Two stage multiple imputation SEM for ordinal variables

1. Polychoric Correlations

Suppose that the rows of $\mathbf{X}(n \times p)$ are n observations of p ordinal variables $x_1, x_2, \ldots, x_p$ with $m_1, m_2, \ldots, m_p$ categories, respectively. Suppose further that these p ordinal variables are the result of the discretization of the underlying p continuous standard normal variables $z_1, z_2, \ldots, z_p$ as such that $\mathbf{z} = \begin{bmatrix} z_1 & z_2 & \ldots & z_p \end{bmatrix}' \sim N(\mathbf{0}, \mathbf{P})$ and

$$\begin{cases} x_i = 1 & \text{if } \tau_{i0} < z_i \leq \tau_{i1} \\ x_i = 2 & \text{if } \tau_{i1} < z_i \leq \tau_{i2} \\ \qquad \vdots \\ x_i = m_i & \text{if } \tau_{i,m_i-1} < z_i \leq \tau_{i,m_i} \end{cases}$$

where $\mathbf{P}$ denotes the population correlation matrix of $\mathbf{z}$ and $-\infty = \tau_{i0} < \tau_{i1} < \tau_{i2} \ldots \leq \tau_{i,m_i} = \infty$ are parameters known as thresholds. The model for the univariate marginal of variable x_i is

$$\pi_{ik} = \int_{\tau_{i,k-1}}^{\tau_{i,}} \phi(u) du$$

where $\phi(.)$ denotes the probability density function of the standard normal distribution. The maximum likelihood estimator of τ_{ik} (Jöreskog, 1994) is given by

$$\widehat{\tau}_{ik} = \Phi^{-1}(p_{i1} + p_{i2} + \ldots + p_{ik})$$

where $\Phi^{-1}(.)$ denotes the inverse of the cumulative distribution function of the standard normal distribution and $p_{i1}, p_{i2}, \ldots, p_{im_i}$ denote the marginal sample proportions for x_i.

The polychoric correlation matrix, $\mathbf{R}$, is a consistent estimator of the population correlation matrix $\mathbf{P}$. The model for the bivariate marginal of variables x_i and x_j is

$$\pi_{ijkl} = \int_{\tau_{i,k-1}}^{\tau_{i,}} \int_{\tau_{i,l-1}}^{\tau_{il}} \phi_2(u, v, \rho_{ij}) du dv$$

where $\phi_2(u, v, \rho_{ij})$ denotes the probability density function of the bivariate standard normal distribution with correlation ρ_{ij}. The maximization of the bivariate likelihood function is equivalent to minimization of the discrepancy function

$$F\left(\rho_{ij}, \hat{\boldsymbol{\tau}}_i, \hat{\boldsymbol{\tau}}_j\right) = \sum_{k=1}^{m_i} \sum_{l=1}^{m_j} p_{ijkl}\left(\ln\left\{p_{ijkl}\right\} - \ln\left\{\pi_{ijkl}\right\}\right)$$

where $\hat{\boldsymbol{\tau}}_i$ and $\hat{\boldsymbol{\tau}}_j$ denote the maximum likelihood estimators of the $m_i - 1$ and $m_j - 1$ thresholds of variables x_i and x_j, respectively and p_{ijkl} is the sample proportion for $x_i = k$ and $x_j = l$. The gradient of $F\left(\cdot\right)$ (Olsson (1979)) may be expressed as

$$g\left(\rho_{ij}, \hat{\boldsymbol{\tau}}_i, \hat{\boldsymbol{\tau}}_j\right) = -\sum_{k=1}^{m_i} \sum_{l=1}^{m_j} \frac{p_{ijkl}}{\pi_{ijkl}}\left[\frac{\partial \pi_{ijkl}}{\partial \rho_{ij}}\right]$$

where (Olsson (1979))

$$\frac{\partial \pi_{ijkl}}{\partial \rho_{ij}} = \phi_2\left(\tau_{ik}, \tau_{il}\right) - \phi_2\left(\tau_{i,k-1}, \tau_{jl}\right) - \phi_2\left(\tau_{i,k}, \tau_{j,l-1}\right) + \phi_2\left(\tau_{i,k-1}, \tau_{j,l-1}\right)$$

where $\phi_2\left(\cdot\right)$ denotes the density function of the bivariate standard normal distribution with correlation ρ_{ij}. The information (Jöreskog, 1994) is given by

$$i\left(\rho_{ij}, \hat{\boldsymbol{\tau}}_i, \hat{\boldsymbol{\tau}}_j\right) = \left[\frac{\partial \pi_{ijkl}}{\partial \rho_{ij}}\right]\frac{1}{\pi_{ijkl}}\left[\frac{\partial \pi_{ijkl}}{\partial \rho_{ij}}\right]$$

The Fisher scoring algorithm is used to minimize $F\left(\cdot\right)$ with respect to ρ_{ij}. Let $\theta = \rho_{ij}$. If $\hat{\theta}^{(t)}$ denotes the tth successive approximation to $\hat{\theta}$, then the $(t+1)^{st}$ approximation is obtained from

$$\hat{\theta}^{(t+1)} = \hat{\theta}^{(t)} - \frac{g\left(\rho_{ij}, \hat{\boldsymbol{\tau}}_i, \hat{\boldsymbol{\tau}}_j\right)}{i\left(\rho_{ij}, \hat{\boldsymbol{\tau}}_i, \hat{\boldsymbol{\tau}}_j\right)}$$

Iteration is terminated when the absolute gradient value is below the tolerance limit $\varepsilon = 10^{-3}$.

The asymptotic covariance matrix, $\boldsymbol{\Upsilon}$, of the $p^* = p(p-1)/2$ polychoric correlations is a $p^*(p^* +1)/2$ matrix. A typical element of $\hat{\boldsymbol{\Upsilon}}$ (Jöreskog, 1994) may be expressed as

$$[\hat{\boldsymbol{\Upsilon}}]_{ij,rs} = \sum_{c=1}^{n} \kappa_{cijrs}\hat{\gamma}_{ijkl}\hat{\gamma}_{rsno} - \hat{\omega}_{ij}\hat{\omega}_{kl}$$

where $\kappa_{cijrs} = \dfrac{1}{n}$ if $x_{ci} = k$, $x_{cj} = l$, $x_{cr} = n$, and $x_{cs} = o$ and 0 otherwise, $\omega_{ij} = \sum_{k=1}^{m_i}\sum_{l=1}^{m_j}\gamma_{ijkl}\pi_{ijkl}$, $\omega_{rs} = \sum_{n=1}^{m_i}\sum_{o=1}^{m_j}\gamma_{rsno}\pi_{rsno}$, and γ_{ijkl} denotes a typical element of

$$\boldsymbol{\Gamma}_{ij} = \boldsymbol{\alpha}_{ij} + \mathbf{B}_i\boldsymbol{\beta}_i\mathbf{1}_j' + \mathbf{1}_i\boldsymbol{\beta}_j'\mathbf{B}_j'$$

where $\mathbf{1}_i$ denotes an $m_i \times 1$ column vector and

$$\mathbf{B}_i' = \left(\mathbf{A}_i'\mathbf{D}_{\pi_i}^{-1}\mathbf{A}_i\right)^{-1}\mathbf{A}_i'\mathbf{D}_{\pi_i}^{-1}$$

$$\mathbf{B}_j' = \left(\mathbf{A}_j'\mathbf{D}_{\pi_i}^{-1}\mathbf{A}_j\right)^{-1}\mathbf{A}_j'\mathbf{D}_{\pi_j}^{-1}$$

where $\mathbf{A}_i$ denotes the $m_i \times (m_i - 1)$ matrix given by

$$\mathbf{A}_i = \begin{bmatrix} \phi(\tau_{1k}) & 0 & \cdots & 0 \\ -\phi(\tau_{1k}) & \phi(\tau_{2k}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & -\phi(\tau_{m_i-1.k}) \end{bmatrix}$$

Typical elements of $\boldsymbol{\alpha}_{ij}$, $\boldsymbol{\beta}_i$, and $\boldsymbol{\beta}_j$ are given by

$$\alpha_{ijkl} = D^{-1} \frac{1}{\pi_{ijkl}} \frac{\partial \pi_{ijkl}}{\partial \rho_{ij}}$$

$$[\boldsymbol{\beta}_i]_k = \sum_{k=1}^{m_i} \sum_{l=1}^{m_j} \alpha_{ijkl} \left(\frac{\partial \pi_{ijkl}}{\partial \tau_{ik}} \right)$$

$$[\boldsymbol{\beta}_j]_l = \sum_{k=1}^{m_i} \sum_{l=1}^{m_j} \alpha_{ijkl} \left(\frac{\partial \pi_{ijkl}}{\partial \tau_{il}} \right)$$

where

$$D = \sum_{k=1}^{m_i} \sum_{l=1}^{m_j} \frac{1}{\pi_{ijkl}} \left(\frac{\partial \pi_{ijkl}}{\partial \rho_{ij}} \right)^2 .$$

The robust DWLS, WLS, or ULS methods can be used to fit structural equation models for ordinal variables to the polychoric correlation matrix by using the estimated asymptotic covariance matrix of the polychoric correlations (Chung and Cai (2019)).

2. Multiple Imputation
2.1 The MCMC method

Suppose now that the n observations of the p ordinal variables include missing data values with k missing data value patterns. The EM algorithm and the MCMC method for multiple imputation of incomplete data are intended for continuous variables and cannot readily be applied to ordinal variables. However, they can be applied to the underlying continuous variables $z_1, z_2, \ldots, z_p$ associated with the ordinal variables $x_1, x_2, \ldots, x_p$. Although no observations for these continuous variables are available, these variables are assumed to have a multivariate standard normal distribution with a population covariance matrix $\boldsymbol{\Sigma}$. As a result, we can simulate data from this distribution by using the polychoric correlation matrix of the complete data of the variables if the number of complete cases is large enough and use either the EM algorithm or the MCMC algorithm to impute the missing data values for the underlying continuous variables. After imputation, the estimated thresholds can be used to replace the missing data values for the corresponding ordinal variables by using the relationship between the ordinal variables, the underlying continuous variables, and the thresholds.

Suppose that the rows of $\mathbf{Z}(n \times p)$ are n observations of the p underlying continuous variables $z_1, z_2, \ldots, z_p$ simulated from the $N(\mathbf{0}, \boldsymbol{\Sigma})$ distribution and that $\mathbf{Z}_o$ denotes the observed data values that corresponds with the observed data values of $\mathbf{X}$. The EM algorithm (Dempster, Laird, and Rubin 1977) can be used to compute the maximum likelihood estimate of $\boldsymbol{\Sigma}$. The minus two observed-data log likelihood may be expressed as

$$-2\ln L(\Sigma\,|\,\mathbf{Z}_o) = \sum_{i=1}^{k} n_i \ln|\Sigma_i| + \sum_{i=1}^{k}\sum_{j=1}^{n_i} \mathbf{z}'_{oij}\Sigma_i^{-1}\mathbf{z}_{oij}$$

where n_i denotes the number of observations of missing data value pattern $i = 1,2,\cdots,k,$ Σ_i denotes the population covariance matrix for missing data value pattern i, and $\mathbf{z}_{oij}$ is the jth vector of observed values of missing data value pattern i.

The initial estimate for the M-step is the sample covariance matrix, $\mathbf{S}$, of the complete data or $\mathbf{I}_p$ if the number of complete observations is too small. In the E-step, the conditional covariance matrices of the missing variables given the observed variables for the missing data value patterns are computed and used to compute an updated estimate, $\hat{\Sigma}^{(t+1)}$ of Σ. Iteration of the consecutive M and E steps is terminated when the absolute difference between $\hat{\Sigma}^{(t+1)}$ and $\hat{\Sigma}^{(t)}$ is below the tolerance limit $\varepsilon = 10^{-5}$.

The correlation matrix of the EM estimate, $\hat{\Sigma}$, of Σ is used as the initial covariance matrix of the multivariate standard normal distribution in the first step of the Monte Carlo Markov Chain (MCMC) method. In the first step (P-step) of the MCMC method, an estimate of Σ is simulated form an inverse Wishart distribution. In the I-step, observations are simulated from the conditional standard normal distributions of the missing variables given the observed k missing data value patterns and used to replace the missing data values. The next estimate of Σ is then obtained by computing the sample correlation matrix of the completed data. The P and I steps are repeated for a fixed number of times.

Let the rows of $\mathbf{Z}_i\,(n\times p)$ contain the observed and the imputed data values for the standard normal variables $z_1, z_2, \ldots, z_p$. The observed data for the ordinal variables are obtained from the corresponding observed data values of $\mathbf{X}$. The missing data values of $\mathbf{X}$ are then replaced by the values obtained from the corresponding imputed data values of $\mathbf{Z}$ and the estimated thresholds by using the relationship between the ordinal variables, the underlying continuous variables, and the thresholds.

2.2 The FCS ordinal logistic regression method

Suppose now that the n observations of the p ordinal variables include missing data values and that a joint (multivariate) distribution of the variables exists. In this case, the Fully Conditional Specified (FCS) ordinal logistic regression method (Brand 1999; Van Buuren 2007) can be used to impute the missing data values. The FCS ordinal logistic regression method performs a fixed number of imputations to impute the missing data values. Each imputation consists of a filled-in phase and an imputation phase. In the filled-in phase, the missing data values are filled-in by using a sequence of ordinal logistic regression analyses for the p ordinal variables. These filled-in data are then used as the initial data for the imputation phase in which the missing data values are imputed by using a sequence of ordinal logistic regression analyses for the p ordinal variables. These imputed data are then used as the initial data for the next iteration of the imputation phase and a fixed number of iterations are executed for each imputation.

The filled-in stage fits the following p ordinal logistic regression models sequentially to the data, namely

$$\text{logit}(\pi_{1k}) = \alpha_{1k}$$
$$\text{logit}(\pi_{2k}) = \alpha_{2k} + \beta_{21}x_1$$
$$\text{logit}(\pi_{3k}) = \alpha_{3k} + \beta_{31}x_1 + \beta_{32}x_2$$
$$\vdots$$
$$\text{logit}(\pi_{pk}) = \alpha_{pk} + \beta_{p1}x_1 + \beta_{p2}x_2 + \cdots + \beta_{p,p-1}x_{p-1}$$

where $\pi_{ijk} = P(x_j \le k \mid x_1, x_2, \ldots, x_{j-1})$, $\quad \text{logit}(\pi_{jk}) = \ln(\pi_{jk}) - \ln(\pi_{im_j})$, and the elements of

$\gamma = \left[\alpha_{11} \ \alpha_{12} \ldots \beta_{21} \ldots \beta_{p,p-1} \right]'$ denote unknown regression weights. The first model is fitted to the complete data for x_1. The corresponding estimates are then used to simulate new parameter values from the posterior distribution of the parameters which in turn is used to fill-in the missing data values for x_1. The second model is then fitted to the complete data for x_2 and the filled-in data for x_1. The final model is fitted to the complete data for x_p and the filled-in data for $x_1, x_2, \ldots, x_{p-1}$. The filled-in data for $x_1, x_2, \ldots, x_p$ are used for the first iteration of the imputation phase. The simulation of the new parameter values from the posterior distribution of the parameters and the imputation of the missing data values for each of the p ordinal logistic regression models use the same steps as outlined next for each iteration of the imputation stage.

For each iteration of the imputation stage, the following ordinal logistic regression models are fitted sequentially either to the filled-in data or the imputed data, namely

$$\text{logit}(\pi_{ijk}) = \alpha_k + \beta_1 x_1 + \cdots + \beta_{j-1}x_{j-1} + \beta_{j+1}x_{j+1} + \cdots + \beta_p x_p$$

where $\pi_{ijk} = P(x_j \le k \mid x_1, \ \ldots, x_{j-1}, x_{j+1} \ldots, x_p)$, $\quad \text{logit}(\pi_{jk}) = \ln(\pi_{jk}) - \ln(\pi_{im_j})$, the elements of

$\gamma_j = \left[\alpha_1 \ \alpha_2 \ldots \alpha_{m_j-1} \ \beta_1 \ldots \beta_{j-1} \beta_{j+1} \ldots \beta_p \right]'$ denote $p + m_j - 1$ unknown regression weights, $j = 1, 2, \ldots, p$, and $k = 1, 2, \ldots, m_j - 1$. Let $\mathbf{V}_j$ denote the estimated covariance matrix of the estimator $\hat{\gamma}$ of γ_j.

New values for the parameters are then simulated from their posterior distribution as

$$\gamma_{jt} = \hat{\gamma}_j + \mathbf{V}'_{hj}\mathbf{z}$$

where $\mathbf{V}'_{hj}$ denotes the upper triangular matrix in the Cholesky decomposition of $\mathbf{V}_j = \mathbf{V}'_{hj}\mathbf{V}_{hj}$, and $\mathbf{z}$ is a $(p + m_j - 1) \times 1$ standard normal vector. These new parameter values are then used to compute the predicted cumulative probability $\hat{\pi}_{jk}$ for $k = 1, 2, \ldots, m_j - 1$. A random uniform variable, u, between 0 and 1 is simulated and the missing data values for x_j are imputed as 1 if $u < \hat{\pi}_{j1}$, as k if $\hat{\pi}_{j,k-1} \le u < \hat{\pi}_{jk}$, and as m_j if $u \ge \hat{\pi}_{jm_j}$.

3. Average moment matrices

Suppose that $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$ are m imputed data sets for the incomplete data matrix, $\mathbf{X}$, of the p ordinal variables $x_1, x_2, \ldots, x_p$ and that $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_m$ and $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_m$ denote the corresponding polychoric correlation matrices and

the estimated asymptotic covariance matrices of the polychoric correlations, respectively. Then, the average polychoric correlation matrix is

$$\bar{\mathbf{R}} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{R}_i$$

and the average estimated asymptotic covariance matrix is

$$\bar{\mathbf{U}} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{U}_i$$

Chung and Cai (2019) point out that $\bar{\mathbf{U}}$ only captures uncertainty based on complete data. As a result, its inverse cannot be used as a weight matrix for the robust DWLS, WLS, and ULS methods for ordinal structural equational modeling. A corrected weight matrix is obtained by correcting for the between-imputation variation in the estimated polychoric correlations and is obtained as the inverse of

$$\hat{\boldsymbol{\Upsilon}} = \bar{\mathbf{U}} + \frac{m+1}{m(m-1)} \left[\sum_{i=1}^{m} (\mathbf{r}_i - \bar{\mathbf{r}})(\mathbf{r}_i - \bar{\mathbf{r}})' \right]$$

where $\mathbf{r}$ denotes the $p \times (p-1)/2$ vector consisting of the nondiagonal and nonduplicated elements of the $p \times p$ symmetric matrix $\mathbf{R}$. $\bar{\mathbf{R}}$ and $\hat{\boldsymbol{\Upsilon}}$ can be used to fit structural equation models to the average polychoric correlation matrix with the robust DWLS, WLS, and ULS methods. The corrected robust DWLS and ULS Chi-square test statistic proposed by Chung and Cai (2019) is given by

$$T_B = (n-1)(\bar{\mathbf{r}} - \boldsymbol{\rho}(\hat{\boldsymbol{\theta}}))' \mathbf{V}(\bar{\mathbf{r}} - \boldsymbol{\rho}(\hat{\boldsymbol{\theta}}))$$

where

$$\mathbf{V} = \hat{\boldsymbol{\Upsilon}}^{-1} - \hat{\boldsymbol{\Upsilon}}^{-1} \hat{\boldsymbol{\Delta}} (\hat{\boldsymbol{\Delta}}' \hat{\boldsymbol{\Delta}})^{-1} \hat{\boldsymbol{\Delta}} \hat{\boldsymbol{\Upsilon}}^{-1}$$

where $\hat{\boldsymbol{\Delta}}$ denotes the Jacobian matrix of $\boldsymbol{\rho}(\boldsymbol{\theta})$ with respect to the unknown parameters, $\boldsymbol{\theta}$, of the structural equation model evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. The small sample adjusted T_B test statistic (Yuan and Bentler 1997) is given by

$$T_{YB} = \frac{T_B}{1 + nT_B / (n-1)}$$

References

Brand, J. P. L. (1999). *Development, Implementation, and Evaluation of Multiple Imputation Strategies for the Statistical Analysis of Incomplete Data Sets*. Unpublished Ph.D. thesis, Erasmus University.

Chung, S, and Cai, L. (2019). Alternative Multiple Imputation Inference for Categorical Structural Equation Modeling. *Multivariate Behavioral Research*, **54(3)**, 323–337.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, Series B, **39**, 1–38.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, **59(3)**, 381–389.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, **44(4)**, 443–460.

Van Buuren, S. (2007). Multiple Imputation of Discrete and Continuous Data by Fully Conditional Specification. *Statistical Methods in Medical Research*, **16**, 219–242.

Yuan, K.-H., & Bentler, P. M. (1997). Mean and covariance structure analysis: Theoretical and practical improvements. *Journal of the American Statistical Association*, **92(438)**, 767–774.