Multiple Imputation of Ordinal Missing Not at Random Data

Dr. Angelina Hammon
DIW Berlin, SOEP

ASI Jahrestagung

TU Dresden, 21.09.23

Motivation

- Standard multiple imputation (MI) assumes ignorable missing data, i.e., missingness only depends on observed information (MAR)
- Unrealistic assumption in many applications
- Non-consideration may lead to invalid inferences

MI and non-ignorable missing data

- MI can be used under every missing-data mechanism, but it has to be incorporated into the imputation model
- Joint distribution $f(Y, R)$ has to be modeled (pattern-mixture or selection modeling)
- Presence of MNAR cannot be tested empirically
- Alternative imputation models rely on unverifiable, external assumptions

$\Rightarrow$ Need of sensitivity analyses!

- **But**: Techniques for non-ignorable missing data hardly used in social sciences $\rightarrow$ lack of suitable implementations usable with typical survey data

MI and non-ignorable missing data

- MI can be used under every missing-data mechanism, but it has to be incorporated into the imputation model
- Joint distribution $f(Y, R)$ has to be modeled (pattern-mixture or selection modeling)
- Presence of MNAR cannot be tested empirically
- Alternative imputation models rely on unverifiable, external assumptions

$\Rightarrow$ Need of sensitivity analyses!

- **But**: Techniques for non-ignorable missing data hardly used in social sciences $\rightarrow$ lack of suitable implementations usable with typical survey data

MI and non-ignorable missing data

- MI can be used under every missing-data mechanism, but it has to be incorporated into the imputation model
- Joint distribution $f(Y, R)$ has to be modeled (pattern-mixture or selection modeling)
- Presence of MNAR cannot be tested empirically
- Alternative imputation models rely on unverifiable, external assumptions

$\Rightarrow$ Need of sensitivity analyses!

- **But**: Techniques for non-ignorable missing data hardly used in social sciences $\rightarrow$ lack of suitable implementations usable with typical survey data

Typical data situation

- Survey variables often not continuous, but categorical
- Usually multivariate missing data
- Survey data often clustered in higher-level units
- Possibly non-ignorable missing data

⇒ Ordered probit model with sample selection and random intercept as (univariate) imputation model in the framework of Fully Conditional Specification

Typical data situation

- Survey variables often not continuous, but categorical
- Usually multivariate missing data
- Survey data often clustered in higher-level units
- Possibly non-ignorable missing data

⇒ Ordered probit model with sample selection and random intercept as (univariate) imputation model in the framework of Fully Conditional Specification

Typical data situation

- Survey variables often not continuous, but categorical
- Usually multivariate missing data
- Survey data often clustered in higher-level units
- Possibly non-ignorable missing data

$\Rightarrow$ Ordered probit model with sample selection and random intercept as (univariate) imputation model in the framework of Fully Conditional Specification

Ordered probit model with sample selection and random intercept (e.g. Greene 2012)

For $j = 1, \dots, J$ clusters consisting of $i = 1, \dots, n_j$ individuals, respectively, $h = 1, \dots, H$ observed categories of Y, and where $x_{Y,ji} \subset x_{R,ji}$ (exclusion criterion) should hold, the model can be specified as follows

$$\text{Joint model} \begin{cases} \text{Selection equation: } r^*_{ji} = \beta_R^\top x_{R,ji} + \alpha_{R,j} + \epsilon_{R,ji} \\ \text{Outcome equation: } y^*_{ji} = \beta_Y^\top x_{Y,ji} + \alpha_{Y,j} + \epsilon_{Y,ji} \end{cases}$$

$$r_{ji} = \mathbf{1}(r^*_{ji} > 0),$$

$$y_{ji} = \begin{cases} h & \text{if } \kappa_{h-1} < y^*_{ji} \leq \kappa_h \ \& \ r_{ji} = 1 \\ \text{NA} & \text{if } r_{ji} = 0. \end{cases}$$

$$\begin{pmatrix} \epsilon_R \\ \epsilon_Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$
$$\begin{pmatrix} \alpha_R \\ \alpha_Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_R^2 & \tau\sigma_R\sigma_Y \\ \tau\sigma_R\sigma_Y & \sigma_Y^2 \end{pmatrix} \right)$$

- κ_h with $h = 1, \dots, H$ are threshold parameters, where $\kappa_0 = -\infty$ and $\kappa_H = +\infty$
- ϵ denotes the individual error terms which are correlated by ρ
- α constitutes the random intercepts with variance σ^2 and correlation coefficient τ

Approximate log likelihood using AGHQ

The double integral in the likelihood function of the selection model can be approximated by applying Adaptive Gauss-Hermite quadrature (e.g. Butler & Moffitt 1982; Hartzel, Agresti & Caffo 2001):

$$
\begin{aligned}
\ln L^{AGHQ} \simeq \sum_{j=1}^{J} \ln \Bigg\{ |\mathbf{\Omega_j}|^{1/2}\, 2 \sum_{p_1=1}^{P} \sum_{p_2=1}^{P} \omega_{p_1}\omega_{p_2} \prod_{i=1}^{n_j} \Bigg[(1-r_{ji})\Phi\left(-(\beta_R^\top x_{R,ji} + \tilde{a}_{jp_1})\right) \\
+ r_{ji} \sum_{h=1}^{H} m_{jih} \Bigg(\Phi_2\left(\beta_R^\top x_{R,ji} + \tilde{a}_{jp_1}, \kappa_h - (\beta_Y^\top x_{Y,ji} + \tilde{a}_{jp_2}), -\rho\right) \\
- \Phi_2\left(\beta_R^\top x_{R,ji} + \tilde{a}_{jp_1}, \kappa_{h-1} - (\beta_Y^\top x_{Y,ji} + \tilde{a}_{jp_2}), -\rho\right) \Bigg) \Bigg] \\
\cdot \phi_2(\tilde{\mathbf{a}}_{\mathbf{jp}} | \mathbf{0}, \mathbf{\Sigma}) \ \exp(\mathbf{a_p}'\mathbf{a_p}) \Bigg\},
\end{aligned}
$$

- With $\tilde{\mathbf{a}}_{\mathbf{jp}} = \hat{\boldsymbol{\mu}}_{\mathbf{j}} + \sqrt{2}\hat{\mathbf{\Omega}}_{\mathbf{j}}^{1/2}\mathbf{a_p} = (\tilde{a}_{jp_1}, \tilde{a}_{jp_2})'$ for cluster j and quadrature step p
- Quadrature points $p_1 = 1, \dots, P$ and $p_2 = 1, \dots, P$
- ω denote the Gauss-Hermite weights, a are nodes from the Hermite polynomial
- $\hat{\boldsymbol{\mu}}$ and $\hat{\mathbf{\Omega}}$ are estimated by the random intercepts' mode and curvature matrix (negative inverse hessian) at the mode (Liu & Pierce 1994)

Imputation algorithm ordinal multilevel data

Let $\theta = (\beta_Y, \beta_R, r, \xi_Y^2, \xi_R^2, z, \delta_l)$ be the model parameters where $l = 1, \dots, H-1$, $r = \operatorname{atanh} \rho$, $\xi_Y^2 = \ln \sigma_Y^2$, $\xi_R^2 = \ln \sigma_R^2$, $z = \operatorname{atanh} \tau$, $\delta_1 = \kappa_1$, and $\delta_l = \ln(\kappa_l - \kappa_{l-1})$ for all $l > 1$.

1. Estimation of model parameters by Maximum Likelihood using AGHQ and the BFGS algorithm
   - $\hat{\theta} = (\hat{\beta}_Y, \hat{\beta}_R, \hat{r}, \hat{\xi}_Y^2, \hat{\xi}_R^2, \hat{z}, \hat{\delta}_l)$,
   - $\hat{\Psi}$, the variance covariance matrix of $\hat{\theta}$.
2. Draw $\dot{\theta} = (\dot{\beta}_Y, \dot{\beta}_R, \dot{r}, \dot{\xi}_Y^2, \dot{\xi}_R^2, \dot{z}, \dot{\delta}_l)'$ from $N(\hat{\theta}, \hat{\Psi})$ and retransform $\dot{\rho} = \tanh \dot{r}$, $\dot{\tau} = \tanh \dot{z}$, $\dot{\sigma}_Y^2 = \exp \dot{\xi}_Y^2$, $\dot{\sigma}_R^2 = \exp \dot{\xi}_R^2$ and $\dot{\kappa}_1 = \dot{\delta}_1$, $\dot{\kappa}_l = \exp(\dot{\delta}_l) + \dot{\kappa}_{l-1}$ for all $l > 1$.
3. Draw the random intercepts $(\dot{\alpha}_Y, \dot{\alpha}_R)'$ for each cluster from $N\left(\begin{pmatrix}\hat{\alpha}_Y \\ \hat{\alpha}_R\end{pmatrix}, \hat{\boldsymbol{\Omega}}\right)$

   where $(\hat{\alpha}_Y, \hat{\alpha}_R)'$ denote the estimated modes of the random intercepts and $\hat{\boldsymbol{\Omega}}$ is calculated by the inverse hessian at the modes.
4. Calculate $\dot{p}_h$ for $h = 1, \dots, H$ categories

$$\dot{p}_h = P(\dot{Y} = h | X_Y, X_R, R = 0) = \frac{\Phi_2(-(X_R\dot{\beta}_R + \dot{\alpha}_R), \dot{\kappa}_h - (X_Y\dot{\beta}_Y + \dot{\alpha}_Y), \dot{\rho})}{\Phi(-(X_R\dot{\beta}_R + \dot{\alpha}_R))} - \frac{\Phi_2(-(X_R\dot{\beta}_R + \dot{\alpha}_R), \dot{\kappa}_{h-1} - (X_Y\dot{\beta}_Y + \dot{\alpha}_Y), \dot{\rho})}{\Phi(-(X_R\dot{\beta}_R + \dot{\alpha}_R))}.$$

5. Draw for each Y_{mis} a replacement from $Multinom(\dot{p}_1, \dots, \dot{p}_H)$.

Simulation study: multilevel data

Table: Mulitlevel simulation results for $\beta_1 = 1$ estimates in 500 simulation runs.

Methods	Mechanism	emp.mean	rel.bias (%)	CR (%)
Before deletion	MAR	0.9987	−0.47	94.4
	MNAR sel.	1.0055	0.87	95.8
	MNAR non-sel.	1.0055	0.21	94.4
MNAR AGHQ	MAR	1.0024	-0.10	94.6
	MNAR sel.	0.9857	-1.12	97.2
	MNAR non-sel.	1.0008	-0.26	97.0
CCA	MAR	1.0002	−0.32	93.4
	MNAR sel.	0.6981	−29.97	21.8
	MNAR non-sel.	0.5016	−50.01	1.2
MAR 2l.pmm	MAR	0.9808	−2.26	91.4
	MNAR sel.	0.6607	-33.72	18.4
	MNAR non-sel.	0.5085	-49.32	3.0

Application to NEPS data

Table: Effects on Higher Aspirations: Analyses Using Different Methods for Handling Missing Values for Maternal Education.

	CCA		MAR		MNAR	
Variable	$\hat{\beta}$	**p-value**	$\hat{\beta}$	**p-value**	$\hat{\beta}$	**p-value**
Grade in Mathematics	−0.121	0.177	−0.121	0.068	−0.124	0.054
Grade in German	−0.580	<0.001	−0.487	<0.001	−0.484	<0.001
Comp. in Mathematics: satisfactory (Ref. Cat.: poor)	0.797	0.006	0.728	0.250	0.682	0.215
Comp. in Mathematics: good (Ref. Cat.: poor)	1.300	<0.001	1.294	0.066	1.221	0.044
Comp. in Reading: satisfactory (Ref. Cat.: poor)	1.287	0.147	0.718	0.635	0.779	0.575
Comp. in Reading: good (Ref. Cat.: poor)	1.647	0.035	1.231	0.404	1.306	0.338
Sex (Ref. Cat.: male)	0.855	<0.001	0.730	<0.001	0.726	<0.001
Migration Background (Ref. Cat.: no)	0.807	<0.001	0.570	<0.001	0.686	<0.001
Maternal educ.: interm. voc. train. (Ref. Cat.: interm. gen.)	−0.012	0.954	0.098	0.566	0.164	0.278
Maternal educ.: high secondary (Ref. Cat.: interm. gen.)	0.718	0.044	0.353	0.168	0.633	0.068
Maternal educ.: tertiary (Ref. Cat.: interm. gen.)	−0.120	0.816	0.120	0.753	0.450	0.331
Variance of Random Effect on Grade Level	0.575		0.807		0.821	
N Students (in Schools)	1250 (138)		3291 (142)		3291 (142)	

Conclusion

- Novel MNAR imputation method outperforms MAR modeling strategies in MNAR settings and shows reasonable performance when the data are MAR
- Sensitivity analysis indispensable in cases where MNAR seems realistic
- Applicability to empirical data could be proven using data of the NEPS
- R package `selMNAR.MI` containing the new methods available on my GitHub site (http://github.com/AngelinaHammon/selMNAR.MI)
- Future research: Consider varying cluster conditions in simulation studies, extension to multinomial data, further MNAR models based on Pattern-mixture modeling

References

Andridge, Rebecca R and Roderick JA Little (2011). "Proxy pattern-mixture analysis for survey nonresponse". In: *Journal of Official Statistics* 27.2, pp. 153–180.

Andridge, Rebecca R et al. (2019). "Indices of non-ignorable selection bias for proportions estimated from non-probability samples". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 68.5, pp. 1465–1483.

Galimard, Jacques-Emmanuel et al. (2015). "Imputation of MNAR missing data using one-step ML selection model". In: *36th Annual Conference of the International Society for Clinical Biostatistics.*

— (2016). "A multiple imputation approach for MNAR mechanisms compatible with Heckman's model". In: *Statistics in Medicine.*

Hammon, Angelina and Sabine Zinn (2020). "Multiple imputation of binary multilevel missing not at random data". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 69.3, pp. 547–564.