Scientific Software International

Two stage multiple imputation SEM for mixed correlations

Contents

1. Correlations

1.1 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

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

where $\mathbf{P}$ denotes the population correlation matrix of $\mathbf{z}$ and $-\infty = \tau_{i0} < \tau_{i1} < \tau_{i2} \ldots \le \tau_{i,m_i} = \infty$ 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

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

where $\mathbf{\Phi}^{-1}(.)$ 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, **R**, is a consistent estimator of the population correlation matrix **P**. The model for the bivariate marginal of variables x_i and x_j is

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

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

$$i\left(\rho_{ij}, \hat{\boldsymbol{\tau}}_i, \hat{\boldsymbol{\tau}}_j\right) = \sum_{k=1}^{m_i}\sum_{l=1}^{m_j} \frac{1}{\pi_{ijkl}}\left[\frac{\partial \pi_{ijkl}}{\partial \rho_{ij}}\right]^2$$

The 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 t^{th} 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 $\varepsilon = 10^{-3}$.

1.2 Pearson product-moment correlations

Suppose that the rows of $\mathbf{X}(n \times p)$ are n observations of p continuous variables $x_1, x_2, \ldots, x_p$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The sample covariance matrix, $\mathbf{S}$, is an unbiased estimator of $\boldsymbol{\Sigma}$ and may be expressed as

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\mathbf{x}_i - \bar{\mathbf{x}}\right)\left(\mathbf{x}_i - \bar{\mathbf{x}}\right)'$$

where $\mathbf{x}_i$ and $\bar{\mathbf{x}}$ denote observation i and the sample mean vector of $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \ldots & x_p \end{bmatrix}'$, respectively.

The correlation matrix, $\mathbf{P}$, of $x_1, x_2, \ldots, x_p$ is the covariance matrix of the standardized variables $z_1, z_2, \ldots, z_p$ where

$$\mathbf{P} = \mathbf{D}_{\boldsymbol{\sigma}}^{-1}\boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{\sigma}}^{-1}$$

and

$$z_i = \frac{x_i - \mu_i}{\sigma_i}$$

where $\mathbf{D}_{\boldsymbol{\sigma}}$ denotes a diagonal matrix with the standard deviations $\sigma_1, \sigma_2, \ldots, \sigma_p$ of $x_1, x_2, \ldots, x_p$ on the diagonal. The sample correlation matrix, $\mathbf{R}$, which contains the Pearson product-moment correlations (Pearson 1896), is an unbiased estimator of $\mathbf{P}$ and may be expressed as

$$\mathbf{R} = \mathbf{D}_{\mathbf{s}}^{-1}\mathbf{R}\mathbf{D}_{\mathbf{s}}^{-1}$$

where $\mathbf{D}_{\mathbf{s}}$ denotes a diagonal matrix with the sample standard deviations $s_1, s_2, \ldots, s_p$ of $x_1, x_2, \ldots, x_p$ on the diagonal.

1.3 Polyserial correlations

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

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

where $\mathbf{P}_o$ denotes the population correlation matrix of $\mathbf{z}$ and $-\infty = \tau_{i0} < \tau_{i1} < \tau_{i2} \ldots \le \tau_{i,m_i} = \infty$ 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

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

where $\mathbf{\Phi}^{-1}(.)$ 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.

If x_i denotes the i^{th} ordinal variable and x_j denotes the j^{th} continuous variable with mean μ_j and standard deviation σ_j and ρ_{ij} is the polyserial correlation of x_i and x_j, the corresponding bivariate log-likelihood function (Olsson, Drasgow, and Dorans 1982) is given by

$$l\left(\rho_{ij}, \hat{\boldsymbol{\tau}}_i, \hat{\mu}_j, \hat{\sigma}_j\right) = \sum_{m=1}^{n} \ln\left(\pi_{ikjm}\right) - \frac{n}{2}\left[\ln(2\pi) + \ln\left[\hat{\sigma}_j\right]\right] - \frac{1}{2}\sum_{m=1}^{n} z_{jm}^2$$

where

$$z_{jm} = \frac{x_{jm} - \hat{\mu}_j}{\hat{\sigma}_i}$$

and

$$\pi_{ijkm} = \Phi\left(\tau_{ikjm}^{*}\right) - \Phi\left(\tau_{i,k-1,jm}^{*}\right)$$

where k denotes the observed category of x_i, Φ denotes the cumulative distribution function of the standard normal distribution, and

$$\tau_{ikjm}^{*} = \frac{\hat{\tau}_{ik} - \rho_{ij} z_{jm}}{\sqrt{1-\rho_{ij}^2}}$$

The maximization of the log-likelihood function is equivalent to minimizing the following discrepancy function

$$F\left(\rho_{ij}, \hat{\boldsymbol{\tau}}_i, \hat{\mu}_j, \hat{\sigma}_j\right) = -\sum_{m=1}^{n} \ln\left(\pi_{ikjm}\right)$$

The gradient of $F(.)$ follows as

$$g\left(\rho_{ij}, \hat{\boldsymbol{\tau}}_i, \hat{\mu}_j, \hat{\sigma}_j\right) = -\sum_{m=1}^{n} \frac{1}{\pi_{ikjm}} \frac{\partial \pi_{ikjm}}{\partial \rho_{ij}}$$

where (Olsson, Drasgow, and Dorans 1982)

$$\frac{\partial \pi_{ikjm}}{\partial \rho_{ij}} = (1-\rho_{ij}^2)^{-\frac{3}{2}}\left[\phi\left(\tau_{ikjm}^{*}\right)\left(\hat{\tau}_{ik}\rho_{ij} - z_{jm}\right) - \phi\left(\tau_{i,k-1,jm}^{*}\right)\left(\hat{\tau}_{i,k-1}\rho_{ij} - z_{jm}\right)\right]$$

where ϕ denotes the probability density function of the standard normal distribution. The information follows as

$$i\left(\rho_{ij},\hat{\boldsymbol{\tau}}_i,\hat{\mu}_j,\hat{\sigma}_j\right)=\pi_{ikjm}^{-2}\left[\frac{\partial\pi_{ikjm}}{\partial\rho_{ij}}\right]^2$$

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

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

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

2. Mixed correlation and asymptotic covariance matrices

Suppose that the rows of $\mathbf{X}(n\times p)=\left[\mathbf{X}_o \quad \mathbf{X}_c\right]$ are n observations of p_o ordinal variables $x_1,x_2,\ldots,x_{p_o}$ with $m_1,m_2,\ldots,m_{p_o}$ categories, respectively and p_c continuous variables $x_1,x_2,\ldots,x_{p_c}$ as such that $p_o+p_c=p$. Let $\mathbf{R}_o(p_o\times p_o)$ denote the polychoric correlation matrix of the p_o ordinal variables, $\mathbf{R}_c(p_c\times p_c)$ denote the Pearson product-moment correlation matrix of the p_c continuous variables $x_1,x_2,\ldots,x_{p_c}$, and $\mathbf{R}_{oc}(p_o\times p_c)$ denote the polyserial correlation matrix of the ordinal and continuous variables. The correlation matrix, $\mathbf{R}$, of the ordinal and continuous variables may then be expressed as

$$\mathbf{R}=\begin{bmatrix}\mathbf{R}_o & \mathbf{R}_{oc}\\ \mathbf{R}'_{oc} & \mathbf{R}_c\end{bmatrix}$$

If F_{ij} denotes the discrepancy function which is minimized with respect to ρ_{ij} to obtain the maximum likelihood estimate of ρ_{ij}, then the asymptotic covariance matrix, $\boldsymbol{\Upsilon}$, of the polychoric, polyserial, and Pearson product-moment correlations (Muthen 1984) may be approximated by the matrix, $\mathbf{U}$, with typical element given by

$$u_{ij,kl}=n^{-1}\sum_{m=1}^{n}g_{ijm}g_{klm}$$

where g_{ijm} denotes the gradient of F_{ij} for observation m evaluated at $\rho_{ij}=r_{ij}$. If r_{ij} is a polychoric correlation, this gradient is given by

$$g_{ijm}=\frac{1}{\pi_{ijkl}}\left[\phi_2\left(\tau_{ik},\tau_{il}\right)-\phi_2\left(\tau_{i,k-1},\tau_{jl}\right)-\phi_2\left(\tau_{i,k},\tau_{j,l-1}\right)+\phi_2\left(\tau_{i,k-1},\tau_{j,l-1}\right)\right]$$

where $\phi_2(\cdot)$ denotes the density function of the bivariate standard normal distribution with correlation ρ_{ij} and k and l denote the observed category of x_i and x_j for observation m, respectively. In the case of a Pearson product-moment correlation, the gradient for observation m may be expressed as

$$g_{ijm}=\frac{z_{im}z_{jm}+\left(1-z_{im}^2-z_{jm}^2\right)r_{ij}+z_{im}z_{jm}r_{ij}^2-r_{ij}^3}{\left(1-r_{ij}^2\right)^2}$$

If r_{ij} denotes the polyserial correlation of ordinal variable x_i and continuous variable x_j, the gradient for observation m is given by

$$g_{ijm} = \frac{\phi(\tau_{ik})(\tau_{ik} r_{ij} - z_{jm}) - \phi(\tau_{i,k-1})(\tau_{i,k-1} r_{ij} - z_{jm})}{(1 - r_{ij}^2)^{3/2}}$$

where ϕ denotes the probability density function of the standard normal distribution and k denotes the observed category of x_i.

3. Multiple Imputation

3.1 The MCMC method

Suppose now that the n observations of the p_o ordinal variables include missing data values with k_o 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 $z_1, z_2, \ldots, z_{p_o}$ associated with the ordinal variables $x_1, x_2, \ldots, x_{p_o}$. 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 $\boldsymbol{\Sigma}_o$. 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 $\mathbf{Z}(n \times p_o)$ are n observations of the p_o underlying continuous variables $z_1, z_2, \ldots, z_{p_o}$ simulated from the $N(\mathbf{0}, \boldsymbol{\Sigma}_o)$ distribution and that $\mathbf{Z}_o$ denotes the observed data values that corresponds with the observed data values of $\mathbf{X}_o$. The EM algorithm (Dempster, Laird, and Rubin 1977) can be used to compute the maximum likelihood estimate of $\boldsymbol{\Sigma}_o$. The minus two observed-data log likelihood may be expressed as

$$-2 \ln L(\boldsymbol{\Sigma}_o \mid \mathbf{Z}_o) = \sum_{i=1}^{k_o} n_i \ln |\boldsymbol{\Sigma}_{oi}| + \sum_{i=1}^{k_o} \sum_{j=1}^{n_i} \mathbf{z}'_{oij} \boldsymbol{\Sigma}_{oi}^{-1} \mathbf{z}_{oij}$$

where n_i denotes the number of observations of missing data value pattern $i = 1,2,\cdots,k_o$, $\boldsymbol{\Sigma}_{oi}$ denotes the population covariance matrix for missing data value pattern i, and $\mathbf{z}_{oij}$ is the j^{th} vector of observed values of missing data value pattern i.

The initial estimate for the M-step is the sample covariance matrix, $\mathbf{S}_{p_o}$, of the complete ordinal data or $\mathbf{I}_{p_o}$ 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, $\hat{\boldsymbol{\Sigma}}_o^{(t+1)}$ of $\boldsymbol{\Sigma}_o$. Iteration of the consecutive M and E steps is terminated when the absolute difference between $\hat{\boldsymbol{\Sigma}}_o^{(t+1)}$ and $\hat{\boldsymbol{\Sigma}}_o^{(t)}$ is below the tolerance limit $\varepsilon = 10^{-5}$.

The correlation matrix of the EM estimate, $\hat{\boldsymbol{\Sigma}}_o$, of $\boldsymbol{\Sigma}_o$ 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 $\boldsymbol{\Sigma}_o$ 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 $\boldsymbol{\Sigma}_o$ 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 $z_1, z_2, \ldots, z_{p_o}$. The observed data for the ordinal variables are obtained from the corresponding observed data values of $\mathbf{X}_o$. The missing data values of $\mathbf{X}_o$ 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.

Suppose further that the n observations of the p_c continuous variables include missing data values with k_c missing data value patterns and that the joint distribution of the variables is a multivariate normal distribution with mean vector $\boldsymbol{\mu}_c$ and covariance matrix $\boldsymbol{\Sigma}_c$. The EM algorithm and the MCMC method for multiple imputation of incomplete data can be used to impute the missing data values of the continuous variables.

Suppose that $\mathbf{X}_{co}$ denote the observed data values. The EM algorithm (Dempster, Laird, and Rubin 1977) can be used to compute the maximum likelihood estimate of $\boldsymbol{\Sigma}_c$. The minus two observed-data log likelihood may be expressed as

$$-2 \ln L(\boldsymbol{\Sigma}_c \mid \mathbf{X}_{co}) = \sum_{i=1}^{k} n_i \ln |\boldsymbol{\Sigma}_{ci}| + \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(\mathbf{x}_{coij} - \boldsymbol{\mu}_{ci}\right)' \boldsymbol{\Sigma}_{ci}^{-1} \left(\mathbf{x}_{coij} - \boldsymbol{\mu}_{ci}\right)$$

where n_i denotes the number of observations of missing data value pattern $i = 1, 2, \ldots, k_c$, $\boldsymbol{\Sigma}_{ci}$ denotes the population covariance matrix of missing data value pattern i, $\boldsymbol{\mu}_{ci}$ denotes the mean vector of missing data value pattern i, and $\mathbf{x}_{coij}$ is the j^{th} vector of observed values of missing data value pattern i.

The initial estimate for the M-step is the sample covariance matrix, $\mathbf{S}_{p_c}$, of the complete data or $\mathbf{I}_{p_c}$ 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 of the missing data value patterns are computed and used to compute an updated estimate $\hat{\boldsymbol{\Sigma}}_c^{(t+1)}$ of $\boldsymbol{\Sigma}_c$. Iteration of the consecutive M and E steps is terminated when the absolute difference between $\hat{\boldsymbol{\Sigma}}_c^{(t+1)}$ and $\hat{\boldsymbol{\Sigma}}_o^{(t)}$ is below the tolerance limit $\varepsilon = 10^{-5}$.

The EM estimate, $\hat{\boldsymbol{\Sigma}}_c$, of $\boldsymbol{\Sigma}_c$ is used as the initial covariance matrix of the multivariate 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 $\boldsymbol{\Sigma}_c$ is simulated from an inverse Wishart distribution. In the I-step, observations are simulated from the conditional normal distributions of the missing variables given the observed k_c missing data value patterns and used to replace the missing data

values. The next estimate of $\boldsymbol{\Sigma}_c$ is then obtained by computing the sample covariance matrix of the completed data. The P and I steps are repeated for a fixed number of times.

3.2 The FCS ordinal logistic regression method

Suppose that the n observations of the p_o 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_o 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_o 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_o ordinal logistic regression models sequentially to the data, namely

$$\begin{aligned}
&\text{logit}(\pi_{1k}) = \alpha_{1k} \\
&\text{logit}(\pi_{2k}) = \alpha_{2k} + \beta_{21}x_1 \\
&\text{logit}(\pi_{3k}) = \alpha_{3k} + \beta_{31}x_1 + \beta_{32}x_2 \\
&\vdots \\
&\text{logit}(\pi_{p_ok}) = \alpha_{p_ok} + \beta_{p_o1}x_1 + \beta_{p_o2}x_2 + \cdots + \beta_{p_o,p_o-1}x_{p_o-1}
\end{aligned}$$

where $\pi_{ijk} = P(x_j \le k \mid x_1, x_2, \ldots, x_{j-1})$, $\text{logit}(\pi_{jk}) = \ln(\pi_{jk}) - \ln(\pi_{im_j})$, and the elements of $\boldsymbol{\gamma} = \left[\alpha_{11}\ \alpha_{12} \ldots \beta_{21} \ldots \beta_{p_o,p_o-1}\right]'$ 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_o} and the filled-in data for $x_1, x_2, \ldots, x_{p_o-1}$. The filled-in data for $x_1, x_2, \ldots, x_{p_o}$ 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

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

where $\pi_{ijk} = P(x_j \le k \mid x_1, \ldots, x_{j-1}, x_{j+1} \ldots, x_{p_o})$, $\text{logit}(\pi_{jk}) = \ln(\pi_{jk}) - \ln(\pi_{im_j})$, the elements of $\boldsymbol{\gamma}_j = \left[\alpha_1\ \alpha_2 \ldots \alpha_{m_i-1}\ \beta_1 \ldots \beta_{j-1}\beta_{j+1} \ldots \beta_{p_o}\right]'$ denote $p_o + m_j - 1$ unknown regression weights, $j = 1, 2, ..., p_o$, and $k = 1, 2, ..., m_j - 1$. Let $\mathbf{V}_j$ denote the estimated covariance matrix of the estimator $\hat{\boldsymbol{\gamma}}_j$ of $\boldsymbol{\gamma}_j$.

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

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

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 $(p_o + m_j - 1)\times 1$ standard normal vector. These new parameter values are then used to compute the predicted cumulative probability $\hat{\pi}_{jk}$ for $k = 1, 2, ..., m_j - 1$. 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 $u \ge \hat{\pi}_{jm_j}$.

3.3 The FCS regression method

Suppose now that the n observations of the p_c continuous variables include missing data values and that a joint (multivariate) distribution of the variables exists. In this case, the Fully Conditional Specified (FCS) regression method (Brand 1999; Van Buuren 2007) can be used to impute the missing data values. The FCS 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 regression analyses for the p_c continuous 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 regression analyses for the p_c continuous 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_c regression models sequentially to the data, namely

$$\begin{aligned}
x_1 &= \beta_{01} + e_1 \\
x_2 &= \beta_{02} + \beta_{21}x_1 + e_2 \\
x_3 &= \beta_{03} + \beta_{31}x_1 + \beta_{32}x_2 + e_3 \\
&\vdots \\
x_{p_c} &= \beta_{0p_c} + \beta_{p_c1}x_1 + \beta_{p_c2}x_2 + \cdots + \beta_{p_c,p_c-1}x_{p_c-1} + e_{p_c}
\end{aligned}$$

where the elements of $\boldsymbol{\beta} = \left[\beta_{01}\ \beta_{02}\ \cdots\ \beta_{p_c,p_c-1}\right]'$ denote unknown regression weights and $e_1, e_2, \ldots, e_{p_c}$ are p_c error variables. 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 distributions 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_c} and the filled-in data for $x_1, x_2, \ldots, x_{p_c-1}$. The filled-in data for $x_1, x_2, \ldots, x_{p_c}$ are used for the first iteration of the imputation phase. The simulation of the new parameter values from the posterior distributions of the parameters and the imputation of the missing data values for each of the p_c 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 regression models are fitted sequentially either to the filled-in data or the imputed data, namely

$$x_j = \beta_0 + \beta_1 x_1 + \cdots + \beta_{j-1} x_{j-1} + \beta_{j+1} x_{j+1} + \cdots + \beta_{p_c} x_{p_c} + e_j$$

where $j = 1, 2, \ldots, p_c$, the elements of $\boldsymbol{\beta}_j = \left[\beta_0 \ \beta_1 \ldots \beta_{j-1} \beta_{j+1} \ldots \beta_{p_c} \right]'$ denote p_c unknown regression weights, and e_j denotes an error variable with variance σ_j^2. The estimated covariance matrix of the estimator $\hat{\boldsymbol{\beta}}_j$ of $\boldsymbol{\beta}_j$ may be expressed as

$$\sigma_j^2 \mathbf{V}_j = \sigma_j^2 \left(\mathbf{X}'_{c(j)} \mathbf{X}_{c(j)} \right)^{-1}$$

where $\mathbf{X}_{c(j)}$ denotes rows $1, 2, \ldots, j-1, j, \ldots, p_c$ of the filled-in or imputed data. New values for the parameters are then simulated from their posterior distributions as

$$\boldsymbol{\beta}_{jt} = \hat{\boldsymbol{\beta}}_j + \sigma_{tj}^2 \mathbf{V}_{hj}^{'} \mathbf{z}$$
$$\sigma_{tj}^2 = \frac{\hat{\sigma}_j^2 \left(n_j - p_c \right)}{c}$$

where $\mathbf{V}_{hj}$ denotes the upper triangular matrix in the Cholesky decomposition of $\mathbf{V}_j = \mathbf{V}'_{hj} \mathbf{V}_{hj}$, $\mathbf{z}$ denotes a $p_c \times 1$ standard normal vector, and c is a Chi-square variable with $n_j - p_c$ degrees of freedom. The missing data values are then imputed as

$$x_{cijm} = \boldsymbol{\beta}'_{jt} \mathbf{x}_{ci(j)} + \sigma_{ij} z$$

where x_{cijm} denotes a missing data value in row i and column j of $\mathbf{X}_c$, $\mathbf{x}_{ci(j)}$ denotes row i of $\mathbf{X}_{c(j)}$, and z is a standard normal variable.

4. Average moment matrices

Suppose that $\mathbf{X}_{1i}, \mathbf{X}_{2i}, \ldots, \mathbf{X}_{mi}$ are m imputed data sets for the incomplete data matrix, $\mathbf{X}$, of the of p_o ordinal variables $x_1, x_2, \ldots, x_{p_o}$ and the p_c continuous variables $x_1, x_2, \ldots, x_{p_c}$ 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 mixed correlation matrices and the estimated asymptotic covariance matrices of the mixed correlations, respectively. Then, the average mixes 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 structural equational modeling. A corrected weight matrix is obtained by correcting for the between-imputation variation in the estimated mixed 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 mixed 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.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, **49(1)**, 115–132.

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

Olsson, U., Drasgrow, F., and Dorans, N.J. (1992). The polyserial correlation coefficient. *Psychometrika*, **47(3)**, 337–347.

Pearson, K. (1896), Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity and Panmixia, *Philosophical Transactions of the Royal Society of London*, **187**, 253-318.

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.