psi}} &= \frac{1}{nq}\sum_{i=1}^{n}\sum_{u=1}^{q}\sum_{v=1}^{q} \boldsymbol{\sigma}^*_{uv}(\mathbf{Y}_{iv} - \bar{\mathbf{Y}}_v)(\mathbf{Y}_{iu} - \bar{\mathbf{Y}}_u)', \\
\widehat{\boldsymbol{\Sigma}} &= \frac{1}{np}\sum_{i=1}^{n}\sum_{u=1}^{p}\sum_{v=1}^{p} \psi^*_{uv}(\mathbf{Y}^*_{iv} - \bar{\mathbf{Y}}^*_v)(\mathbf{Y}^*_{iu} - \bar{\mathbf{Y}}^*_u)',
\end{aligned}$$

where $\mathbf{Y}_{iv}$ and $\mathbf{Y}^*_{iv}$ are v-th row and v-th column of $\mathbf{Y}_i$, respectively,
ψ^*_{uv} and $\boldsymbol{\sigma}^*_{uv}$ are (u, v)-th elements of $\boldsymbol{\Psi}^{-1}$ and $\boldsymbol{\Sigma}^{-1}$, respectively,
$\bar{\mathbf{Y}}_v = \frac{1}{n}\sum_{i=1}^{n}\mathbf{Y}_{iv}$ and $\bar{\mathbf{Y}}^*_v = \frac{1}{n}\sum_{i=1}^{n}\mathbf{Y}^*_{iv}$

Two-level model with separable structure -MLEs

MANOVA model: $\mathbf{M} = \mathbf{XB}$

- MLEs (presented in Mitchell et al. (2006))

$$
\begin{aligned}
\widehat{\mathbf{B}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \\
\widehat{\mathbf{\Psi}} &= \frac{1}{nq}\sum_{i=1}^{n}(\mathbf{Y}_i - \widehat{\mathbf{M}}_i)'\widehat{\mathbf{\Sigma}}^{-1}(\mathbf{Y}_i - \widehat{\mathbf{M}}_i), \\
\widehat{\mathbf{\Sigma}} &= \frac{1}{np}\sum_{i=1}^{n}(\mathbf{Y}_i - \widehat{\mathbf{M}}_i)\widehat{\mathbf{\Psi}}^{-1}(\mathbf{Y}_i - \widehat{\mathbf{M}}_i)',
\end{aligned}
$$

where $\widehat{\mathbf{M}}_i$ is the i-th reshaped column of $\mathbf{X}\widehat{\mathbf{B}} = \mathbf{P}_X\mathbf{Y}$

Two-level model with separable structure -MLEs

GCM model: $\mathbf{M} = \mathbf{1}_n \operatorname{vec}'(\mathbf{ZBU}')$

- MLEs (presented in Srivastava et al. (2009)): let denote $\mathbf{V}_{q\times pn} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$, then

$$\begin{aligned}
\widehat{\mathbf{X}} &= \tfrac{1}{n}(\mathbf{Z}'\mathbf{S}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{S}^{-1}\mathbf{V}(\mathbf{1}_n \otimes \widehat{\mathbf{\Psi}}^{-1}\mathbf{U}(\mathbf{U}'\widehat{\mathbf{\Psi}}^{-1}\mathbf{U})^{-1}),\\
\widehat{\mathbf{\Sigma}} &= \tfrac{1}{np}(\mathbf{V} - \mathbf{Z}\widehat{\mathbf{B}}(\mathbf{1}_n' \otimes \mathbf{U}'))(\mathbf{I}_n \otimes \widehat{\mathbf{\Psi}}^{-1})(\mathbf{V} - \mathbf{Z}\widehat{\mathbf{B}}(\mathbf{1}_n' \otimes \mathbf{U}'))',\\
\widehat{\mathbf{\Psi}} &= \tfrac{1}{nq}\sum_{i=1}^{n}(\mathbf{Y}_i - \mathbf{Z}\widehat{\mathbf{B}}\mathbf{U}')'\widehat{\mathbf{\Sigma}}^{-1}(\mathbf{Y}_i - \mathbf{Z}\widehat{\mathbf{B}}\mathbf{U}'),
\end{aligned}$$

where $\mathbf{S} = \mathbf{V}(\mathbf{I}_n \otimes \widehat{\mathbf{\Psi}}^{-1} - \mathbf{P}_n \otimes \widehat{\mathbf{\Psi}}^{-1}\mathbf{U}(\mathbf{U}'\widehat{\mathbf{\Psi}}^{-1}\mathbf{U})^{-1}\mathbf{U}'\widehat{\mathbf{\Psi}}^{-1})\mathbf{V}'$.

Two-level model with separable structure -MLEs

Filipiak et al. (2018) - Block trace and Partial trace operators:

Definition

For any matrix $\mathbf{A} = (\mathbf{A}_{ij})$ of order qp we define

(i) **block trace matrix** $\mathrm{BTr}_p\,\mathbf{A}$ as the sum of all diagonal $p \times p$ blocks, i.e.

$$\mathrm{BTr}_p\,\mathbf{A} = \sum_{i=1}^{p} \mathbf{A}_{ii}.$$

(ii) **partial trace matrix** $\mathrm{PTr}_q\,\mathbf{A}$ as the matrix of traces of all $q \times q$ blocks, i.e.

$$\mathrm{PTr}_q\,\mathbf{A} = (\mathrm{Tr}\,\mathbf{A}_{ij}) \quad i, j = 1, 2, \ldots, p.$$

Two-level model with separable structure -MLEs

MLEs of $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ using BTr and PTr

- MLEs (using BTr and PTr operators)

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{np}\mathrm{BTr}_q[(\widehat{\boldsymbol{\Psi}}^{-1} \otimes \mathbf{I}_q)\mathbf{S}],$$
$$\widehat{\boldsymbol{\Psi}} = \frac{1}{nq}\mathrm{PTr}_q[(\mathbf{I}_p \otimes \widehat{\boldsymbol{\Sigma}}^{-1})\mathbf{S}],$$

where $\mathbf{S} = (\mathbf{Y} - \widehat{\mathbf{M}})(\mathbf{Y} - \widehat{\mathbf{M}})'$.

- To find the solution so-called "flip-flop" algorithm can be used.

Two-level model with separable structure -MLEs

generalized GCM model: $\mathcal{M} = [\![\mathcal{B}; \mathbf{X}, \mathbf{Z}, \mathbf{U}]\!]$

- MLEs (presented in Filipiak and Klein (2017))

$$\begin{aligned}
\widehat{\mathcal{B}} &= [\![\mathcal{Y}; (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}', (\mathbf{Z}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\widehat{\boldsymbol{\Sigma}}^{-1}, (\mathbf{U}'\mathbf{S}_1^{-1}\mathbf{U})^{-1}\mathbf{U}'\mathbf{S}_1^{-1}]\!], \\
nq\widehat{\boldsymbol{\Psi}} &= \mathbf{S}_1 + \mathbf{Q}_{U;S_1}\mathbf{S}_2\mathbf{Q}'_{U;S_1}, \\
np\widehat{\boldsymbol{\Sigma}} &= \sum_{j=1}^{p}\sum_{\ell=1}^{p} \mathrm{PTr}_n\left(s_{j\ell}\widehat{\boldsymbol{\Xi}}_{j\ell}\right),
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{S}_1 &= \mathrm{PTr}_{nq}\left[\{\mathbf{I}_p \otimes (\widehat{\boldsymbol{\Sigma}}^{-1/2} \otimes \mathbf{I}_n)\mathbf{Q}_{(\widehat{\Sigma}^{-1/2}Z\otimes X)}(\widehat{\boldsymbol{\Sigma}}^{-1/2} \otimes \mathbf{I}_n)\} \operatorname{vec}\mathcal{Y} \operatorname{vec}'\mathcal{Y}\right], \\
\mathbf{S}_2 &= \mathrm{PTr}_{nq}\left[\{\mathbf{I}_p \otimes (\widehat{\boldsymbol{\Sigma}}^{-1/2} \otimes \mathbf{I}_n)\mathbf{P}_{(\widehat{\Sigma}^{-1/2}Z\otimes X)}(\widehat{\boldsymbol{\Sigma}}^{-1/2} \otimes \mathbf{I}_n)\} \operatorname{vec}\mathcal{Y} \operatorname{vec}'\mathcal{Y}\right]
\end{aligned}$$

and $s_{j\ell}$ and $\widehat{\boldsymbol{\Xi}}_{j\ell}$ are the (j,ℓ)-th element of $\widehat{\boldsymbol{\Psi}}^{-1}$ and the $nq \times nq$ block of

$$\operatorname{vec}\left(\mathcal{Y} - [\![\widehat{\mathcal{B}}; \mathbf{X}, \mathbf{Z}, \mathbf{U}]\!]\right) \operatorname{vec}'\left(\mathcal{Y} - [\![\widehat{\mathcal{B}}; \mathbf{X}, \mathbf{Z}, \mathbf{U}]\!]\right).$$

Generalized GCM example

Table: Selected data from the integrated monitoring of the effects of liming project at SLU, Sweden: temperature measured in three years (1990, 2000, 2009) (p=3) at three depth levels (0.5 m, 5 m, 15 m) (q=3); first seven lakes are on the north, remaining 10 are on the south (n=17)

depth 0.5				depth 5				depth 15			
lake	1990	2000	2009	lake	1990	2000	2009	lake	1990	2000	2009
1	7.7	5.1	7.4	1	7.6	5.0	7.2	1	6.9	5.0	6.8
2	15.7	11.8	11.2	2	9.9	11.8	11.0	2	8.5	12.0	10.5
3	13.2	9.4	10.8	3	11.3	9.4	10.0	3	10.1	9.3	8.5
4	15.6	12.7	12.1	4	13.4	12.7	11.8	4	12.6	12.6	11.4
5	15.0	13.0	12.9	5	14.7	12.8	12.4	5	8.9	9.6	8.2
6	18.5	16.5	19.6	6	13.7	11.5	14.5	6	5.2	4.3	4.7
7	15.6	13.5	13.3	7	9.9	12.7	12.0	7	6.4	6.3	7.6
8	15.7	18.5	19.0	8	15.6	14.9	14.6	8	14.9	13.8	13.0
9	19.2	19.0	18.3	9	15.9	17.2	16.0	9	5.6	4.9	3.8
10	19.0	15.7	16.3	10	11.0	9.5	8.8	10	6.0	4.7	3.8
11	18.4	16.7	17.6	11	11.8	12.6	11.8	11	4.8	5.4	4.1
12	17.8	15.0	15.9	12	17.1	13.5	14.3	12	8.0	4.2	4.0
13	20.2	15.7	14.6	13	14.2	15.6	13.7	13	6.2	5.0	4.6
14	18.7	16.6	15.9	14	16.4	12.5	9.7	14	8.4	6.6	6.0
15	17.5	14.3	15.9	15	16.0	14.1	14.4	15	9.2	5.8	5.7
16	18.5	16.1	15.1	16	14.5	15.5	9.5	16	8.4	5.8	5.8
17	16.2	15.1	14.2	17	16.2	15.0	13.3	17	6.9	5.2	5.3

Generalized GCM example

- The measurements from one lake can be arranged into a $q \times p$ matrix $\mathbf{Y}_i$, $i = 1, \ldots, n$, representing a spatio-temporal measurements.
- Assuming a polynomial trend of order $q_1 - 1$ in depth and a polynomial trend of order $p_1 - 1$ in time we have a model (Srivastava et al., 2009)

$$\mathrm{E}\left(\mathbf{Y}_i\right) = \mathbf{Z}\mathbf{B}_i\mathbf{U}' = [\![\mathbf{B}_i; \mathbf{Z}, \mathbf{U}]\!]$$

- The model in tensor notation

$$\mathrm{E}\left(\mathcal{Y}\right) = [\![\mathcal{B}; \mathbf{1}_n, \mathbf{Z}, \mathbf{U}]\!]$$

$$\mathbf{Z} = \begin{pmatrix} 1 & z_1 & \cdots & z_1^{q_1-1} \\ 1 & z_2 & \cdots & z_2^{q_1-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_p & \cdots & z_q^{q_1-1} \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} 1 & u_1 & \cdots & u_1^{p_1-1} \\ 1 & u_2 & \cdots & u_2^{p_1-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & u_p & \cdots & u_p^{p_1-1} \end{pmatrix}$$

- $\mathcal{B}_{1 \times q_1 \times p_1}$ consists of only one horizontal slice which reflects the unknown parameters for the mean of each horizontal slice of measurement tensor $\mathcal{Y}$, i.e., for each lake.

Generalized GCM example

- We want to model also the trend between the lakes from different regions, which may have different mean.
- The number of horizontal slices in tensor $\mathcal{B}$ will correspond to the number of different regions n_1
- The model

$$\mathrm{E}\left(\mathcal{Y}\right) = [\![\mathcal{B}; \mathbf{X}, \mathbf{Z}, \mathbf{U}]\!], \quad \mathcal{B} : n_1 \times q_1 \times p_1$$

- $\mathbf{X} = diag(\mathbf{1}_{a_1}, \mathbf{1}_{a_2}, \ldots, \mathbf{1}_{a_{n_1}})$, $\sum_{i=1}^{n_1} a_i = n$

Generalized GCM example

- The variable of interest - temperature
- number of lakes - $n = 17$

  $a_1 = 7$ - lakes above the 60th parallel

  $a_2 = 10$ - lakes below the 60th parallel
- $q = 3$ - number of depth levels, where measurements were taken (0.5, 5, and 15 m)
- $p = 3$ - number of time points, when measurements were taken (1990, 2000 and 2009)

Generalized GCM example

- The design matrices

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_7 & \mathbf{0}_7 \\ \mathbf{0}_{10} & \mathbf{1}_{10} \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} 1 & 0.5 \\ 1 & 5 \\ 1 & 15 \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} 1 & 0 \\ 1 & 10 \\ 1 & 19 \end{pmatrix}$$

- Applying the flip-flop algorithm we obtained tensor $\boldsymbol{\mathcal{B}}$ represented as matrix $\widehat{\mathbf{B}}_{ij1}$ - the front slice - and matrix $\widehat{\mathbf{B}}_{ij2}$ - the back one:

$$\left(\widehat{\mathbf{B}}_{ij1} \mid \widehat{\mathbf{B}}_{ij2}\right) = \left(\begin{array}{cc|cc} 13.949 & -0.370 & -0.054 & 0.003 \\ 18.348 & -0.708 & -0.092 & -0.001 \end{array}\right)$$

$$\widehat{\boldsymbol{\Psi}} = \begin{pmatrix} 1.000 & 0.381 & 0.015 \\ 0.381 & 0.988 & 0.185 \\ 0.015 & 0.185 & 0.574 \end{pmatrix}, \qquad \widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 7.000 & 5.543 & 5.255 \\ 5.543 & 8.527 & 6.585 \\ 5.255 & 6.585 & 6.992 \end{pmatrix}$$

Generalized GCM example

- Interpretation

$$\left(\widehat{\mathbf{B}}_{ij1} \,\middle|\, \widehat{\mathbf{B}}_{ij2}\right) = \left(\begin{array}{cc|cc} 13.949 & -0.370 & -0.054 & 0.003 \\ 18.348 & -0.708 & -0.092 & -0.001 \end{array}\right)$$

$\hat{b}_{111}$ $= 13.949$ – intercept for northern lakes (average temperature at the beginning of the experiment in 1990, at depth zero),
$\hat{b}_{121}$ $= -0.37$ – linear trend in depth for northern lakes,
$\hat{b}_{112}$ $= -0.054$ – linear trend in years for northern lakes,
$\hat{b}_{122}$ $= 0.003$ – interaction between depth and year for northern lakes,

- We have the following regression for northern lakes (d - depth, t - time):

$$y = 13.949 - 0.054\,t - 0.37\,d + 0.003\,t\,d$$

Generalized GCM example

- Interpretation

$$\left(\widehat{\mathbf{B}}_{ij1} \mid \widehat{\mathbf{B}}_{ij2}\right) = \left(\begin{array}{cc|cc} 13.949 & -0.370 & -0.054 & 0.003 \\ 18.348 & -0.708 & -0.092 & -0.001 \end{array}\right)$$

$\hat{b}_{211} = 18.348$ – intercept for southern lakes (average temperature at the beginning of the experiment in 1990, at depth zero),
$\hat{b}_{221} = -0.708$ – linear trend in depth for southern lakes,
$\hat{b}_{212} = -0.092$ – linear trend in years for southern lakes,
$\hat{b}_{222} = -0.001$ – interaction between depth and year for southern lakes,

- We have the following regression for southern lakes (d - depth, t - time):

$$y = 18.348 - 0.708\, d - 0.092\, t - 0.001\, t\, d$$

Testing the covariance structure in two-level multivariate data

Introduction

Assume

$$\mathbf{Y}_{n,qp} \sim N_{n,qp}(\mathbf{1}_n\boldsymbol{\mu}', \mathbf{I}_n, \boldsymbol{\Omega})$$

where $\boldsymbol{\mu} \in \mathbb{R}^{qp}$.

We want to test test the hypothesis

$$H_0 : \boldsymbol{\Omega} = \boldsymbol{\Omega}_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\Omega} \text{ unstructured}$$

where $\boldsymbol{\Omega}_0$ is

- $\boldsymbol{\Psi}_{\text{UN}} \otimes \boldsymbol{\Sigma}$ - test of separability,
- $\boldsymbol{\Psi}_{\text{CS}} \otimes \boldsymbol{\Sigma}$ - test of separability with CS component,
- $\boldsymbol{\Psi}_{\text{AR}} \otimes \boldsymbol{\Sigma}$ - test of separability with AR(1) component,
- $\boldsymbol{\Gamma} = \mathbf{I}_p \otimes (\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1) + \mathbf{J}_p \otimes \boldsymbol{\Sigma}_1$ - test of BCS structure.

Testing hypotheses for large samples

Most common hypotheses testing procedures for large samples (referred to as *Holy trinity* in statistical literature):

- likelihood ratio test (Wilks, 1938)
- Wald test (Wald, 1943)
- Rao score test (Rao, 1948)

All tests are equivalent to the first-order asymptotics, but differ to some extent in the second-order properties.

Widely used even for small samples n, since exact test are not always available $\rightarrow$ this result in erroneous conclusions as χ^2 distribution is generally an inadequate approximation.

One can use the empirical null distribution of the test statistic.

Testing covariance structure

Testing separable covariance structure with both unstructured components:

- Dutilleul (1999), Naik and Rao (2001), Roy and Khattree (2003), Lu and Zimmerman (2005), Roy (2007), Mitchell et al. (2006), Srivastava et al. (2008), Werner et al. (2008)

Testing separable covariance structure with one additionally structured component:

- Roy and Khattree (2003, 2005, 2007), Roy and Leiva (2008)

BCS structure was discussed by Rao (1945, 1953) for discriminating genetically different groups. Several hypotheses about the BCS covariance structure were discussed by

- Szatrowski (1982) and Roy and Leiva (2011)

Likelihood ratio test (LRT)

The tests discussed in the literature used LRT statistic for testing various permutations of patterns of separable covariance structures.

Based on comparison of the maximum likelihoods under the null and alternative hypotheses

$$\Lambda = \frac{\max\limits_{H_0} L}{\max\limits_{H_A} L}.$$

The exact distribution of Λ usually unknown - statistics used in practice

$$\mathrm{LR} = -2\ln(\Lambda)$$

Likelihood ratio test (LRT)

Under null hypothesis statistics LR known to have asymptotically χ^2_ν distribution

- ν is the difference between the number of unknown parameters under alternative and null hypotheses
- commonly used in practice, since the exact null distribution of LRT statistics is usually unknown.

Works under large sample asymtotics: dimension (qp) is fixed, $n \to \infty$

Many authors considered modification of LRT statistic to match the χ^2 distribution in small sample case, e.g., Mitchell et al. (2006) in context of separability test.

There are results for exact distribution for LRT statistic, e.g. Coelho and Roy (2017)

- complex numerical procedures - may not be achievable to the researchers.

Likelihood ratio test (LRT)

In case of testing separability and BCS:

$$\mathrm{LR} = -n \ln |\widehat{\boldsymbol{\Omega}}_0^{-1} \widehat{\boldsymbol{\Omega}}| = -n \ln |\widehat{\boldsymbol{\Omega}}_0^{-1}| - n \ln |\widehat{\boldsymbol{\Omega}}|$$

- LRT cannot be performed, when n is smaller than the dimension qp - common in many real data analyses.

Rao score test (RST)

Filipiak et al. (2016, 2017) and Roy et al. (2018) - Rao score test was proposed and studied

RST is an alternative or competitor to LRT - proposed by Rao (1948), also based on the likelihood principles.

RST statistic RS can be expressed in terms of the trace of $\widehat{\mathbf{\Omega}}_0^{-1}\widehat{\mathbf{\Omega}}$

$$\mathrm{RS} = \frac{n}{2}\,\mathrm{Tr}\left[\widehat{\mathbf{\Omega}}_0^{-1}\widehat{\mathbf{\Omega}} - \mathbf{I}_{qp}\right]^2$$

Exact null distribution of the test statistic is unknown, RS the test statistic has asymtotically χ^2 distribution.

Test comparison

LRT requires $\widehat{\boldsymbol{\Omega}}$ to be regular, in contrary to RST

- RST can be performed in the so-called high-dimensional case - $n < qp$.

Minimum sample size

	LRT	**RST**
$\boldsymbol{\Psi}_{\text{UN}} \otimes \boldsymbol{\Omega}$	qp+1	$\max\{q, p\} + 1$
$\boldsymbol{\Psi}_{\text{UN}} \otimes \boldsymbol{\Omega}$		q+1
$\boldsymbol{\Psi}_{\text{UN}} \otimes \boldsymbol{\Omega}$		
BCS		

Test comparison

Simulation studies - quicker convergence to the limiting χ^2 distribution of RST

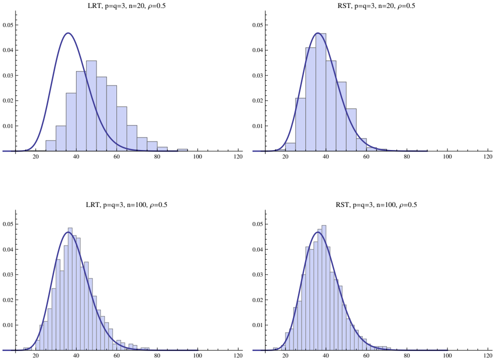


Figure: Plots of the empirical histogram and its limiting χ^2 distribution for LRT and RST statistics for sample sizes 20 and 100 for $p = 3$.

Test comparison

Simulation studies - quicker convergence to the limiting χ^2 distribution of RST

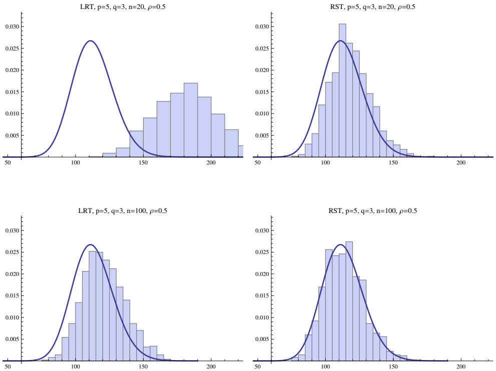
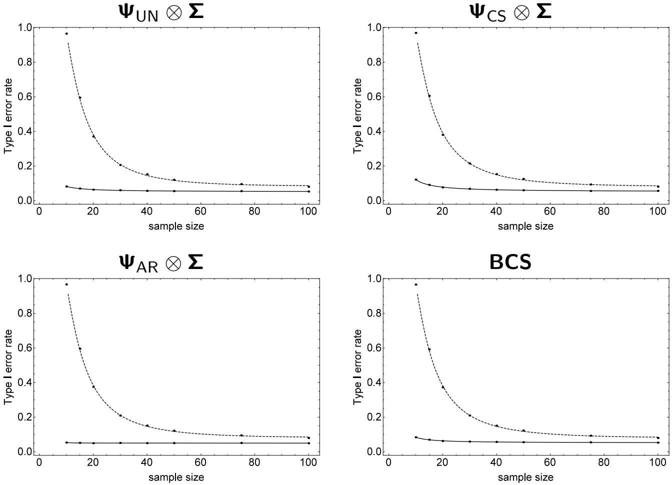


Figure: Plots of the empirical histogram and its limiting χ^2 distribution for LRT and RST statistics for sample sizes 20 and 100 for $p = 5$.

Test comparison

Speed of convergence of Type I error when the limiting χ^2 distribution is used



Small sample size

Both tests, LRT as well as RST, works under large sample asymtotics:

$$\text{dimension (say } p\text{) is fixed, } n \to \infty$$

- relatively big sample size comparing to p is needed

It has been observed that several well-known methods in multivariate analysis become inefficient or even misleading when the data dimension p is close to sample size n

- Dempster (1958) - proposed *non-exact test* for testing the mean vectors equality in two normal populations for high-dimensional case
- Bai and Saranadasa (1996) - proved analytically that Dempster's test is more powerful than the well-defined Hotelling's test for p close to n.

Small sample size

To deal with such large-dimensional data - area in asymptotic statistics has been developed where the data dimension p is no longer fixed but tends to infinity together with the sample size n

- large-dimensional asymptotics: $n, p \to \infty$, $\frac{p}{n} \to c$, $c \in (0, 1)$
- high-dimensional asymptotics: $n, p \to \infty$, $\frac{p}{n} \to c$, $c > 1$

Yao J., Zheng S., Bai Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge University Press.

The sphericity test

Assume $\mathbf{Y} \sim N_{n,p}(\mathbf{1}_n\boldsymbol{\mu}, \boldsymbol{\Omega})$. The hypothesis:

$$H_0 : \boldsymbol{\Omega} = \sigma^2 \mathbf{I}_p \quad \text{vs.} \quad H_1 : \boldsymbol{\Omega} \quad \text{unstructured}$$

LRT statistic

$$\Lambda = \left(\frac{p^p |\mathbf{S}|}{(\mathrm{Tr}[\mathbf{S}])^p} \right)^{n/2} = \left(\frac{(\ell_1 \cdots \ell_p)^{1/p}}{\frac{1}{p}(\ell_1 + \cdots + \ell_p)} \right)^{np/2},$$

where ℓ_i are the eigenvalues of $\mathbf{S}$.

Under large-sample asymtotics:

$$-2 \ln \Lambda \xrightarrow{\mathcal{D}} \chi^2_\nu, \quad \nu = p(p+1)/2$$

Box-Bartlett correction (BBLRT) - expansion of the distribution function of $-2\ln\Lambda$

The sphericity test

John (1971) - so called *John's* test

$$T_2 = \frac{n}{2}\,\mathrm{Tr}\left[\frac{p}{\mathrm{Tr}[\mathbf{S}]}\mathbf{S} - \mathbf{I}_p\right]^2 = \frac{np}{2}\cdot\frac{\frac{1}{p}\sum(\ell_i - \bar{\ell})^2}{\bar{\ell}^2}, \quad \bar{\ell} = \frac{1}{p}\sum \ell_i$$

Under large-sample asymtotics:

$$T_2 \xrightarrow{\mathcal{D}} \chi^2_\nu$$

with $\nu = p(p+1)/2$

Nagao (1973) - expansion of the distribution function of T_2

The sphericity test

Performance of BBLRT and Nagao's test with growing dimension p:

- $n = 64$, various p, $\alpha = 0.05$

p	4	8	16	32	48	56	60
BBLRT	0.0483	0.0523	0.0491	0.0554	0.1262	0.3989	0.7605
Nagao's test	0.0485	0.0495	0.0478	0.0518	0.0518	0.0513	0.0495

When $\frac{p}{n} < \frac{1}{2}$ - empirical significance level close to the nominal one for both tests

With growing ratio $\frac{p}{n}$ - BBLRT becomes quickly biased, while Nagao's test keeps correct empirical significance level

The sphericity test

LRT large-dimensional distribution - corrected LRT (CLRT): let $\mathrm{LR} = -(2/n)\ln\Lambda$ and let $\frac{p}{n-1} = c_n \to c \in (0,1)$, then

$$\mathrm{LR} + (p-n-1)\ln(1-c_n) - p \xrightarrow{\mathcal{D}} N\left(-\frac{1}{2}\ln(1-c), -2\ln(1-c) - 2c\right)$$

Limiting distribution of the test crucially depends on the limiting dimension-to-sample size ratio c through the factor -log(1 - c)

- when c approaches 1 - the asymptotic variance blow up quickly, so the power will seriously break down

The sphericity test

Ledoit and Wolf (2002)

John's test large-dimensional distribution - corrected John's test (CJT): let $\frac{p}{n-1} = c_n \to c > 0$, then

$$\frac{2}{nc_n} T_2 - p \xrightarrow{\mathcal{D}} N(1, 4)$$

Remarks to CJT

- valid also for high-dimensional case ($p > n$) - in contrast to CLRT, where this ratio should be kept smaller than 1 to avoid null eigenvalues
- John's test is in fact Rao score test - CJT is corrected RST (CRST)

Test of BCS structure

The advantage of testing the BCS structure: the likelihood estimators of unknown parameter are given in explicit form

$$\widehat{\mathbf{\Delta}}_1 = \tfrac{1}{n}\,\mathrm{BTr}_q[(\mathbf{P}_p \otimes \mathbf{I}_q)\mathbf{S}] \sim W_q\left(\tfrac{1}{n}\,\mathbf{\Delta}_2, n-1\right),$$

$$\widehat{\mathbf{\Delta}}_2 = \tfrac{1}{n(p-1)}\,\mathrm{BTr}_q[(\mathbf{Q}_p \otimes \mathbf{I}_q)\mathbf{S}] \sim W_q\left(\tfrac{1}{n(p-1)}\,\mathbf{\Delta}_2, (n-1)(p-1)\right).$$

Large-dimensional distribution of LRT and RST?

Test of BCS structure

LRT distribution of the test statistic $W = -n \ln \Lambda$

- Mitsui et al. (2015) - asymptotic normality under $\frac{qp}{n-1} \to c \in (0, 1)$
- Sun and Xie (2020) - allowed $c = 1$ and also under mild asymptotic restrictions
- Coelho and Roy (2017) - derived a near-exact approximation of the distribution

Large-dimensional distribution of RST statistic RS

$$RS = \frac{n}{2} \operatorname{Tr} \left[\left(\mathbf{P}_p \otimes \widehat{\boldsymbol{\Delta}}_1^{-1} + \mathbf{Q}_p \otimes \widehat{\boldsymbol{\Delta}}_2^{-1} \right) \mathbf{S} - \mathbf{I}_{qp} \right]^2$$

only for $q = 1$

Test of CS structure

Theorem

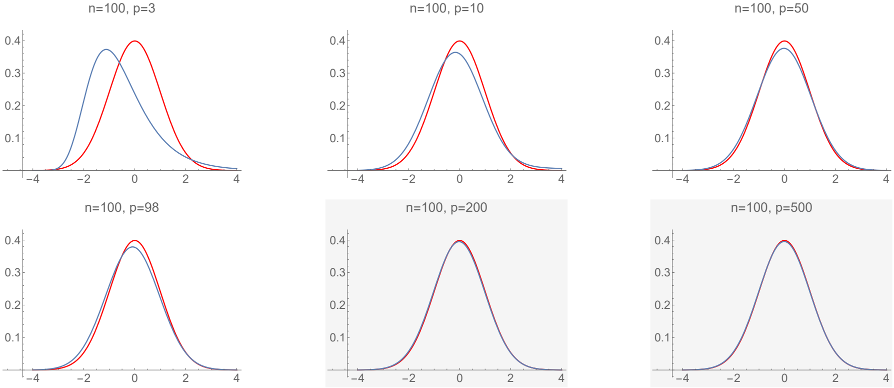
Under model $\mathbf{Y} \sim N_{n,p}(\mathbf{1}_n\boldsymbol{\mu}, \mathbf{I}_n, \boldsymbol{\Omega})$ *and hypothesis* $H_0 : \boldsymbol{\Omega} = \boldsymbol{\Omega}_{CS}$*, and when* $c_n = \frac{p-1}{n-1} \to c > 0$*, the following holds:*

$$\frac{1}{nc_n} RS - \frac{p+2}{2} \quad \xrightarrow{\mathcal{D}} \quad N\left(0,1\right),$$

where RS *is the Rao score test statistic.*

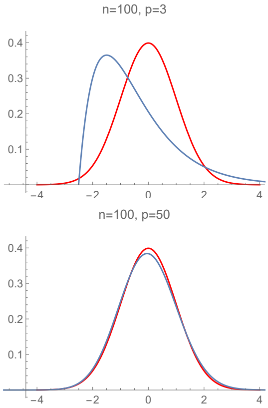
Test of CS structure

RST large- and high-dimensional distribution - simulated empirical null distribution of RS test statistic (blue) together with approximation by standard normal distribution (red)



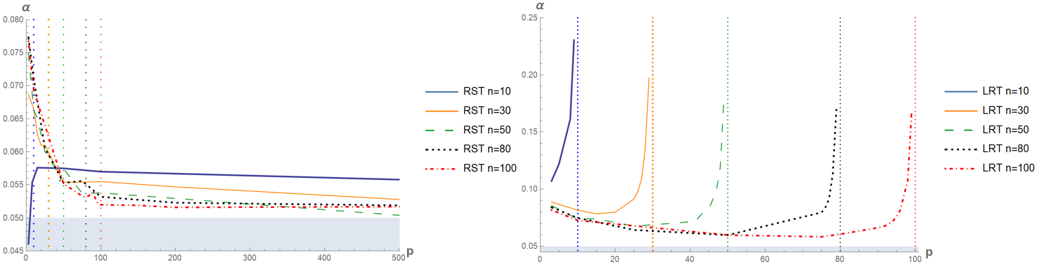
Test of CS structure

LRT large- and high-dimensional distribution - transformed exact null distribution of LR test statistic (blue) together with approximation by standard normal distribution (red)



Test of CS structure

Simulations - empirical Type I error, $\widehat{\alpha}$



Type I error as a function of p based on simulations for $n = 10, 30, 50, 80, 100$

Thank you for your attention!