# Model Selection and Multimodel Inference in Reliability Estimation

Ariel Alonso Abad[1] and Annouschka Laenen[2]

ariel.alonso@maastrichtuniversity.nl

[1] Maastricht University. The Netherlands

[2] Katholieke Universiteit Leuven. Belgium

# Estimating Reliability in a Longitudinal Framework

- *Advantages*
  - $\Rightarrow$ Study of evolutions over time
  - $\Rightarrow$ More information about the patients $\Rightarrow$ more reliable results?
- *Disadvantages*
  - $\Rightarrow$ Different sources of variation and correlation
  - $\Rightarrow$ Systematic changes over time
  - $\Rightarrow$ Individualized evolutions
  - $\Rightarrow$ Serial correlation and heterogenous variance components

## Modeling Framework

Laenen *et al* (2007, 2009)

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i + \boldsymbol{\tau}_i + \boldsymbol{\varepsilon}_i,$$

where

$\Rightarrow$ $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{ip})^T$ and $Y_{ij}$ outcome of subject $i$ at time $j$

$\Rightarrow$ $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{ip})^T$ is a general mean vector

$\Rightarrow$ $\boldsymbol{\tau}_i = (\tau_{i2}, \tau_{i2}, \ldots, \tau_{ip})^T$ is the vector of true scores

$\Rightarrow$ $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{ip})^T$ measurement error component

Problem: Number of parameters increases with the number of time points and subjects

# Modeling Framework: A Second Level

Laenen *et al* (2007, 2009)

$\Rightarrow$ $\boldsymbol{\mu}_i = \boldsymbol{X}_i\boldsymbol{\beta}$ where $\boldsymbol{X}_i$ is a matrix and $\boldsymbol{\beta}$ a vector of fixed effects

$\Rightarrow$ $\boldsymbol{\tau}_i = \boldsymbol{Z}\boldsymbol{b}_i$ where $\boldsymbol{Z}$ is a matrix and $\boldsymbol{b}_i$ are subject-specific effects

$\Rightarrow$ $\tau_{ij} = b_{i0} + b_{i1} t_j \log t_j$:

- $\boldsymbol{z}_j = (1 \quad t_j \log t_j)$ is the $j^{th}$ row of the $\boldsymbol{Z}$
- $\boldsymbol{b}_i = (b_{i0}, b_{i1})^T$

$\Rightarrow$ $\boldsymbol{\varepsilon}_i = \boldsymbol{\varepsilon}_{(1)i} + \boldsymbol{\varepsilon}_{(2)i}$

$\Rightarrow$ $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D})$ , $\boldsymbol{\varepsilon}_{(1)i} \sim N(\boldsymbol{0}, \tau^2 \boldsymbol{H}_\rho)$, $\boldsymbol{\varepsilon}_{(2)i} \sim N(\boldsymbol{0}, \boldsymbol{R})$

# Quantifying Reliability

$$\text{Var}(\boldsymbol{Y}_i) = \boldsymbol{V}$$

# Quantifying Reliability

$$\text{Var}(\boldsymbol{Y}_i) = \boldsymbol{V}, \text{Var}(\boldsymbol{\tau}_i) = \boldsymbol{\Sigma}_D = \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}'$$

# Quantifying Reliability

Var$(\boldsymbol{Y}_i) = \boldsymbol{V}$, Var$(\boldsymbol{\tau}_i) = \boldsymbol{\Sigma}_D = \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}'$ and Var$(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma} = \tau^2 \boldsymbol{H}_\rho + \boldsymbol{R}$

# Quantifying Reliability

$\text{Var}(\boldsymbol{Y}_i) = \boldsymbol{V}$, $\text{Var}(\boldsymbol{\tau}_i) = \boldsymbol{\Sigma}_D = \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}'$ and $\text{Var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma} = \tau^2 \boldsymbol{H}_\rho + \boldsymbol{R}$

$$\boxed{\boldsymbol{V} = \boldsymbol{\Sigma}_D + \boldsymbol{\Sigma}}$$

# Quantifying Reliability

$$\text{Var}(\boldsymbol{Y}_i) = \boldsymbol{V}, \text{ Var}(\boldsymbol{\tau}_i) = \boldsymbol{\Sigma}_D = \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}' \text{ and } \text{Var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma} = \tau^2\boldsymbol{H}_\rho + \boldsymbol{R}$$

$$\boxed{\boldsymbol{V} = \boldsymbol{\Sigma}_D + \boldsymbol{\Sigma}}$$

Average Reliability

$$R_T = 1 - \frac{\text{tr}(\boldsymbol{\Sigma})}{\text{tr}(\boldsymbol{V})}$$

# Quantifying Reliability

$\text{Var}(\boldsymbol{Y}_i) = \boldsymbol{V}$, $\text{Var}(\boldsymbol{\tau}_i) = \boldsymbol{\Sigma}_D = \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}'$ and $\text{Var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma} = \tau^2 \boldsymbol{H}_\rho + \boldsymbol{R}$

$$\boxed{\boldsymbol{V} = \boldsymbol{\Sigma}_D + \boldsymbol{\Sigma}}$$

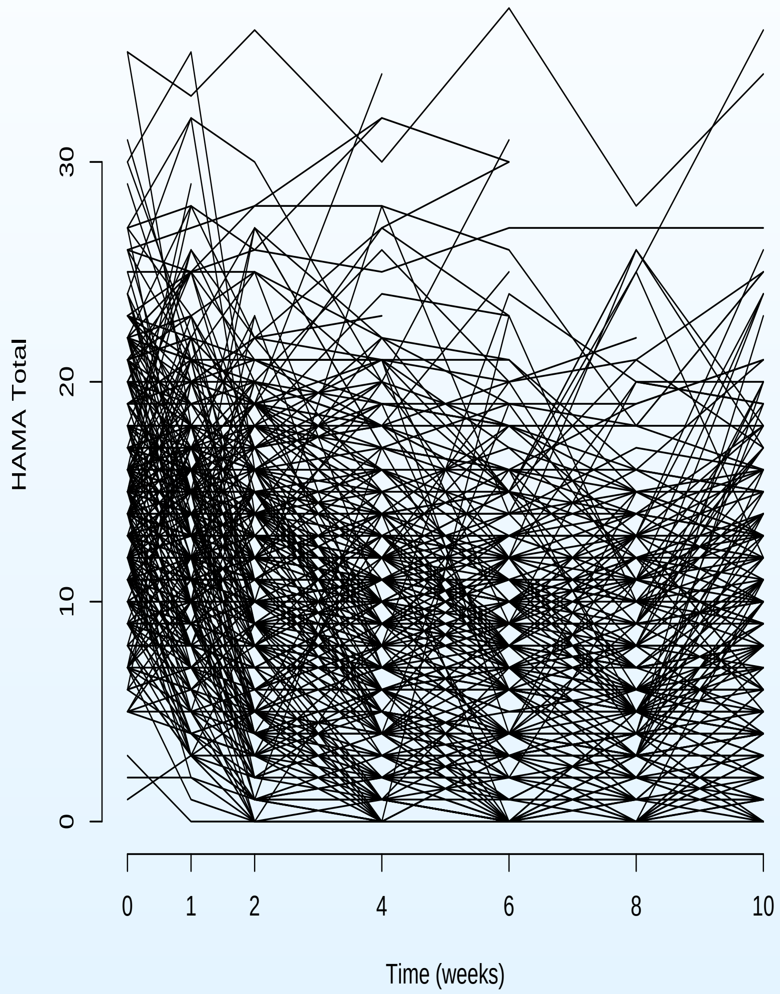
Average Reliability

$$R_T = 1 - \frac{\text{tr}(\boldsymbol{\Sigma})}{\text{tr}(\boldsymbol{V})}$$

Total Reliability

$$R_\Lambda = 1 - |\boldsymbol{\Sigma}\boldsymbol{V}^{-1}| = 1 - \frac{|\boldsymbol{\Sigma}|}{|\boldsymbol{V}|}$$

# Case Study




- Trial: 353 individuals
- Major depressive disorders:2 treatments
- Hamilton Anxiety Rating Scale (HAMA)
- Six fixed time points: $1, 2, 4, 6, 8$ and $10$ weeks

Maastricht University Leading in Learning!

# Case Study: Results

| | Random effects | Residual covariance | AIC | Akaike weights |
|---|---|---|---|---|
| 1 | Quadratic | autoregressive | 11234.1 | 0.56155 |
| 2 | Quadratic | Toeplitz 2 bands | 11235.1 | 0.34060 |
| 3 | Quadratic | heterogeneous autoregressive | 11237.9 | 0.08399 |
| 4 | Linear | autoregressive | 11242.7 | 0.00762 |
| 5 | Quadratic | simple | 11243.1 | 0.00624 |

# Case Study: Results

| | Random effects | Residual covariance | AIC | Akaike weights |
|---|---|---|---|---|
| 1 | Quadratic | autoregressive | 11234.1 | 0.56155 |
| 2 | Quadratic | Toeplitz 2 bands | 11235.1 | 0.34060 |
| 3 | Quadratic | heterogeneous autoregressive | 11237.9 | 0.08399 |
| 4 | Linear | autoregressive | 11242.7 | 0.00762 |
| 5 | Quadratic | simple | 11243.1 | 0.00624 |

$$R_T = 0.675 \quad CI_{0.95} = (0.601, 0.741)$$

$$R_\Lambda = 0.968 \quad CI_{0.95} = (0.930, 0.986)$$

# Modeling Building: A Quiet Scandal in Statistics

Data used two times

1. To construct/chose the model

2. To make the inferences

Consequences of ignoring 1

$\Rightarrow$ Distribution of post-model-selection estimators generally complex and may depend on the unknown value of the parameters.

$\Rightarrow$ Final model typically wrong

*Burnham and Anderson 2002, Claeskens and Hjort 2008*
*Leeb and Potscher 2005-2006*

# Simulation Study

## True Data Generating Mechanism

$$\begin{aligned} Y_{ij} &= \mu_{ij} + \tau_{ij} + \varepsilon_{ij1} + \varepsilon_{ij2}, \\ \mu_{ij} &= \beta_0 + \beta_1 t_j + \beta_2 t_j^2 + \beta_3 t_j^3 + \beta_4 \log(t_j), \\ \tau_{ij} &= b_{0i} + b_{1i} t_j + b_{2i} t_j^2 + b_{3i} \log(t_j), \end{aligned}$$

where $\boldsymbol{\beta} = (40, 20, -3, -0.05, -5)'$ and $\boldsymbol{b}_i \sim N(\mathbf{0}, \boldsymbol{D})$ with

$$\boldsymbol{D} = \begin{pmatrix} 15 & 8 & \text{-}5 & 0.5 \\ 8 & 15 & \text{-}2.1 & 0.2 \\ \text{-}5 & \text{-}2.1 & 6 & 0.2 \\ 0.5 & 0.2 & 0.2 & 0.8 \end{pmatrix}$$

$\boldsymbol{\varepsilon}_{i1} \sim N(\mathbf{0}, \boldsymbol{R})$ with $\boldsymbol{R} = \text{diag}(\sigma_j^2)$, $\sigma_1^2 = 15$ and $\sigma_{j+1}^2 = \sigma_j^2 + 5$

$\boldsymbol{\varepsilon}_{i2} \sim N(\mathbf{0}, \tau^2 \boldsymbol{H}_\rho)$ with $\tau^2 = 20$ and $\boldsymbol{H}_\rho$ autoregressive with $\rho = 0.2$

# Simulation Study

Five models fitted to the data

| Model | Systematic eff. | Rand. eff. | Error structure |
|---|---|---|---|
| 1 | saturated | intercept | simple |
| 2 | saturated | intercept | autoregressive |
| 3 | saturated | intercept | heterogeneous autoregressive |
| 4 | saturated | intercept, time | simple |
| 5 | saturated | intercept, time | autoregressive |

Three strategies for the analysis

$\Rightarrow$ Best model approach (AIC): naive/corrected

$\Rightarrow$ Model average approach

*(Burnham and Anderson 2002)*

# Simulation Study: Results

Probability of each model to be selected as the best model

| | 4 | | | 6 | | |
|---|---|---|---|---|---|---|
| Model | 50 | 100 | 300 | 50 | 100 | 300 |
| 1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 2 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 3 | 0.132 | 0.140 | 0.146 | 0.034 | 0.038 | 0.002 |
| 4 | 0.794 | 0.758 | 0.508 | 0.510 | 0.320 | 0.020 |
| 5 | 0.072 | 0.102 | 0.346 | 0.456 | 0.642 | 0.978 |

# Simulation Study: Results

Probability of each model to be selected as the best model

| | 4 | | | 6 | | |
|---|---|---|---|---|---|---|
| Model | 50 | 100 | 300 | 50 | 100 | 300 |
| 1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 2 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 3 | 0.132 | 0.140 | 0.146 | 0.034 | 0.038 | 0.002 |
| 4 | 0.794 | 0.758 | 0.508 | 0.510 | 0.320 | 0.020 |
| 5 | 0.072 | 0.102 | 0.346 | 0.456 | 0.642 | 0.978 |

# Simulation Study: Results

Probability of each model to be selected as the best model

| | 4 | | | 6 | | |
|---|---|---|---|---|---|---|
| Model | 50 | 100 | 300 | 50 | 100 | 300 |
| 1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 2 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 3 | 0.132 | 0.140 | 0.146 | 0.034 | 0.038 | 0.002 |
| 4 | 0.794 | 0.758 | 0.508 | 0.510 | 0.320 | 0.020 |
| 5 | 0.072 | 0.102 | 0.346 | 0.456 | 0.642 | 0.978 |

# Simulation Study: $R_T$ Point Estimates

| | $p$ | 4 | | | 6 | | |
|---|---|---|---|---|---|---|---|
| | $n$ | 50 | 100 | 300 | 50 | 100 | 300 |
| | $R_T$ | .606 | .606 | .606 | .780 | .780 | .780 |
| BM | $\overline{\widehat{R}_T}$ | .601 | .573 | .531 | .744 | .733 | .744 |
| | $\text{CP}_{na}$ | .736 | .690 | .696 | .920 | .896 | .668 |
| | $\text{CP}_{ad}$ | .848 | .820 | .836 | .940 | .926 | .736 |
| MA | $\overline{\widehat{R}_T}$ | .596 | .573 | .536 | .745 | .735 | .745 |
| | CP | .830 | .814 | .834 | .940 | .828 | .742 |

Maastricht University Leading in Learning!

# Simulation Study: $R_T$ Point Estimates

| | $p$ | 4 | | | 6 | | |
|---|---|---|---|---|---|---|---|
| | $n$ | 50 | 100 | 300 | 50 | 100 | 300 |
| | $R_T$ | .606 | .606 | .606 | .780 | .780 | .780 |
| BM | $\overline{\widehat{R}_T}$ | .601 | .573 | .531 | .744 | .733 | .744 |
| | $\text{CP}_{na}$ | .736 | .690 | .696 | .920 | .896 | .668 |
| | $\text{CP}_{ad}$ | .848 | .820 | .836 | .940 | .926 | .736 |
| MA | $\overline{\widehat{R}_T}$ | .596 | .573 | .536 | .745 | .735 | .745 |
| | CP | .830 | .814 | .834 | .940 | .828 | .742 |

$\Rightarrow$ Slight underestimation: At leat two random effects omitted

# Simulation Study: $R_T$ Point Estimates

| | $p$ | 4 | | | 6 | | |
|---|---|---|---|---|---|---|---|
| | $n$ | 50 | 100 | 300 | 50 | 100 | 300 |
| | $R_T$ | .606 | .606 | .606 | .780 | .780 | .780 |
| BM | $\overline{\widehat{R}_T}$ | .601 | .573 | .531 | .744 | .733 | .744 |
| | CP$_{na}$ | .736 | .690 | .696 | .920 | .896 | .668 |
| | CP$_{ad}$ | .848 | .820 | .836 | .940 | .926 | .736 |
| MA | $\overline{\widehat{R}_T}$ | .596 | .573 | .536 | .745 | .735 | .745 |
| | CP | .830 | .814 | .834 | .940 | .828 | .742 |

$\Rightarrow$ Slight underestimation: At leat two random effects omitted

$\Rightarrow$ Coverage probabilities much lower than $95\%$

# Simulation Study: Confidence Intervals

Averages of lower and upper limits of $95\%$ confidence intervals for $R_T$

| | | $n$ | 50 | 100 | 300 |
|---|---|---|---|---|---|
| 4 | $\text{BM}_{na}$ | $R_T$ | (.484 , .752) | (.476 , .716) | (.450 , .656) |
| | $\text{BM}_{ad}$ | $R_T$ | (.358 , .844) | (.362 , .785) | (.368 , .693) |
| | MA | $R_T$ | (.390 , .802) | (.395 , .750) | (.399 , .672) |
| 6 | $\text{BM}_{na}$ | $R_T$ | (.639 , .836) | (.656 , .811) | (.698 , .786) |
| | $\text{BM}_{ad}$ | $R_T$ | (.638 , .850) | (.654 , .811) | (.700 , .789) |
| | MA | $R_T$ | (.643 , .848) | (.657 , .812) | (.701 , .789) |

# Simulation Study: Best Ranked Models

Average $\widehat{R}_T$ for ranked models

| $p$ | 4 | | | 6 | | |
|---|---|---|---|---|---|---|
| $n$ | 50 | 100 | 300 | 50 | 100 | 300 |
| $R_T$ | .606 | .606 | .606 | .780 | .780 | .780 |
| Best | .601 | .573 | .531 | .744 | .733 | .744 |
| Second best | .569 | .561 | .583 | .755 | .773 | .776 |
| Third best | .332 | .260 | .206 | .441 | .453 | .459 |
| Fourth best | .377 | .359 | .289 | .455 | .444 | .471 |
| Fifth best | .467 | .476 | .449 | .494 | .499 | |

# Case Study: Reanalysis

Reliability estimates based on MA and on the five best models.

| | $R_T$ | $R_\Lambda$ |
|---|---|---|
| MA | 0.680 (0.610, 0.749) | 0.970 (0.924, 1.000) |
| 1 | 0.675 (0.601, 0.741) | 0.968 (0.930, 0.986) |
| 2 | 0.697 (0.657, 0.734) | 0.978 (0.595, 0.999) |
| 3 | 0.650 (0.547, 0.741) | 0.953 (0.905, 0.977) |
| 4 | 0.560 (0.467, 0.649) | 0.875 (0.765, 0.938) |
| 5 | 0.737 (0.705, 0.766) | 0.991 (0.987, 0.994) |

# Conclusions

- $\Rightarrow$ Model averaging: Alternative to the best model analysis
- $\Rightarrow$ Point estimates: BM$=$MA
- $\Rightarrow$ Confidence intervals: MA$=$wider, MA$\approx$BM$_{ad}$
- $\Rightarrow$ Good approximating model is all what is needed
- $\Rightarrow$ Similar fitting models lead to similar results