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

This document gives technical details of the full conditional distributions used to draw regression coefficients, random effects, and covariance matrices (or variance estimates) for the two-level imputation scheme from the Enders, Keller, and Levy (2018) paper in *Psychological Methods*. Additional estimation 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 scalar notation from the paper 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. Assuming a uniform prior, the full conditional distribution is as follows.

$$\begin{gathered} \boldsymbol{\beta} \sim \text{MVN}\left(\hat{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}\right) \\ \hat{\boldsymbol{\beta}} = \left(\sum_{j=1}^{J} \mathbf{X}_j' \mathbf{X}_j\right)^{-1} \sum_{j=1}^{J} \mathbf{X}_j' (\mathbf{y}_j - \mathbf{Z}_j \mathbf{u}_j) \\ \boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} = \sigma^2_{\varepsilon_j} \left(\sum_{j=1}^{J} \mathbf{X}_j' \mathbf{X}_j\right)^{-1} \end{gathered} \tag{3}$$

We use a j subscript on the within-cluster residual variance $\sigma^2_{\varepsilon_j}$ to allow for the possibility of heterogeneous within-cluster variances (discussed below), noting that all $\sigma^2_{\varepsilon_j}$ 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.

$$\begin{gathered} \mathbf{u}_j \sim \text{MVN}\left(\hat{\mathbf{u}}_j, \mathbf{V}_{\mathbf{u}_j}\right) \\ \mathbf{V}_{\mathbf{u}_j} = \left(\sigma^{-2}_{\varepsilon_j}\ \mathbf{Z}_j' \mathbf{Z}_j + \boldsymbol{\Sigma}_u^{-1}\right)^{-1} \\ \hat{\mathbf{u}}_j = \sigma^{-2}_{\varepsilon_j}\ \mathbf{V}_{\mathbf{u}_j} \mathbf{Z}_j' (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}) \end{gathered} \tag{4}$$

Step 3: We define $1/\sigma^2_{\varepsilon}$ as a gamma random variable. Then, for a homogeneous within-cluster variance (the HOV keyword of the OPTION command, the default), draw the reciprocal of the residual variance from a gamma distribution, conditional on the current parameter estimates, level-2 residuals, and imputations.

$$1/\sigma^2_\varepsilon \sim \text{gamma}\left(\frac{N + df_p}{2}, \frac{S + S_p}{2}\right)$$
$$S = \sum_{j=1}^{J} \boldsymbol{\varepsilon}'_j \boldsymbol{\varepsilon}_j \tag{5}$$
$$\boldsymbol{\varepsilon}_j = \mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta} - \mathbf{Z}_j\mathbf{u}_j$$

The df_p and S_p terms are the prior distribution's hyperparameters. The default setting (the PRIOR1 keyword of the OPTIONS command) specifies $S_p = 1$ and $df_p = 2$ (i.e., a gamma(1,.5) prior). Specifying the PRIOR2 keyword of the OPTIONS command sets $S_p = 0$ and $df_p = -2$, which is equivalent to a uniform prior on the variance. Finally, the PRIOR3 keyword of the OPTIONS command sets $S_p = 0$ and $df_p = 0$, which gives a Jeffreys prior. 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: We define the inverse of the covariance matrix (i.e., the precision matrix, $\boldsymbol{\Sigma}_u^{-1}$) as a Wishart random variable. The level-2 precision matrix is sampled from a Wishart distribution, conditional on the current parameter estimates, level-2 residuals, and imputations.

$$\boldsymbol{\Sigma}_\mathbf{u}^{-1} \sim \text{W}((\mathbf{S} + \mathbf{S}_p^{-1})^{-1}, J + df_p) \tag{6}$$
$$\mathbf{S} = \mathbf{u}'_j\mathbf{u}_j$$

The $\mathbf{S}_p$ can be viewed as the *inverse* of the prior sums of squares matrix based on df_p degrees of freedom (i.e., prior observations). As such, $\mathbf{S} + \mathbf{S}_p^{-1}$ is a sums of squares and cross products matrix based on $J + df_p$ observations. The default setting (the PRIOR1 keyword of the OPTIONS command) specifies $\mathbf{S}_p^{-1} = \mathbf{I}$ and $df_p = p + 1$, where p is the dimension of $\boldsymbol{\Sigma}_u$. This prior corresponds to marginal uniform priors between -1 and 1 for all correlations and a marginal inverse gamma prior IG(1,.5) for variance elements. Specifying the PRIOR2 keyword of the OPTIONS command sets $\mathbf{S}_p^{-1} = 0$ and $df_p = -p - 1$, which is equivalent to a uniform prior on the elements in $\boldsymbol{\Sigma}_\mathbf{u}$. Finally, the PRIOR3 keyword sets $\mathbf{S}_p^{-1} = 0$ and $J_p = 0$.

For random intercept models with a single level-2 variance component, we define we define the reciprocal of the variance, $1/\sigma^2_u$, as a gamma random variable.

$$1/\sigma_u^2 \sim \text{gamma}\left(\frac{N + df_p}{2}, \frac{S + S_p}{2}\right) \tag{7}$$
$$S = \sum_{j=1}^{J} u_j^2$$

The priors are the univariate analogs of those given for the Wishart. The default setting (the PRIOR1 keyword of the OPTIONS command) specifies $S_p = 1$ and $df_p = 2$ (i.e., a gamma(1.,.5) prior). Specifying the PRIOR2 keyword of the OPTIONS command sets $S_p = 0$ and $df_p = -2$, which is equivalent to a uniform prior on the variance. Finally, the PRIOR3 keyword of the OPTIONS command sets $S_p = 0$ and $df_p = 0$, which gives a Jeffreys prior.

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

$$y_{ij(mis)} \sim \text{N}(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{u}_j, \sigma_\varepsilon^2) \tag{8}$$

Level-1 Gibbs Sampler Steps for Categorical Variables

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 converges 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 the manuscript (e.g., the discrete response has the largest latent score). Both situations are described in the body of the manuscript, so we do not repeat that information here. The sampling steps for $\boldsymbol{\beta}$, $\mathbf{u}_j$, and $\boldsymbol{\Sigma}_\mathbf{u}$ are the same as those in Equations 3, 4, and 6, and no sampling step is needed for σ_ε^2 because this term is fixed at unity. The primary computational difference is that $\mathbf{y}_j^*$ (a vector of latent variable scores for cluster j) replaces $\mathbf{y}_j$ in all expressions. 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 described in the paper.

Level-2 Gibbs Sampler Steps for Continuous Variables

After completing a single iteration of level-1 imputation, Blimp creates a level-2 data set with J records, such that each row contains level-2 scores and the manifest-variable cluster means of the level-1 variables (i.e., the arithmetic average of scores for cluster j). The program then applies single-level imputation scheme to the incomplete level-2 variables. The remainder of the document describes the sampling steps for the single-level regression model in Equation 2. Note that 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. Assuming a uniform prior, the full conditional distribution is as follows.

$$\begin{aligned} \boldsymbol{\beta} &\sim \text{MVN}\left(\hat{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}\right) \\ \hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ \boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} &= \sigma_u^2 \ (\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \tag{8}$$

Step 2: We define $1/\sigma_u^2$ as a gamma random variable. Assuming a Jeffreys prior, the full conditional distribution is.

$$1/\sigma_u^2 \sim \text{gamma}\left(\frac{J}{2}, \frac{S}{2}\right)$$

$$S = \sum_{j=1}^{J} u_j^2 \tag{9}$$
$$u_j = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$$

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

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

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 the paper (e.g., the discrete response has the largest latent score). Third, draw regression coefficients from the distribution in Equation 8 where $\mathbf{y}^*$ (the vector of latent variable scores) 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 given in the paper.