Joint modelling of a multilevel factor analytic model and a multilevel covariance regression model

Baoyue Li, Luk Bruyneel and Emmanuel Lesaffre

Department of Biostatistics, Erasmus MC, Rotterdam

Bayes 2014, London May 22, 2014

Outline

- RN4CAST project and research questions
- Multilevel covariance regression (MCR) model
- Multilevel higher-order factor (MHOF) model
- Conclusions

The RN4CAST project

- Registered Nurse Forecasting (Sermeus et al., 2011)
- Nurse survey across Europe (2009-2011)
- 33,731 nurses, 2,169 nursing units, 486 hospitals and 12 countries
- **Aim**: Study the impact of system-level features of nursing care on nurse wellbeing and patient safety outcomes

Variables of interest

- 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” (EE)
  - A: 0-never; 1-a few times a year or less; ...; 6-every day

Covariates

- Covariates of interest:
  - Nurse level: *working experience (yrs), fulltime/part-time*
  - Nursing unit level: *work environment, size, surgical/medical, work load*
  - Hospital level: *teaching, technical, size*
  - Country level: aggregated variables

Research questions

- **RQ 1**: Are the **means** of the three burnout dimensions correlated with the organizational-level and individual-level characteristics?

- **RQ 2**: Are the **variances/correlations** among the three burnout dimensions stable across hospitals, nursing units and nurses, after taking into account a rich set of confounders at different levels?
  - Original English-written questionnaire was translated into several languages
  - Might be some variability in interpreting the questions
  - Reflected by not only the mean of burnout, could be also the correlation of the the burnout dimensions

Proposed solutions

- A multivariate multilevel model for
  - The mean structure
  - The covariance structure
- Different types of responses:
  - Sum scores with each dimension (continuous)
    - Multilevel covariance regression (MCR) (Li et al., 2014b)
  - The original 22 burnout items (ordinal)
    - Multilevel higher-order factor (MHOF) model (Li et al., 2014a)

MCR model: Specification

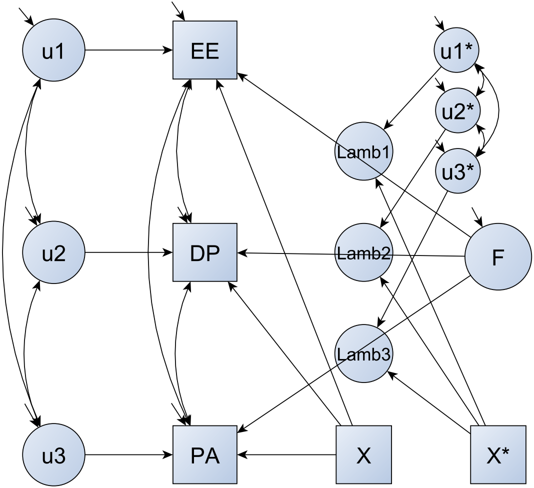
- A 2-level MCR model:

$$\boldsymbol{y}_{ij} = \boldsymbol{B}\boldsymbol{x}_{ij} + \boldsymbol{u}_j + \boldsymbol{\delta}_{ij}, \qquad \longrightarrow \text{A random intercept model}$$

$$\boldsymbol{\delta}_{ij} = \boldsymbol{\lambda}_{ij} F_{ij} + \boldsymbol{\varepsilon}_{ij},\ \boldsymbol{\lambda}_{ij} = \boldsymbol{B}^* \boldsymbol{x}^*_{ij} + \boldsymbol{u}^*_j, \ \longrightarrow \text{A factor model}$$

$$\boldsymbol{u}_j \sim N(\mathbf{0}, \Sigma_u), \quad \boldsymbol{u}^*_j \sim N(\mathbf{0}, \Sigma^*_u),$$
$$F_{ij} \sim N(0,1), \quad \boldsymbol{\varepsilon}_{ij} \sim N(\mathbf{0}, \Sigma_\varepsilon)$$

MCR model: Specification



MCR model: Implied marginal distribution

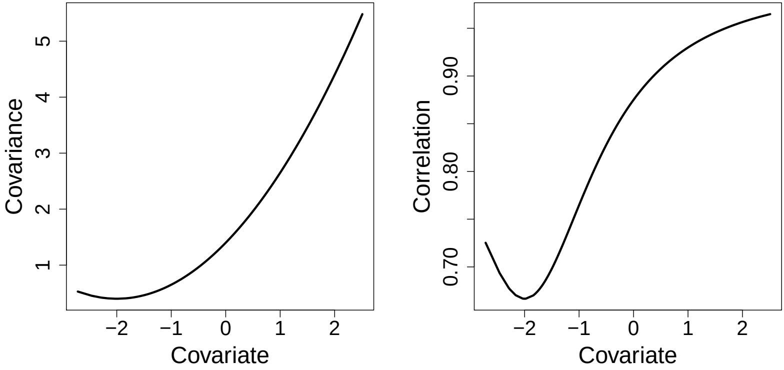
- The marginal covariance matrix of the responses is:

$$\Psi_{ij} = (\boldsymbol{B}^* \boldsymbol{x}_{ij}^* + \boldsymbol{u}_j^*)(\boldsymbol{B}^* \boldsymbol{x}_{ij}^* + \boldsymbol{u}_j^*)^T + \Sigma_\varepsilon$$

- Both level covariates could be added
- The factor model guarantees a positive definite covariance matrix Ψ_{ij}
- Easy interpretation: Ψ_{ij} depends on covariates quadratically

MCR model: Implied marginal distribution

- Relationship between covariance/correlation and covariate



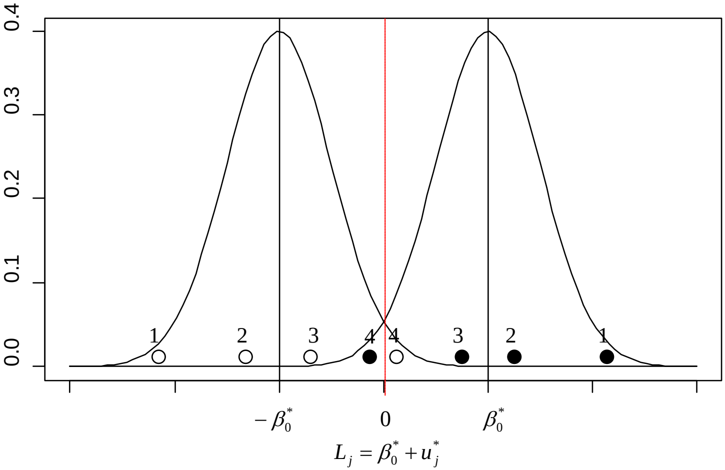
MCR model: Implied marginal distribution

- The marginal distribution of the response is not normal
- Skewness is zero if all random effects are independent
- (excess) Kurtosis is always non-negative
- The ability to handle heavy-tailed distributions

MCR model: Identification issues

- ”Flipping states” issue in factor model
- More complex with random effects in the loadings
- Leads to biased estimates if not properly taken care of
- Solution: use mixture prior for the random loadings
- Mixture of two multivariate normal distributions

MCR model: Identification issues



MCR model for the RN4CAST study

- A 3-variate 4-level MCR model:

$$\boldsymbol{y}_{ijkl} = \boldsymbol{B}\boldsymbol{x}_{ijkl} + \boldsymbol{u}_{jkl} + \boldsymbol{u}_{kl} + \mathbf{u}_l + \boldsymbol{\delta}_{ijkl}$$

$$\boldsymbol{\delta}_{ijkl} = \boldsymbol{\lambda}_{ijkl} F_{ijkl} + \boldsymbol{\varepsilon}_{ijkl}, \quad \boldsymbol{\lambda}_{ijkl} = \boldsymbol{B}^* \boldsymbol{x}^*_{ijkl} + \boldsymbol{u}^*_{jkl} + \boldsymbol{u}^*_{kl} + \boldsymbol{u}^*_l$$

$$\boldsymbol{u}_{jkl} \sim N(\mathbf{0}, \Sigma_u), \quad \boldsymbol{u}_{kl} \sim N(\mathbf{0}, \Sigma_h), \quad \boldsymbol{u}_l \sim N(\mathbf{0}, \Sigma_c),$$
$$\boldsymbol{u}^*_{jkl} \sim N(\mathbf{0}, \Sigma^*_u), \quad \boldsymbol{u}^*_{kl} \sim N(\mathbf{0}, \Sigma^*_h), \quad \boldsymbol{u}^*_l \sim N(\mathbf{0}, \Sigma^*_c),$$
$$F_{ijkl} \sim N(0, 1), \quad \boldsymbol{\varepsilon}_{ijkl} \sim N(\mathbf{0}, \Sigma_\varepsilon)$$

Computational aspects

- Bayesian approach (MCMC method) was used
  - Large number of random effects
  - Various distributions for the random effects
  - Various transformations of parameters
- JAGS through R packages *rjags/dclone*
- Model comparison: DIC and PSBF (Pseudo Bayes Factor)
- Convergence check: BGR plots and PSRF (Potential Scale Reduction Factor)
- Goodness of fit: PPC (Posterior Predictive Check) with χ^2 discrepancy function

MCR model: Main results

- Mean part of burnout:
  - Fulltime nurses $\Longrightarrow$ more burnout
  - Experienced nurses $\Longrightarrow$ less burnout
  - Better work environment $\Longrightarrow$ less burnout
  - Heavier work load $\Longrightarrow$ more burnout

MCR model: Main results

- Covariance part of burnout:
  - Experienced nurses have a larger variance of burnout
    - Burnout feelings get pronounced over years
  - Random effects: variance of burnout differs across units
    - Measure the degree of equality of work load – ”harmonic burden”

Working on 22 items directly

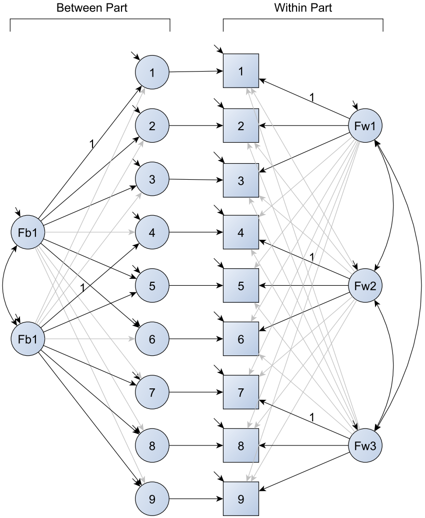
- Burnout was originally measured through 22 items
- The three dimensions was proposed by Maslach and Jackson (1981) on different population
- These dimensions might change
- Alternative: model the original 22 items directly
- MCR not efficient to hand high-dimensional response
- A factor model to find the correct burnout dimensions (MFA)
- Jointly estimate the MFA and MCR – MHOF model (Li et al., 2014a)

MFA model: Specification

- Find the latent factors underlying a group of variables in a multilevel context
- A two level MFA model is:

$$\begin{gathered}
\boldsymbol{y}_{ij} = \boldsymbol{\mu} + \boldsymbol{L}_B \boldsymbol{f}_j + \boldsymbol{u}_j + \boldsymbol{L}_W \boldsymbol{f}_{ij} + \boldsymbol{\varepsilon}_{ij}, \\
\boldsymbol{f}_j \sim N(\mathbf{0}, \Sigma_{fB}), \quad \boldsymbol{u}_j \sim N(\mathbf{0}, \Sigma_u), \\
\boldsymbol{f}_{ij} \sim N(\mathbf{0}, \Sigma_{fW}), \quad \boldsymbol{\varepsilon}_{ij} \sim N(\mathbf{0}, \Sigma_\varepsilon), \\
i = 1, 2, ..., n_j;\ j = 1, 2, ..., k,
\end{gathered}$$

MFA model: Specification



MHOF model

- The MFA part:
    - Factor structure at the lowest level only
    - Estimate the whole covariance matrix at higher levels
- The MCR part:
    - Use the nurse-level factor scores as responses
    - Include covariates at each level

MHOF model: Specification

$The\ MFA\ part:$

$$\boldsymbol{y}_{ij} = \boldsymbol{\mu} + \boldsymbol{b}_j + \boldsymbol{L}\boldsymbol{z}_{ij} + \boldsymbol{\varepsilon}_{ij}^{FA},$$

$$\boldsymbol{b}_j \sim N(\mathbf{0}, \Sigma_u), \quad \boldsymbol{\varepsilon}_{ij}^{FA} \sim N(\mathbf{0}, diag(\sigma_1^2, \sigma_2^2, ...\sigma_p^2))$$

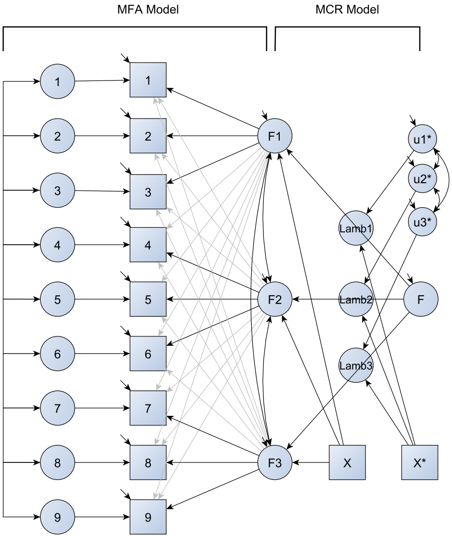
$The\ MCR\ part:$

$$\boldsymbol{z}_{ij} = \boldsymbol{B}\boldsymbol{x}_{ij} + \boldsymbol{\delta}_{ij},$$

$$\boldsymbol{\delta}_{ij} = \boldsymbol{\Lambda}_{ij} F_{ij} + \boldsymbol{\varepsilon}_{ij}^{CR}, \quad \boldsymbol{\Lambda}_{ij} = \boldsymbol{B}^* \boldsymbol{x}_{ij}^* + \boldsymbol{u}_j^*,$$

$$\boldsymbol{u}_j^* \sim N(\mathbf{0}, \Sigma_u^*), \quad F_{ij} \sim N(0, 1), \quad \boldsymbol{\varepsilon}_{ij}^{CR} \sim N(\mathbf{0}, \Sigma_\varepsilon)$$

MHOF model: Specification



MHOF model: Identification issues

- All the cross loadings in the MFA part are estimated, but with informative priors (e.g. $N(0, 0.05)$) (BSEM, Muthén and Asparouhov (2012))
- Same as MCR model, mixture prior is applied to the factor loading parameters in the MCR part

MHOF model: Implied marginal distribution

- The MCR part of the MHOF model shares the same properties with the MCR model
- The lower-order factor scores have a non-normal distribution with zero skewness and non-zero kurtosis

Apply to the RN4CAST study

- Applied to Belgian part of RN4CAST study
  - 2809 nurses, 268 nursing units, 55 hospitals
- A 3-level MFA model based on 22 items and a 3-variate 3-level MCR model are jointly estimated
  - Quite similar findings as from the MCR model: Mean part and covariance part

Conclusions

- MCR model provides a novel way of modeling covariance matrix hierarchically
- Model both the mean and the covariances simultaneously
- MHOF model could handle high-dimensional data well
- It can be seen as a multilevel SEM, with a complex structural part not being done before
- Find ”hidden” information:
    - Variance varies across unit, indicating different degrees of work load equality

Main references

Li, B., Bruyneel, L., and Lesaffre, E. (2014a). Multilevel Higher-Order Factor Model: joint modeling of a multilevel factor analytic model and a multilevel covariance regression model. *Structural Equation Modeling: A Multidisciplinary Journal*. under review.

Li, B., Bruyneel, L., and Lesaffre, E. (2014b). A multivariate multilevel Gaussian model with a mixed effeccts structure in the mean and covariance part. *Statistics in Medicine*, 33(11):1877–1899.

Maslach, C. and Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Organizational Behavior*, 2(2):99–113.

Muthén, B. and Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17(3):313–335.

Sermeus, W., Aiken, L., Van den Heede, K., Rafferty, A., Griffiths, P., Moreno-Casbas, M., Busse, R., Lindqvist, R., Scott, A., Bruyneel, L., et al. (2011). Nurse forecasting in Europe (RN4CAST): Rationale, design and methodology. *BMC Nursing*, 10(1):6.

Thanks for your time!