Streamlining the Applied Mathematics Studies at Faculty of Science of Palacký University in Olomouc
CZ.1.07/2.2.00/15.0243

International Conference
Olomoucian Days of Applied Mathematics

ODAM 2011

Department of Mathematical analysis
and Applications of Mathematics
Faculty of Science
Palacký University Olomouc

european
social fund in the
czech republic

EUROPEAN UNION

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

OP Education
for Competitiveness

INVESTMENTS
IN EDUCATION
DEVELOPMENT

CONTINUATION OF PROFESSOR KUBÁČEK'S RESEARCH WORK: SOME MISCELLANEOUS EXAMPLES

Gejza Wimmer, Viktor Wítkovský, Gejza Wimmer jr.

Introduction

- Longitudinal Data
- Linear Mixed Model (LMM)
- Parameter Estimation
  - regression parameters
  - covariance parameters
- Problem of the construction of confidence regions for regression parameters
  - in small sample case
  - if it is necessary to estimate also the covariance parameters of the model
  - simulation results

Longitudinal data

- the defining feature is that measurements of the same subjects are taken repeatedly through time
  - obtain vectors $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$ of repeated measurements on each subject ($i = 1, \ldots, N$)
  - vectors of outcomes are independent between subjects
  - repeated measurements done on the single subject exhibit some form of (positive) correlation
- the primary goal of a longitudinal study is to characterize the change on response over time and the factors that influence change
  - to describe a common feature which defines the behavior of all subjects in time - fixed effects
  - in addition, every subject has some individual effects on his repeated measurements - individual effects

Linear model model for longitudinal data

- for longitudinal data proposed by Laird a Ware (1982)
- useful model for such type of data, because it reflects both -common and individual- effects of each subject on his repeated measurements
- the response vector $\mathbf{Y}_i$ for i-th subject, $(i = 1, \ldots, I)$, can be written as

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i$$

  - $\mathbf{X}_i$ are $(n_i \times p)$-dimensional known matrices
  - $\boldsymbol{\beta}$ is p-dimensional vector of the unknown regression parameters
  - - the same for all subjects
  - - fixed effects
  - $\mathbf{Z}_i$ are $(n_i \times r)$-dimensional known matrices
  - $\boldsymbol{\eta}_i$ are r-dimensional vectors of unknown parameters
  - - mutually independent
  - - random vectors from $N(\mathbf{0}, \mathbf{D})$
  - - random effects of the individuals on their repeated measurements
  - $\boldsymbol{\varepsilon}_i$ is n_i-dimensional vector of errors for i-th subject
  - - independent from $\boldsymbol{\eta}_i$
  - - $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i)$

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_i \\ \vdots \\ \mathbf{Y}_I \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_i \\ \vdots \\ \mathbf{X}_I \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{Z}_1 & 0 & \dots & 0 & \dots & 0 \\ 0 & \mathbf{Z}_2 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{Z}_i & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & \mathbf{Z}_I \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_i \\ \vdots \\ \boldsymbol{\eta}_I \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_I \end{bmatrix}$$

$\mathbf{Y} = (\mathbf{Y}_1', \mathbf{Y}_2', \dots, \mathbf{Y}_I')'$ and $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2', \dots, \mathbf{X}_I')'$
covariance matrix of the random response vector **Y** is

$$Var(\mathbf{Y}) = \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Sigma}_i & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \boldsymbol{\Sigma}_I \end{bmatrix},$$

where each $\boldsymbol{\Sigma}_i$ is the covariance matrix for the response vector of the i-th subject $\mathbf{Y}_i$

$$Var(\mathbf{Y}_i) = \boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \mathbf{R}_i$$

we assume that the covariance matrix of the random response vector $\mathbf{Y}$ is some known function of r-dimensional vector of parameters $\boldsymbol{\theta}$, $Var(\mathbf{Y}) = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$

- vector of parameters $\boldsymbol{\theta}$ is known, hence the covariance matrix $\boldsymbol{\Sigma}$ is also known and providing that there exist the inverse matrix to it, the estimator of fixed effects is

$$\widetilde{\boldsymbol{\beta}} = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}$$

  - we assume the existence of the matrix $(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} = \boldsymbol{\Phi}$

- such an estimator of fixed effects corresponds to the solution of the generalized least square method of the estimation of the unknown regression parameters and with the given assumptions it is known that

$$\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \sim N\left(\mathbf{0}, \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\right)$$

  - best linear unbiased estimator (*BLUE*) of the vector of fixed effects

Estimation of the unknown covariance parameters

- the primary interest in the longitudinal data analysis is to estimate the unknown fixed effects $\boldsymbol{\beta}$
- it is often necessary to estimate the unknown covariance parameters of the model too, $\boldsymbol{\theta}$ - nuisance
  - it is largely used the REML likelihood function, where the logarithm of REML likelihood function is in form

$$l_{REML}(\boldsymbol{\theta}; \mathbf{Y}) = -\frac{1}{2}(n-r)\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \frac{1}{2}\ln\left|\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}\right|$$
$$-\frac{1}{2}\mathbf{Y}'\left\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}\left[\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\mathbf{X}\right]^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right\}\mathbf{Y},$$

    where $n = \sum_{i=1}^{I} n_i$
- for the estimator of covariance parameters $\widehat{\boldsymbol{\theta}}$
  - asymptotically distributed (for $n \to \infty$)

$$\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \sim N(\mathbf{0}, \mathbf{W}),$$

    where $\mathbf{W}$ is the inverse of Fisher information matrix

$$\mathbf{W} = \left\{E\left[\frac{\partial l_{REML}}{\partial \boldsymbol{\theta}}\frac{\partial l_{REML}}{\partial \boldsymbol{\theta}'}\right]\right\}^{-1} = \left\{-E\left[\frac{\partial^2 l_{REML}}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right]\right\}^{-1}$$

    and $\boldsymbol{\theta}_0$ is the true vector of covariance parameters of the model

Some properties of the estimator of fixed effects

- unknown are not only the fixed effects but also the covariance parameters of the model $\boldsymbol{\theta}$
  - $\boldsymbol{\Sigma}$ unknown
- we have the REML estimator of the covariance matrix $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$
- a simply approach to get an estimator of the unknown vector of fixed effects$\boldsymbol{\beta}$

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^{'}\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{X}\right)^{-1}\mathbf{X}^{'}\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{Y}$$

  - empirical best linear unbiased estimator - *EBLUE*
- is an unbiased estimator of fixed effects of the model $\boldsymbol{\beta}$

$$E\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$$

- the covariance matrix of $\widehat{\boldsymbol{\beta}}$ is in general not known
- often is estimated as

$$\widehat{\boldsymbol{\Phi}} = \left(\mathbf{X}^{'}\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\theta}})\mathbf{X}\right)^{-1}$$

- for "small" sample sizes is such an approximation of $Var\left(\widehat{\beta}\right)$ using $\widehat{\mathbf{\Phi}}$ inappropriate
  - do not take into account the uncertainty included in the estimating of unknown covariance parameters of the model $\boldsymbol{\theta}$
  - $\widehat{\mathbf{\Phi}}$ is not an unbiased estimator of $\mathbf{\Phi}$
- Kackar, Harville (1984) showed that the approximate covariance matrix of the *EBLUE* of the fixed effects β is

$$Var\left(\widehat{\beta}\right) = \mathbf{\Phi} + \mathbf{\Lambda}_1,$$

where

$$\mathbf{\Lambda}_1 \approx \mathbf{\Phi}\left\{\sum_{k=1}^{r}\sum_{l=1}^{r}\{\mathbf{W}\}_{kl}\left(\mathbf{Q}_{kl} - \mathbf{P}_k\mathbf{\Phi}\mathbf{P}_l\right)\right\}\mathbf{\Phi},$$

$$\mathbf{P}_k = -\mathbf{X}'\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\theta_k}\mathbf{\Sigma}^{-1}\mathbf{X},$$

$$\mathbf{Q}_{kl} = \mathbf{X}'\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\theta_k}\mathbf{\Sigma}^{-1}\frac{\partial\mathbf{\Sigma}}{\partial\theta_l}\mathbf{\Sigma}^{-1}\mathbf{X}$$

- is function of the unknown covariance parameters of the model
  - it is necessary to estimate it

$$\widehat{Var}\left(\widehat{\beta}\right) = \widehat{\mathbf{\Phi}} + \widehat{\mathbf{\Lambda}}_1$$

- Kenward, Roger (1997), based on Harville, Jeske (1992) and Prasad, Rao (1990), derived an approximation of the matrix $\widehat{Var}\left(\widehat{\boldsymbol{\beta}}\right)$ such that this estimator includes the bias in $\widehat{\boldsymbol{\Phi}}$

$$\widehat{Var}\left(\widehat{\boldsymbol{\beta}}\right) =$$

$$\widehat{\boldsymbol{\Phi}} + 2\widehat{\boldsymbol{\Phi}}\left\{\sum_{k=1}^{r}\sum_{l=1}^{r}\left\{\widehat{\mathbf{W}}\right\}_{kl}\left(\widehat{\mathbf{Q}}_{kl} - \widehat{\mathbf{P}}_{k}\widehat{\boldsymbol{\Phi}}\widehat{\mathbf{P}}_{l} - \frac{1}{4}\widehat{\mathbf{R}}_{kl}\right)\right\}\widehat{\boldsymbol{\Phi}} \equiv \widehat{\boldsymbol{\Phi}}_{mod},$$

  where

$$\mathbf{R}_{kl} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\left[\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_k\partial\theta_l}\right]\boldsymbol{\Sigma}^{-1}\mathbf{X}$$

- this method assumes unbiasedness of the estimator of covariance parameters of the model $\boldsymbol{\theta}$
- REML estimator of covariance parameters is in general not unbiased
  - the bias is more significant in the cases which assume nonlinear structures of the covariance matrix of the response vector **Y**

- Kenward, Roger (2009) derived the bias of the REML estimator of $\widehat{\boldsymbol{\theta}}$

$$Bias(\widehat{\theta}_k) = E\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)_k\right] =$$

$$-\frac{1}{4}\sum_{l=1}^{r}\sum_{i=1}^{r}\sum_{j=1}^{r}\{\mathbf{W}\}_{ij}\{\mathbf{W}\}_{kl}\, trace\left[\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_k\partial\theta_l}\mathbf{C}_k\right],$$

where

$$\mathbf{C}_k = \left(\left(\mathbf{I} - \boldsymbol{\Sigma}^{-1}\left(\mathbf{X}\boldsymbol{\Phi}\mathbf{X}'\right)\right)\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\left(\mathbf{I} - \boldsymbol{\Sigma}^{-1}\left(\mathbf{X}\boldsymbol{\Phi}\mathbf{X}'\right)\right)'\right)$$

- this relationship was included to the method of achieving the estimator of the covariance matrix of the estimator of fixed effects from the previous article, what leads to modified estimator of the covariance matrix of the estimator of fixed effects

$$\widehat{Var}\left(\widehat{\beta}\right) = \widehat{\boldsymbol{\Phi}}_{mod} + \widehat{\mathbf{B}} \equiv \widehat{\boldsymbol{\Phi}}^{*}_{mod1},$$

where

$$\mathbf{B} = \frac{1}{4}\sum_{k=1}^{r}\sum_{l=1}^{r}\sum_{i=1}^{r}\sum_{j=1}^{r}\{\mathbf{W}\}_{ij}\{\mathbf{W}\}_{kl}\, trace\left[\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_k\partial\theta_l}\mathbf{C}_k\right]\left[\frac{\partial\boldsymbol{\Phi}}{\partial\theta_k}\right] =$$

$$-\sum_{k=1}^{r} Bias(\widehat{\theta}_k)\left[\frac{\partial\boldsymbol{\Phi}}{\partial\theta_k}\right]$$

Some proposal to improve the previous method

- to implement the bias into the methods for obtaining the covariance matrix of $\widehat{\boldsymbol{\theta}}$ given in Kackar, Harville (1984)
  - by setting the term $\boldsymbol{\Lambda}_1$
  - including this bias into the estimator of the covariance matrix $\widehat{\boldsymbol{\beta}}$ not only through the correction term $\widehat{\mathbf{B}}$ of the matrix $\widehat{\boldsymbol{\Phi}}^*_{mod1}$ in comparison to matrix $\widehat{\boldsymbol{\Phi}}_{mod}$, but also in the term
  $$2\widehat{\boldsymbol{\Phi}}\left\{\sum_{k=1}^r\sum_{l=1}^r\left\{\widehat{\mathbf{W}}\right\}_{kl}\left(\widehat{\mathbf{Q}}_{kl}-\widehat{\mathbf{P}}_k\widehat{\boldsymbol{\Phi}}\widehat{\mathbf{P}}_l-\tfrac{1}{4}\widehat{\mathbf{R}}_{kl}\right)\right\}\widehat{\boldsymbol{\Phi}}$$
- omitting the assumption of the unbiasedness of the REML estimator of the covariance parameters $\boldsymbol{\theta}$
  $$\widehat{Var}(\widehat{\boldsymbol{\beta}}) = \widehat{\boldsymbol{\Phi}}+\widehat{\mathbf{B}}+2\widehat{\boldsymbol{\Phi}}\left\{\sum_{k=1}^r\sum_{l=1}^r\widehat{w}_{kl}\left(\widehat{\mathbf{Q}}_{kl}-\widehat{\mathbf{P}}_k\widehat{\boldsymbol{\Phi}}\widehat{\mathbf{P}}_l-\tfrac{1}{4}\widehat{\mathbf{R}}_{kl}\right)\right\}\widehat{\boldsymbol{\Phi}}\equiv\widehat{\boldsymbol{\Phi}}_{mod1},$$
  where
  $$w_{kl}=E\left[\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)_k\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)_l\right]=\{\mathbf{W}\}_{kl}+Bias(\widehat{\theta}_k)Bias(\widehat{\theta}_l)$$

Confidence regions for the fixed effects

- the statistical inference about l linear combinations of the elements of $\boldsymbol{\beta}$
  - construction of the confidence regions for some linear combinations $\mathbf{L}'\boldsymbol{\beta}$, where $\mathbf{L}$ is some known $(p \times l)$-dimensional matrix

Covariance parameters of the model $\boldsymbol{\theta}$ are known

- the distribution of the $\widetilde{\boldsymbol{\beta}}$ is known and

$$\chi^2 = (\mathbf{L}'\widetilde{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})'(\mathbf{L}'\boldsymbol{\Phi}\mathbf{L})^{-1}(\mathbf{L}'\widetilde{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})$$

- asymptotically exact confidence regions for $\mathbf{L}\boldsymbol{\beta}$

Covariance parameters $\boldsymbol{\theta}$ are unknown

- replace $\boldsymbol{\theta}$ with its REML estimator $\widehat{\boldsymbol{\theta}}$
  - $\boldsymbol{\Phi} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$ is substituted by $\widehat{\boldsymbol{\Phi}} = (\mathbf{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}$
- to trust that $\chi^2_* = (\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})'(\mathbf{L}'\widehat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})$ has χ^2 distribution with l degrees of freedom too
  - is used in the major part of the basic literature about longitudinal data
  - appropriate for "large" sample sizes

- in the cases of sufficient sample sizes is previous approach appropriate

$\boldsymbol{\theta}$ is unknown; "small" sample sizes

- uncertainty due the estimating of the covariance parameters of the model $\boldsymbol{\theta}$ have to be involved into the inferences about the elements of $\boldsymbol{\beta}$
  - using appropriate estimator of the covariance matrix of $\widehat{\boldsymbol{\beta}}$
  - using F distribution instead of standard χ^2 distribution
- distribution of

$$F = \frac{1}{l}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})'(\mathbf{L}'\widehat{\boldsymbol{\Phi}}^*\mathbf{L})^{-1}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})$$

  - $\widehat{\boldsymbol{\Phi}}^*$ is either the "naive" estimator of the covariance matrix of the vector of fixed effects of the model $\widehat{\boldsymbol{\Phi}}$, or some of its modification, e.g. $\widehat{\boldsymbol{\Phi}}_{mod}$
  - for $l = 1$ - Satterthwaite (1941)
  - for $l > 1$ - more complex
    - take into account the random structure of $\mathbf{L}\widehat{\boldsymbol{\Phi}}_{mod}\mathbf{L}'$
    - shows to be advantageous to assume $F_{l,m}$ distribution - number of degrees of freedom m

Fai-Cornelius method

- method to determine the degrees of freedom m
- assumption
$$F_{FC} = \frac{1}{l}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})'(\mathbf{L}'\widehat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})$$
has approximately $F_{l,m}$ distribution
  - Fai-Cornelius method use as an estimator of the covariance matrix of the *EBLUE* of the fixed effects the matrix $\widehat{\boldsymbol{\Phi}}$
- using the spectral decomposition of the matrix $(\mathbf{L}'\widehat{\boldsymbol{\Phi}}\mathbf{L})^{-1}$
- using Satterthwaite approximation

Fai-Cornelius approximative $(1-\alpha)\cdot 100\%$ confidence region ($\alpha \in (0,1)$) is such a set of vectors of fixed effects $\boldsymbol{\beta}^{+}$, for which

$$\frac{1}{l}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}^{+})'(\mathbf{L}'\widehat{\boldsymbol{\Phi}}\mathbf{L})^{-1}(\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}^{+}) < F_{l,\widehat{m}}(1-\alpha),$$

where $F_{l,\widehat{m}}(1-\alpha)$ is the $(1-\alpha)$-quantile of the Fisher-Snedecor distribution with l and $\widehat{m}$ degrees of freedom

Kenward-Roger method

- method to estimate the number of degrees of freedom m
- based on the assumption that
$$F_{KR} = \lambda \frac{1}{l} (\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})' (\mathbf{L}'\widehat{\boldsymbol{\Phi}}_{mod}\mathbf{L})^{-1} (\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})$$
has approximately $F_{l,m}$ distribution
  - use as an estimator of the covariance matrix of the *EBLUE* of the fixed effects the matrix $\widehat{\boldsymbol{\Phi}}_{mod}$ from Kenward, Roger (1997)
  - in addition there is some "scale" factor λ - used to be estimated
- due the second-order Taylor series expansion of the matrix $(\mathbf{L}'\widehat{\boldsymbol{\Phi}}_{mod}\mathbf{L})^{-1}$ around the true vector of covariance parameters of the model $\boldsymbol{\theta}_0$
- using this expansion in the calculation of the first and second moment of the random variable $F = \frac{1}{\lambda} F_{KR}$

Kenward-Roger approximative $(1-\alpha) \cdot 100\%$ confidence region $(\alpha \in (0,1))$ is such a set of vectors of fixed effects $\boldsymbol{\beta}^+$, for which

$$\widehat{\lambda} \frac{1}{l} (\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}^+)' (\mathbf{L}'\widehat{\boldsymbol{\Phi}}_{mod}\mathbf{L})^{-1} (\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}^+) < F_{l,\widehat{m}}(1-\alpha),$$

where $F_{l,\widehat{m}}(1-\alpha)$ is the $(1-\alpha)$-quantile of the Fisher-Snedecor distribution with l and $\widehat{m}$ degrees of freedom

Modified Kenward-Roger method

- let us suppose that
$$F^* = \lambda \frac{1}{l} (\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})' (\mathbf{L}'\widehat{\boldsymbol{\Phi}}_{mod1}\mathbf{L})^{-1} (\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta})$$
has approximately Fisher-Snedecor distribution with l and m degrees of freedom
    - estimator of the covariance matrix of the *EBLUE* of the fixed effects the matrix is $\widehat{\boldsymbol{\Phi}}_{mod1}$
- based on the Alnosaier (2007) and Kenward, Roger (1997) where derived relationships to an approximately calculation of the "scale" factor λ and degrees of freedom m

Modified Kenward-Roger approximative $(1-\alpha)\cdot 100\%$ confidence region ($\alpha \in (0,1)$) is such a set of vectors of fixed effects $\boldsymbol{\beta}^+$, for which

$$\widehat{\lambda} \frac{1}{l} (\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}^+)' (\mathbf{L}'\widehat{\boldsymbol{\Phi}}_{mod1}\mathbf{L})^{-1} (\mathbf{L}'\widehat{\boldsymbol{\beta}} - \mathbf{L}'\boldsymbol{\beta}^+) < F_{l,\widehat{m}}(1-\alpha),$$

where $F_{l,\widehat{m}}(1-\alpha)$ is the $(1-\alpha)$-quantile of the Fisher-Snedecor distribution with l and $\widehat{m}$ degrees of freedom

Small simulation study

- to compare all mentioned approaches for different numbers of subjects with 8 repeated measurements on each subject in the linear mixed model with with AR(1) errors
  - "small" number of subjects
- it were computed the empirical probabilities of coverage of the real values $\mathbf{L}^{'}\beta$ from 10000 simulations of the 95% confidence regions taken by all mentioned methods

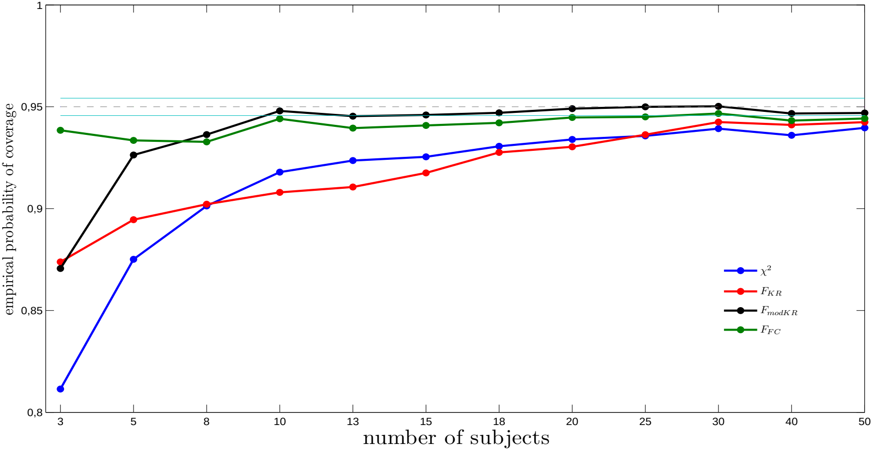
based on the example given in the Zerbe (1979) we considered the LMM model with

- 3-dimensional vector of fixed effects $\boldsymbol{\beta}_0 = [4.5, -0.7, 0.3]'$
- 1-dimensional vector of individual effects $\eta_i \sim N(0, \sigma_\eta^2)$, $i = 1, 2, \ldots, I$
- AR(1) errors

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \eta_i + \boldsymbol{\varepsilon}_i$$

$$\mathbf{X}_i = \begin{bmatrix} 1 & 0.0 & 0 \\ 1 & 0.5 & 0 \\ 1 & 1.0 & 0 \\ 1 & 1.5 & 0 \\ 1 & 2.0 & 0 \\ 1 & 2.0 & 1 \\ 1 & 2.0 & 2 \\ 1 & 2.0 & 3 \end{bmatrix}$$

- $\mathbf{Z}_i = [1, 1, 1, 1, 1, 1, 1, 1]'$, $i = 1, 2, \ldots, I$
- covariance parameters of the model were sets as $\sigma_0^2 = 0.16$, $\sigma_{\eta\ 0}^2 = 0.25$ and AR(1) parameter $\rho_0 = 0.7$



Empirical probabilities of coverage of $\mathbf{L}'\beta_0$ for different number of subjects with the same range of the repeated measurements, 8, on each subject for the nominal confidence level 0.95.

Conclusion

It turns out that using of the modified Kenward-Roger method improves the results obtained by the Kenward-Roger method given in Kenward, Roger (1997) and with the increasing number of subject the proposed modified Kenward-Roger confidence region is approaching the theoretical value faster than the Kenward-Roger method. From the figure it should be also deduced that the modified Kenward-Roger method gives almost the same results as the Fai-Cornelius method.

Thanks