Supplement B: MCMC Sampling Steps and Distributions for Two-Level Imputation

This document contains the supplemental online material from Enders, Keller, and Levy (2017) paper in *Psychological Methods*, the full citation for which is as follows.

Enders, C.K., Keller, B.T., & Levy, R. (2017, Advanced Online Publication). A fully conditional specification approach to multilevel imputation of categorical and continuous variables. *Psychological Methods*.

The document gives technical details of the full conditional distributions used to draw regression coefficients, random effects, and covariance matrices (or variance estimates). Additional details are widely available in the literature, as these distributions largely borrow from established Bayesian estimation procedures for multilevel models (Browne & Draper, 2000; Cowles, 1996; Gelman et al., 2014; Goldstein et al., 2009; Kasim & Raudenbush, 1998; Schafer, 1997; Schafer & Yucel, 2002; Sinharay et al., 2001; van Buuren, 2012; Yucel, 2008). For the remainder of the document, we abandon the previous scalar notation in favor of a more succinct matrix representation of the multilevel model

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\varepsilon}_j \tag{1}$$

where $\mathbf{y}_j$ is the vector of outcome scores for cluster j, $\mathbf{X}_j$ is the corresponding matrix of predictor variables (level-1 or level-2), including a unit vector for the intercept, $\mathbf{Z}_j$ is a subset of the level-1 variables in $\mathbf{X}_j$ that have a random influence on the outcome (including a unit vector for the intercept), $\mathbf{u}_j$ is the column vector of level-2 residuals for cluster j, and $\boldsymbol{\varepsilon}_j$ is a vector of within-cluster residuals. In the context of FCS, $\mathbf{y}$ is an incomplete variable that is the target of imputation at a particular step, and $\mathbf{X}$ and $\mathbf{Z}$ contain complete and previously imputed variables. Level-2 imputation applies a single-level regression model to a cluster-level data set with *J* records. In matrix format, the model is as follows.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{2}$$

Two algorithmic options are available for the Bayesian estimation sequence that provides the necessary parameter estimates for imputation. Following derivations from Rubin (1987, pp. 162-166), the MICE computational option is consistent with the original formulation of fully conditional specification (van Buuren, 2012; van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006), whereby only those cases with observed data on the variable to be imputed are used to estimate the imputation model parameters. In contrast, the GIBBS option uses a conventional Gibbs sampler that draws parameters from a distribution that conditions on the observed and imputed data. The Gibbs sampler for missing data imputation is described in various Bayesian analysis texts (Jackman, 2009; Lynch, 2007).

Level-1 Gibbs Sampler Steps for Continuous Variables

Step 1: Draw regression coefficients from a multivariate normal distribution, conditional on the current random effects, parameter estimates, and imputations.

$$\boldsymbol{\beta} \sim \text{MVN}\left(\hat{\boldsymbol{\beta}}, \Sigma_{\hat{\boldsymbol{\beta}}}\right)$$

$$\hat{\boldsymbol{\beta}} = \left(\sum_{j=1}^{J} \mathbf{X}_j^T \mathbf{X}_j\right)^{-1} \left(\sum_{j=1}^{J} \mathbf{X}_j^T \left(\mathbf{y}_j - \mathbf{Z}_j \mathbf{u}_j\right)\right) \tag{3}$$

$$\Sigma_{\hat{\boldsymbol{\beta}}} = \left(\sum_{j=1}^{J} \mathbf{X}_j^T \mathbf{X}_j \sigma_{\varepsilon_j}^{-2}\right)^{-1}$$

We use a *j* subscript on the within-cluster residual variance to allow for the possibility of heterogeneous within-cluster variances (discussed below), noting that all $\sigma_{\varepsilon_j}^2$ are the same in the homogeneous case, which is the default in Blimp.

Step 2: Draw cluster-specific random effects from a multivariate normal distribution, conditional on the current parameter estimates and imputations.

$$\mathbf{u}_j \sim \text{MVN}\left(\hat{\mathbf{u}}_j, \mathbf{V}_j\right)$$

$$\mathbf{V}_j = \left(\boldsymbol{\Sigma}_u^{-1} + \frac{\mathbf{Z}_j^T \mathbf{Z}_j}{\sigma_{\varepsilon_j}^2}\right)^{-1} \tag{4}$$

$$\hat{\mathbf{u}}_j = \mathbf{V}_j \left(\sigma_{\varepsilon_j}^{-2} \mathbf{Z}_j^T\right)\left(\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}\right)$$

Step 3: For a homogeneous within-cluster variance (the HOV keyword of the OPTION command, which is the default), draw a residual variance from an inverse Gamma distribution, conditional on the current parameter estimates, level-2 residuals, and imputations.

$$
\begin{aligned}
&\sigma_\varepsilon^2 \sim \mathrm{IG}(a,b) \\
&\boldsymbol{\varepsilon}_j = \mathbf{y}_j - \left(\mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{u}_j\right) \\
&a = \frac{\upsilon}{2} \quad b = \frac{S}{2} \\
&S = \sum_{j=1}^{J} \boldsymbol{\varepsilon}_j^T \boldsymbol{\varepsilon}_j + S_p \\
&\upsilon = N + \upsilon_p
\end{aligned} \tag{5}
$$

The sum of squares (scale) value S is the sum of a component based on computed level-1 residuals and the prior distribution's sum of squares, S_p. Similarly, the degrees of freedom value is a sum based on the data and the degrees of freedom for the prior, υ_p. The Blimp application offers two common sets of hyperparameters for the prior distribution: $S_p = 0$ and $\upsilon_p = -2$ (the PRIOR2 keyword of the OPTIONS command), and $S_p = 1$ and $\upsilon_p = 2$ (the PRIOR1 keyword of the OPTIONS command, the default). For a heteroscedastic within-cluster variance (the HEV keyword of the OPTIONS command), we implement the procedure described in Kasim and Raudenbush (1998).

Step 4: Draw the level-2 covariance matrix from an inverse Wishart distribution, conditional on the current parameter estimates, level-2 residuals, and imputations.

$$
\begin{aligned}
&\boldsymbol{\Sigma}_u = \mathrm{IW}(\mathbf{S}, \upsilon) \\
&\mathbf{S} = \sum_{j=1}^{J} \mathbf{u}_j^T \mathbf{u}_j + \mathbf{S}_p \\
&\upsilon = J + \upsilon_p
\end{aligned} \tag{6}
$$

The scale (sum of squares and cross-products) matrix **S** is the sum of a component based on the level-2 residuals from Step 2 and the prior distribution's scale matrix, $\mathbf{S}_p$. Similarly, the degrees of freedom value is a sum based on the data and the degrees of freedom for the prior, υ_p. Blimp offers two common sets of hyperparameters for the prior distribution: $\mathbf{S}_p = \mathbf{0}$ and $\upsilon_p = -p - 1$ (the PRIOR2 keyword of the OPTIONS command), and $\mathbf{S}_p = \mathbf{I}$ and $\upsilon_p = p + 1$ (the PRIOR1 keyword of the OPTIONS command), where p is the number of random effects. Blimp uses the PRIOR1 hyperparameters as the default because our simulations suggest that this option gives better performance when the number of clusters is small.

Step 5: Draw the imputation for case i from a univariate normal distribution, conditional on the current parameter estimates, level-2 residual terms, and previously.

$$Y_{ij(\text{mis})} \sim N\left(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{u}_j, \sigma^2_{\varepsilon_j}\right) \tag{7}$$

Level-1 Gibbs Sampler Steps for Categorical Variables

The sampling steps for categorical variables are expressed symbolically in Equation (9). The initial sampling step that draws threshold parameters, conditional on the underlying latent scores and current parameter values, applies only to ordinal variables with $K > 2$ categories. Albert and Chib (1993) described an approach for updating thresholds, but this procedure equation reference goes hereconverges very slowly. Instead, Blimp implements the procedure described by Cowles (1996). Cowles' procedure uses a Metropolis-Hastings procedure within the Gibbs sampler to draw each threshold from a normal proposal distribution, and it accepts the threshold draws at some prespecified probability. In the interest of space, we refer readers to Cowles (1996) for details on sampling threshold parameters, as the procedure is rather involved

The second step draws latent variable scores for the complete cases. For ordinal variables, latent values are drawn from a truncated normal distribution. For nominal variables, latent scores are drawn that conform to the necessary rank and magnitude conditions given in Equation (13). Both situations are described in the body of the manuscript, so we do not repeat that information here. The sampling steps for $\boldsymbol{\beta}$, $\mathbf{u}$, and $\boldsymbol{\Sigma}_u$ are identical to those in Equations (3), (4) and (6), except that $\mathbf{y}_j^*$ (the vector of latent variable scores in cluster *j*, comprised of for the complete and incomplete cases) replaces $\mathbf{y}_j$ in the equations. For nominal variables, these sampling steps are repeated for each of the $K - 1$ latent variable difference scores, whereas they are performed only once for ordinal variables. After drawing parameter values and level-2 residual terms, latent variable imputations for the incomplete cases are drawn from an unrestricted normal distribution, as described in text. The final step converts the latent imputes to discrete values using the functions from Equations (8) or (13), depending on whether a variable is ordinal or nominal, respectively.

Level-2 Gibbs Sampler Steps for Continuous Variables

After completing a single iteration of level-1 imputation, Blimp aggregates all level-1 variables, creating a *J*-record data set where each row contains the cluster means of level-1 variable (complete and imputed) and level-2 scores for cluster *j*. The program then applies single-level FCS to the incomplete level-2 variables. The remainder of the document describes the sampling steps for the single-level regression model in Equation (2). To reiterate, we use $\mathbf{u}$ to denote the residuals in this single-level model because these terms reflect between-cluster variation.

Step 1: Draw regression coefficients from a multivariate normal distribution, conditional on the current parameter values and imputations.

$$\begin{aligned} \boldsymbol{\beta} &\sim \mathrm{MVN}\left(\hat{\boldsymbol{\beta}}, \Sigma_{\hat{\boldsymbol{\beta}}}\right) \\ \hat{\boldsymbol{\beta}} &= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \\ \Sigma_{\hat{\boldsymbol{\beta}}} &= \sigma_\varepsilon^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1} \end{aligned} \tag{8}$$

Step 2: Draw a residual variance from an inverse Gamma distribution, conditional on the current parameter estimates and imputations.

$$\begin{aligned} \sigma_u^2 &\sim \mathrm{IG}(a,b) \\ \mathbf{u} &= \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \\ a = \frac{\upsilon}{2} &\quad b = \frac{S}{2} \\ S &= \mathbf{u}^T\mathbf{u} \\ \upsilon &= N \end{aligned} \tag{9}$$

The sum of squares and degrees of freedom values follow from adopting a standard non-informative prior from Bayesian linear regression (Lynch, 2007, p. 170).

Step 3: Draw an imputation for cluster *j* from a univariate normal distribution, conditional on the current parameter values and data.

$$Y_{j(\mathrm{mis})} \sim \mathrm{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma_u^2\right) \tag{10}$$

Level-2 Gibbs Sampler Steps for Categorical Variables

The level-2 Gibbs steps for categorical variables are as follows. First, draw threshold parameters for ordinal variables with $K > 2$ response options. This step follows Cowles (1996), as described previously. Second, draw latent variable scores for the complete cases. For ordinal variables, latent values are drawn from a truncated normal distribution. For nominal variables, latent scores are drawn that conform to the necessary rank and magnitude conditions given in Equation (13). Third, draw regression coefficients from the distribution in Equation (8), where $\mathbf{y}^*$ (the vector of latent scores for the full sample) replaces $\mathbf{y}$. For nominal variables, this step is repeated for each of the K - 1 latent difference scores. Fourth, draw latent imputations for the incomplete cases from an unrestricted normal distribution, as described in the manuscript. Finally, convert the latent imputes to discrete values using the functions in Equations (8) or (13).