A Multivariate Multilevel Gaussian Model with a Mixed Effects Structure in the Mean and Covariance Part

Baoyue Li, Luk Bruyneel and Emmanuel Lesaffre

Biostatistics department, Erasmus MC

Bayes 2013 May 23, 2013

Outline

- Data description and research questions
- Multivariate multilevel factor model
- Bayesian estimation and identification issue
- Application to RN4CAST data
- Some future work

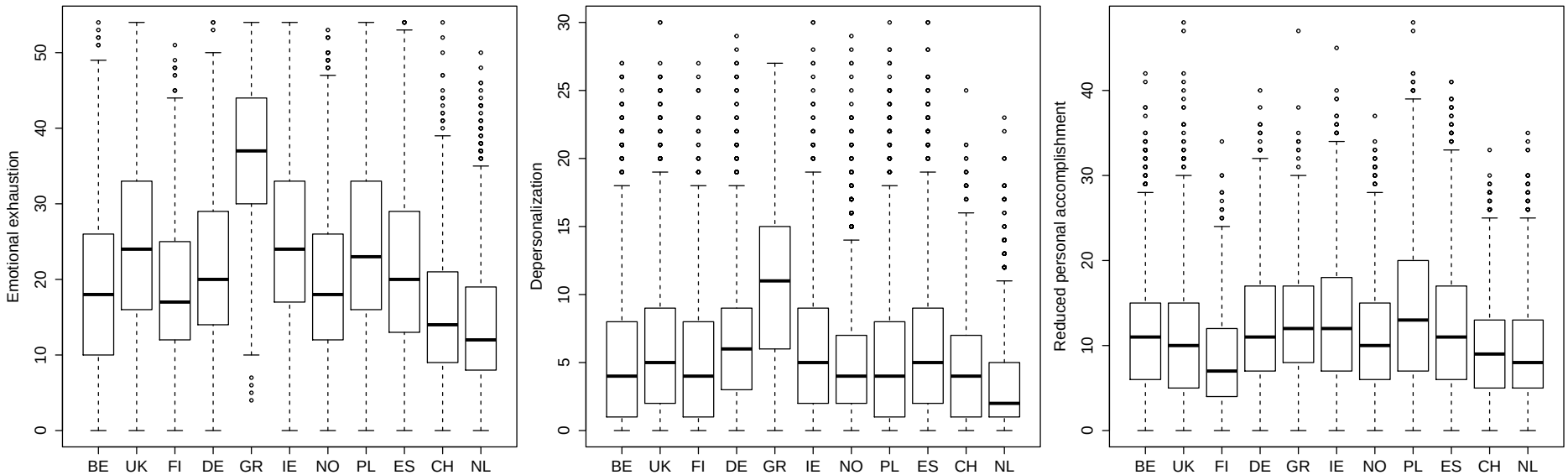
Data description and research questions

- The RN4CAST project
  - Registered Nurse Forecasting project
  - Nurse survey across Europe (2009-2011)
  - 33,731 nurses in 2,169 nursing units in 486 hospitals in 12 countries
  - Study the impact of system-level features of nursing care on nurse wellbeing and patient safety outcomes
    - Burnout, job satisfaction, turnover, etc.

Data description and research questions

- Three dimensions of burnout
  - Emotional exhaustion (EE)
  - Depersonalization (DP)
  - Reduced personal accomplishment (PA)
- Measured using the 22-item Maslach Burnout Inventory
  - Q: ”I feel emotionally drained from my work”
  - A: 0-never; 1-a few times a year or less; ...; 6-every day
- Sum of all items within each dimension as the outcome

Data description and research questions



Distribution of burnout at each country

Data description and research questions

- Some covariates:
  - Working experience (yrs)
  - Work environment
  - Hospital size and nursing unit size
  - Teaching hospital, technology hospital
  - Type of nursing unit (surgical or medical)

Data description and research questions

- Research questions:
  - Relationship of burnout and covariates at different levels
  - If the correlations among the 3 burnout dimensions remain the same (check for inter cultural differences)
    - At each level
    - For different levels of covariates

Multivariate multilevel factor model

- Basic idea: combining two models:
  - Gaussian multivariate mixed model: to estimate the mean structure
  - Factor model: rebuild COV via the factor loadings
    - Add covariates
    - Add random effects

Multivariate multilevel factor model

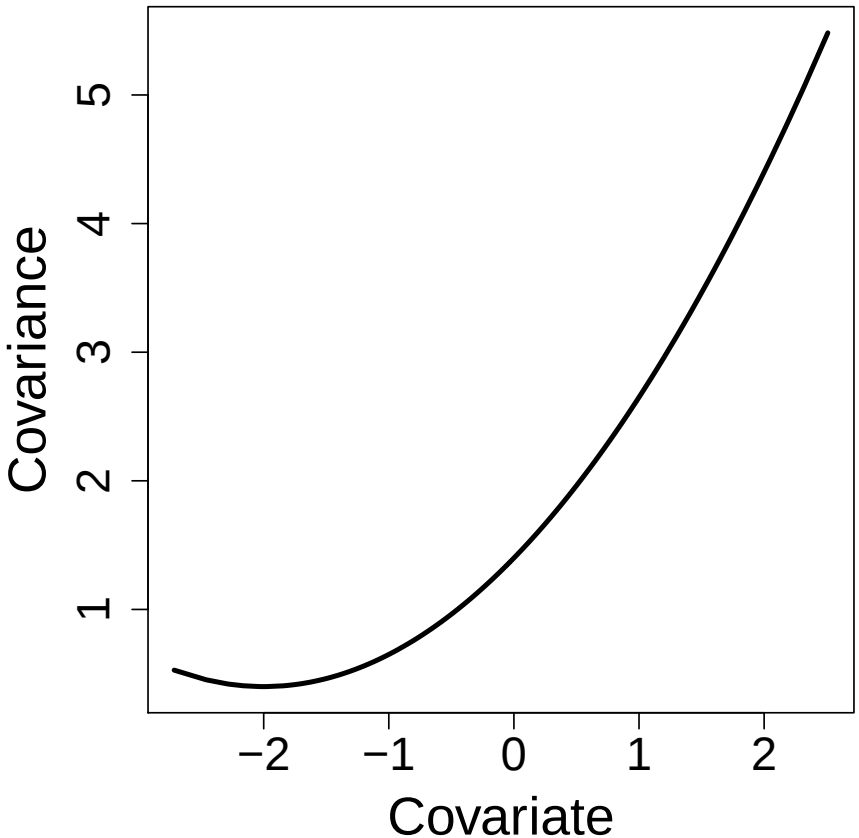
- The 2-level single-factor model:

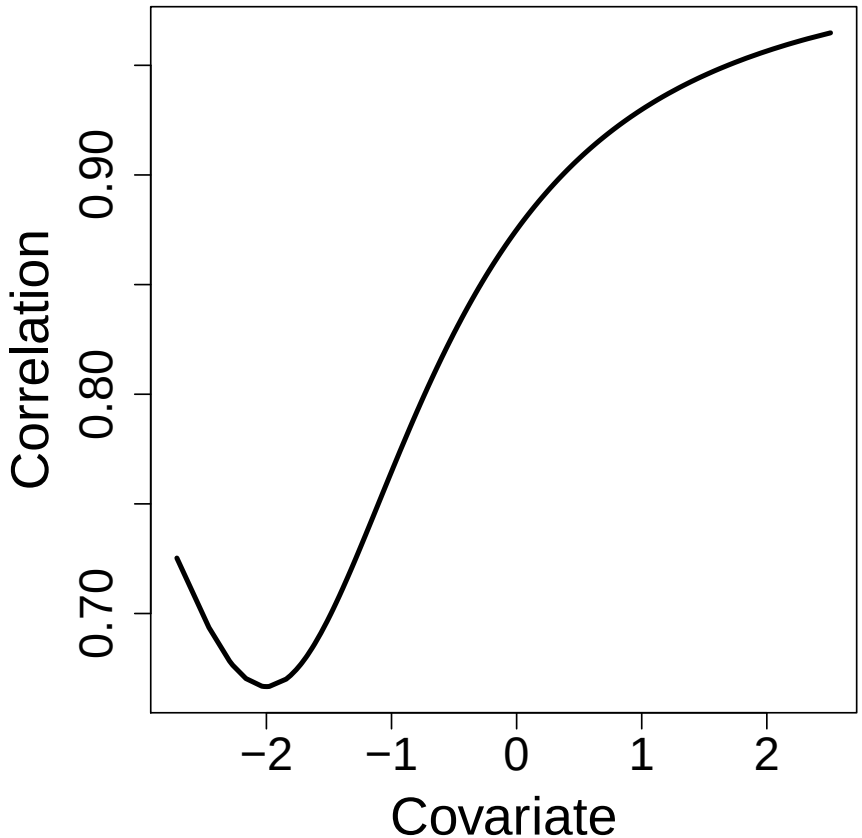
$$\begin{aligned}
\boldsymbol{y}_{ij} &= \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_{ij} + \boldsymbol{u}_j + \boldsymbol{\delta}_{ij}, \\
\boldsymbol{\delta}_{ij} &= (\boldsymbol{\beta}_0^* + \boldsymbol{\beta}_1^* x_{ij}^* + \boldsymbol{u}_j^*) F_{ij} + \boldsymbol{\varepsilon}_{ij} \\
\boldsymbol{u}_j &\sim N(\mathbf{0}, \Sigma_u), \quad \boldsymbol{u}_j^* \sim N(\mathbf{0}, \Sigma_u^*), \\
F_{ij} &\sim N(0,1), \quad \boldsymbol{\varepsilon}_{ij} \sim N(\mathbf{0}, \Sigma_\varepsilon)
\end{aligned}$$

- The conditional COV (on random effects):

$$\Sigma = \Sigma_\varepsilon + (\boldsymbol{\beta}_0^* + \boldsymbol{\beta}_1^* x_{ij}^* + \boldsymbol{u}_j^*)(\boldsymbol{\beta}_0^* + \boldsymbol{\beta}_1^* x_{ij}^* + \boldsymbol{u}_j^*)^T$$

Multivariate multilevel factor model





Relationship between covariance/correlation and covariate

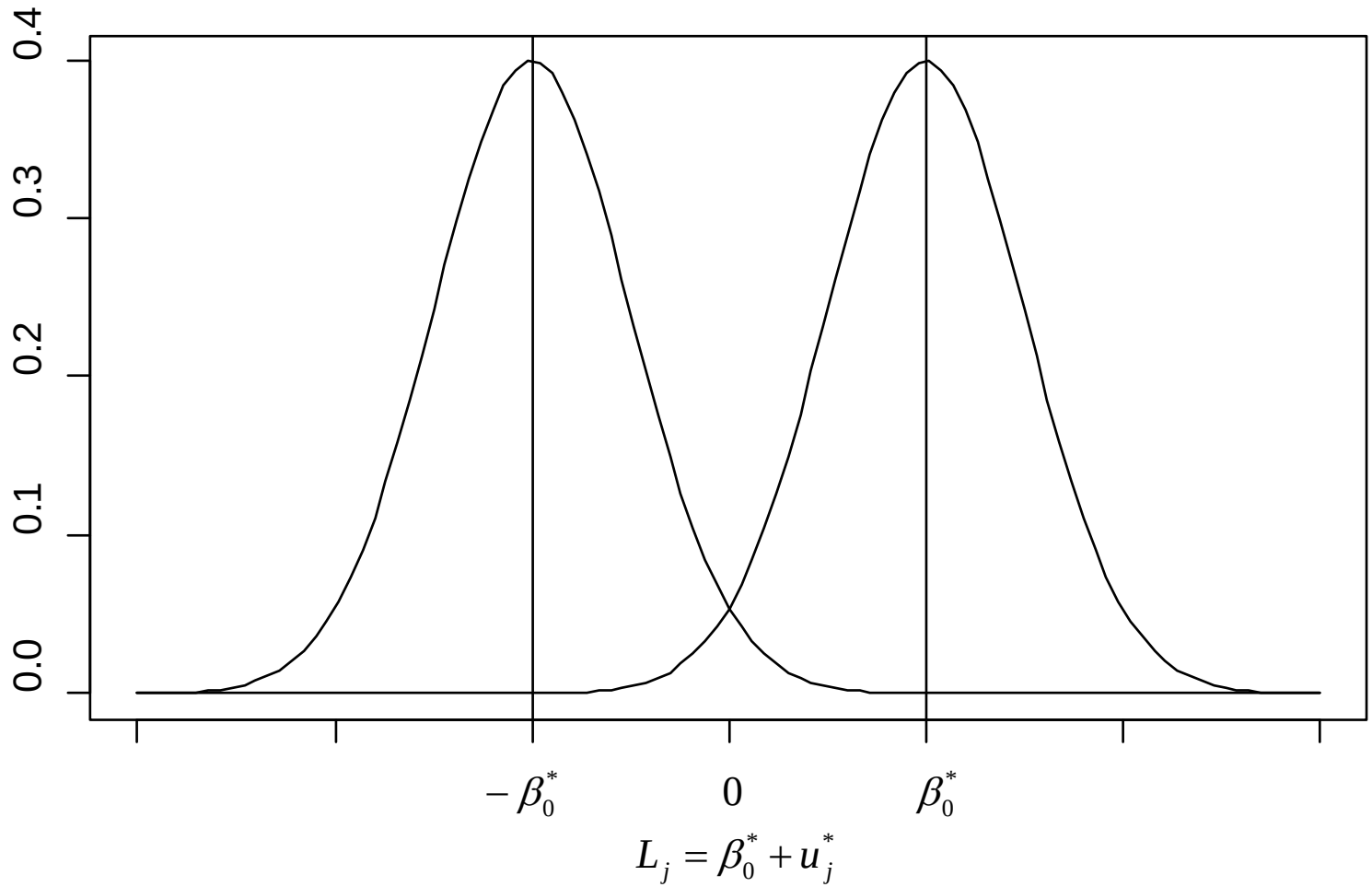
Bayesian estimation and identification issue

- Frequentist method may not be efficient
  - High dimensionality of random effects - Numeric problem for integration
  - Lack of software/packages to model COV appropriately
- Bayesian method
  - Avoid numeric problem by MCMC sampling
  - Quite flexible for complex modeling

Bayesian estimation and identification issue

- Identification issue in Bayesian estimation
  - "flipping states" issue: $\Lambda F \Longleftrightarrow (-\Lambda)(-F)$
  - A lesser problem for models without random effects (u_j^*) in loadings
  - This issue will be mixed up with the random effects, MCMC run will never converge
  - Solution: assign a mixture normal distribution to the loadings

Bayesian estimation and identification issues



The 2 normal distributions that form the mixture distribution for the factor loadings

Application to RN4CAST data

- 3-variate 4-level factor model
- Include all covariates in both the mean and loadings
- Random intercept at each level in both the mean and loadings
- Computational details
  - *dclone* package in R, 3 chains using 3 cores
  - 70,000 burnin $+$ 30,000 iterations
  - Convergence: Brooks-Gelman-Rubin plots, $Rhat < 1.1$; $MCMC\ error / SD < 5\%$
  - Model comparison: DIC (defined by Martyn Plummer) and PSBF

Application to RN4CAST data

- Better work environment, more working experience lead to less burnout
- Nurses working in a surgical nursing unit are more inclined to burnout than in a medical nursing unit.
- Adding covariates and random effects to COV improved the model fit largely
  - COV is different among countries, hospitals and nursing units
  - The more experienced the nurse is, the more correlation between EE and PA

Some future work

- Use the latent factor score for burnout instead of sum of the items
- Model COV at higher levels
- Relax some model assumptions:
  - Correlated random effects in the mean and loadings
  - Replace multivariate normal distribution of the random effects among the outcomes with multivariate t distribution

It is over!!!