Institute for Employment Research

The Research Institute of the Federal Employment Agency

Discussing Effects of Different MAR-Settings

Research Seminar, Department of Statistics, LMU Munich

Munich, 11.07.2014

Matthias Speidel

Jörg Drechsler

Joseph Sakshaug

Outline

- What we basically want to do: hierarchical imputation
- What we encounter: effects of MAR-settings
- Questions: explanation of these effect

Our research

- DFG-Project
- Reason for hierarchical imputation:
  - Item-Nonresponse → nonresponse bias
  - Imputation → reduce or unmake nonresponse bias
  - Hierarchical data → hierarchical imputation
- Aim for unbiased linear mixed model parameter estimates
- Which imputation method works well, which not? And why?
  - Focus on *dummy variables imputation*

Notation

Hierarchical data in our case:

$$y_{j,i} = \beta_0 + b_{0_j} + (\beta_1 + b_{1_j}) \cdot x_{j,i} + \varepsilon_{j,i}$$

$$i = 1, \ldots, n_j$$

$$j = 1, \ldots, m$$

$$n_j \mathrel{\hat{=}} \textit{size of cluster } j$$

$$m \mathrel{\hat{=}} \textit{number of clusters}$$

$$\begin{pmatrix} b_{0_j} \\ b_{1_j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \Sigma_b := \begin{pmatrix} \sigma^2_{b_0} & \sigma_{b_1 b_0} \\ \sigma_{b_0 b_1} & \sigma^2_{b_1} \end{pmatrix}\right)$$

$$\varepsilon_{j,i} \sim N(0, \sigma^2_\varepsilon)$$

$$\beta_j = \beta + b_j$$

Dummy variables imputation

- Dummy imputation only hierarchical imputation method implemented in SAS, SPSS and Stata.
- Estimates cluster specific intercept and slope imputation parameters (assumed fixed not random)
- Under which data and missing generating setting biased results?
- Explanations for the bias?

Shortly: the main results

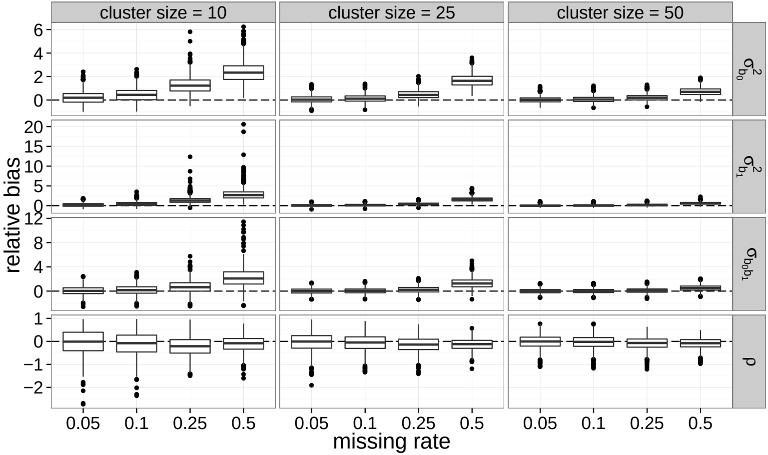

Figure: Relative bias after dummy imputation for different settings.

Side results

- Not only data setting important, but also shape of *missing function*.
- Don't mean *MCAR* vs. *MAR* vs. *MNAR* (Rubin (1976)). Mean different missing functions within MAR.

General missing function

Missing function for standardized X_i, desired missing rate $MR \leq 0.5$ and $-1 \leq s \leq 1$:

$$P(Y_i = NA | X_i,\ MR,\ s) = MR \cdot (1 - s) + 2 \cdot MR \cdot s \cdot logit(X_i)^{-1} \tag{1}$$

- s controls strength and direction of X's influence on the missing probabilities.
  - MCAR if $s = 0$
- Missing probabilities on average equal to MR if X symmetrically distributed around 0 (proof missing).

Illustrations missing mechanism: varying *MR*

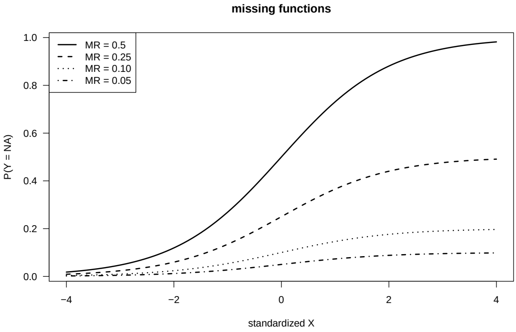

Figure: Illustration of missing probabilities for different missing rates

Illustrations missing mechanism: varying s

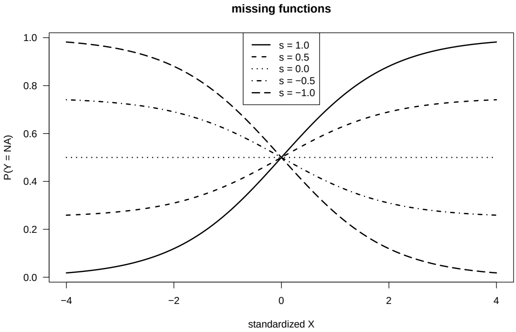

Figure: Five different shapes of missing functions ($s = 1,\ 0.5,\ 0,\ -0.5\ ,\ -1$). All lead to MAR, except MCAR of course.

Results when estimating Σ_b after dummy imputation

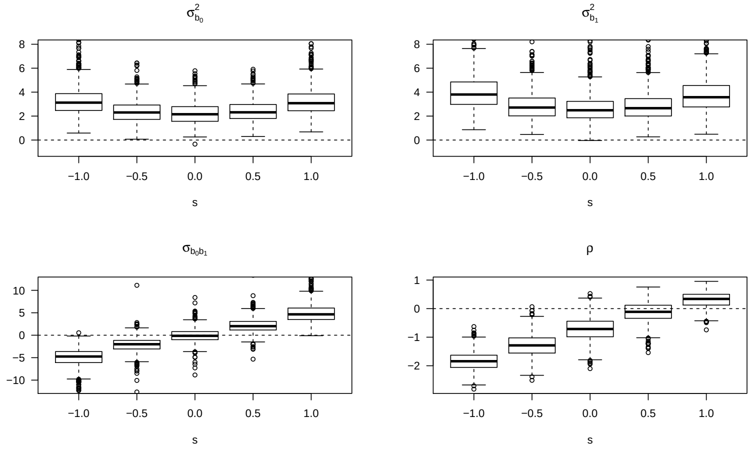

Figure: Relative empirical biases under different missing functions.

Questions

1. Explanation for differences in parameter estimates for different MAR settings?
2. Sufficient to look at variance matrix from imputation parameters (instead of variances after imputation)?
   - Sufficient to look at variance matrices of a single cluster?
3. If bias in $var(\hat{\beta}_j)^{fix}$ not clearly visible, sufficient to compare it to the unbiased $var(\hat{\beta}_j)^{rand}$?
   - How to say "matrix A is larger than matrix B"? Elementwise comparison?
   - Multiplicative or additive matrix? Interpretation?

Imputation parameters variances

Main interest in

$$Var(\hat{\beta}_j|\beta,\ \Sigma_b,\ \sigma_\varepsilon^2)^{fix} \cong (readability)\, Var(\hat{\beta}_j)^{fix} = \sigma_\varepsilon^2 \cdot (X^T X)^{-1} \tag{2}$$

Maybe compare it to what can be found in (Goldstein, 2011: p. 69)

$$Var(\hat{\beta}_j|\beta,\ \Sigma_b,\ \sigma_\varepsilon^2)^{rand} \cong Var(\hat{\beta}_j)^{rand} = \sigma_\varepsilon^2 \cdot \left(X^T X + \sigma_\varepsilon^2 \cdot \Sigma_b^{-1}\right)^{-1} \tag{3}$$

Beside elementwise comparison, multiplicative and additive term

$$\gamma \cdot Var(\hat{\beta}_j)^{rand} = Var(\hat{\beta}_j)^{fix} \tag{4}$$

$$\delta + Var(\hat{\beta}_j)^{rand} = Var(\hat{\beta}_j)^{fix} \tag{5}$$

Multiplicative disparity

Properties and interpretation of γ?

$$
\begin{aligned}
\gamma \cdot Var(\hat{\beta}_j)^{rand} &= Var(\hat{\beta}_j)^{fix} \\
\gamma \cdot \sigma_\varepsilon^2 \cdot \left(X^T X + \sigma_\varepsilon^2 \cdot \Sigma_b^{-1}\right)^{-1} &= \sigma_\varepsilon^2 \cdot (X^T X)^{-1} \\
\gamma &= (X^T X)^{-1} \cdot \left(X^T X + \sigma_\varepsilon^2 \cdot \Sigma_b^{-1}\right) \\
&= (X^T X)^{-1} \cdot X^T X + (X^T X)^{-1} \cdot \sigma_\varepsilon^2 \cdot \Sigma_b^{-1} \\
&= I + (X^T X)^{-1} \cdot \sigma_\varepsilon^2 \cdot \Sigma_b^{-1} \\
&= I + Var(\hat{\beta}_j)^{fix} \cdot \Sigma_b^{-1}
\end{aligned} \tag{6}
$$

Additive disparity

Properties and interpretation of δ?

$$
\begin{aligned}
\gamma \cdot Var(\hat{\beta}_j)^{rand} &= Var(\hat{\beta}_j)^{fix} \\
(I + Var(\hat{\beta}_j)^{fix} \cdot \Sigma_b^{-1}) \cdot Var(\hat{\beta}_j)^{rand} &= Var(\hat{\beta}_j)^{fix} \qquad (7) \\
Var(\hat{\beta}_j)^{rand} + Var(\hat{\beta}_j)^{fix} \cdot \Sigma_b^{-1} \cdot Var(\hat{\beta}_j)^{rand} &= Var(\hat{\beta}_j)^{fix}
\end{aligned}
$$

Now, we have the matrix δ:

$$
\begin{aligned}
\delta &= Var(\hat{\beta}_j)^{fix} \cdot \Sigma_b^{-1} \cdot Var(\hat{\beta}_j)^{rand} \\
\delta &= (X^T X)^{-1} \cdot \sigma_\varepsilon^2 \cdot \Sigma_b^{-1} \cdot \sigma_\varepsilon^2 \cdot \left(X^T X + \sigma_\varepsilon^2 \cdot \Sigma_b^{-1}\right)^{-1} \qquad (8)
\end{aligned}
$$

References I

Goldstein, Harvey (2011): Multilevel Statistical Models. Chichester (UK): Wiley, 4 ed..

Rubin, Donald B. (1976): Inference and Missing Data. In: Biometrika, Vol. 63, No. 3, p. 581 – 592, URL http://biomet.oxfordjournals.org/content/63/3/581.short.

Additional material

Dummy imputation:

$$Var(\hat{\beta}_j)^{fix} = \frac{\sigma_\varepsilon^2}{n \cdot \sum x_i^2 - (\sum x_i)^2} \cdot \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} \tag{9}$$

$$= \frac{\sigma_\varepsilon^2}{n^2 \cdot var(x)} \cdot \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} \tag{10}$$

Mixed effects imputation:

$$Var(\hat{\beta}_j)^{rand} = \sigma_\varepsilon^2 \cdot \left(\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} + \frac{\sigma_\varepsilon^2}{\sigma_0^2 \cdot \sigma_1^2 - \sigma_{01}^2} \cdot \begin{pmatrix} \sigma_1^2 & -\sigma_{01} \\ -\sigma_{01} & \sigma_0^2 \end{pmatrix} \right)^{-1} \tag{11}$$

To shorten the expression, I define $a := \frac{\sigma_\varepsilon^2}{\sigma_0^2 \cdot \sigma_1^2 - \sigma_{01}^2}$.

$$Var(\hat{\beta}_j)^{rand} = \sigma_\varepsilon^2 \cdot \begin{pmatrix} n + a \cdot \sigma_1^2 & \sum x_i - a \cdot \sigma_{01} \\ \sum x_i - a \cdot \sigma_{01} & \sum x_i^2 + a \cdot \sigma_0^2 \end{pmatrix}^{-1} \tag{12}$$

Details of imputation parameters variance (cont.)

$$Var(\hat{\beta}_j)^{rand} = \frac{\sigma_\varepsilon^2}{(n + a \cdot \sigma_1^2) \cdot (\sum x_i^2 + a \cdot \sigma_0^2) - (\sum x_i - a \cdot \sigma_{01})^2} \cdot \begin{pmatrix} \sum x_i^2 + a \cdot \sigma_0^2 & -\sum x_i + a \cdot \sigma_{01} \\ -\sum x_i + a \cdot \sigma_{01} & n + a \cdot \sigma_1^2 \end{pmatrix} \qquad (13)$$

Details on *delta*

$$
\begin{aligned}
\delta =& Var(\hat{\beta}_j)^{fix} \cdot \Sigma_b^{-1} \cdot Var(\hat{\beta}_j)^{rand} \\
=& \frac{\sigma_\varepsilon^2}{n^2 \cdot var(x)} \cdot \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} \\
& \cdot \frac{1}{\sigma_0^2 \cdot \sigma_1^2 - \sigma_{01}^2} \cdot \begin{pmatrix} \sigma_1^2 & -\sigma_{01} \\ -\sigma_{01} & \sum \sigma_0^2 \end{pmatrix} \\
& \cdot \frac{\sigma_\varepsilon^2}{(n + a \cdot \sigma_1^2) \cdot (\sum x_i^2 + a \cdot \sum \sigma_0^2) - (\sum x_i - a \cdot \sigma_{01})^2} \\
& \cdot \begin{pmatrix} \sum x_i^2 + a \cdot \sum \sigma_0^2 & -\sum x_i + a \cdot \sigma_{01} \\ -\sum x_i + a \cdot \sigma_{01} & n + a \cdot \sigma_1^2 \end{pmatrix}
\end{aligned} \tag{14}
$$

Our simulation parameters

Hierarchical data in our case:

$$y_{j,i} = 62.1 + b_{0_j} + (0.8 + b_{1_j}) \cdot x_{j,i} + \varepsilon_{j,i}$$

$$\begin{pmatrix} b_{0_j} \\ b_{1_j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \Sigma_b := \begin{pmatrix} 6.2 & 0.4 \\ 0.4 & 0.1 \end{pmatrix}\right)$$

$$\varepsilon_{j,i} \sim N(0, 92.2)$$

$$x_{j,i} \sim N(0, 81)$$

$$n_j \textit{ dependent on setting}: \; 10, \; 25, \; \textit{or } 50$$

$$m = 100$$

Individuals' class affiliation known and variable in data set. Parameters from NEPS data (and similar to CILS4EU data).

Matthias Speidel
Matthias.Speidel@iab.de

Jörg Drechsler
Jörg.Drechsler@iab.de

Joseph Sakshaug
Joseph.Sakshaug@iab.de