Performance of Sequential Imputation Method in Multilevel Applications

Enxu Zhao[1], Recai M. Yucel[2]
[1]New York State Department of Health, 800 N. Pearl St., Albany, NY 12204
[2]University at Albany, SUNY, 1 University Place, Rensselaer, NY 12144

Abstract
In missing-data problems complicated not only by the underlying data-generating mechanism (e.g. multi-stage surveys) but also skip patterns, censored items and a diverse set of variable types, a variable-by-variable imputation strategy has been increasingly popular among practitioners. While this approach offers certain advantages, it does not necessarily follow a conventional joint modeling approach. Our work assesses the performance and compatibility of sequential approach to a joint modeling approach in multilevel settings. This sequential approach uses a set of hierarchical regression models each of which follows the appropriate format for the underlying variable subject to missingness. Computational techniques used to approximate the conditional posterior predictive distributions are based on Markov Chain Monte Carlo and/or numerical integration techniques to overcome the problem of intractability. We present a simulation study assessing the performance of the sequential approach. In most realistic applications, our simulations suggest that the sequential method lead to well-calibrated estimates and in some settings the performance is even better than the more conventional methods with well-defined joint model in some scenarios.

Key Words: missing-data, multiple imputation, sequential imputation, hierarchical models, random-effects, clustered data

1. Introduction

1.1 Problems
Missing data is one of the most pervasive problems in survey research. Many survey data have arbitrary structures of missingness due to item nonresponse. Further, they usually consist of variables measured on diverse set of scales (e.g. continuous, binary or count). Further structures that need to be incorporated in any missing-data technique include sampling design (e.g. clustering or multi-stage selection), skip patterns and censored items.

Figure 1 depicts data structure we assume for the methods of this paper. The survey sample design includes clusters reflecting multi-stage selection. Y-variables can be of continuous, categorical and/or count nature, which are allowed to be missing (denoted by "?") under a missing-at-random (MAR) mechanism. Such missingness is due to item non-response and the skip patterns (denoted by boxes) further complicate the data structure. When devising a methodology for dealing with such incomplete data, one must consider these special structures and relationships among variables within a cluster for each unit, relationship between incompletely observed variables and any additional covariates, and variation in any variable within a cluster across the units.

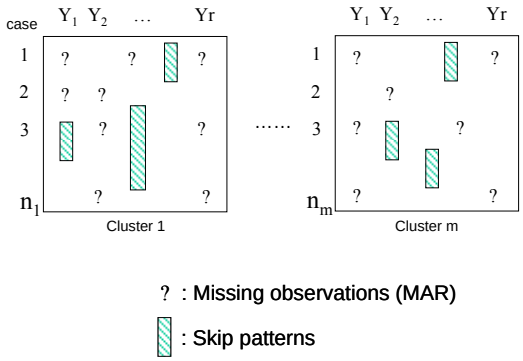

Figure 1: Typical data structure for complex surveys

Traditional imputation methods such as matching, regression imputation may work well in simple missing data problems where missingness is confined to a single variable or single item, or where the missing data follow a monotone pattern. However, the implementation of these methods in multilevel settings with the structured stated above often leads to serious inferential problems such as bias and failure to account for uncertainty due to missing data.

More principled methods such as maximum likelihood (ML) estimation have also been popular using expectation-maximization (EM) algorithm (Dempster, Laird and Rubin, 1977) or its variations (Little and Rubin, 2002), which incorporate uncertainty due to missing data. Unfortunately, there are several obstacles preventing practitioners from using EM-based ML inferences. First is the complexity of the models and potential for computational failure in structures considered here. Second, such methods tend to be problem-specific and thus lack the attractiveness of multiple imputation which is the ability to serve multiple analyses.

1.2 Multiple Imputation

Multiple imputation (MI) (Rubin, 1987) is a simulation-based approach to missing data. MI retains much of the attractiveness of single imputation from a conditional distribution but solves the problem of understating uncertainty. In MI, each missing value is replaced by a set of m>1 plausible values drawn from their predictive distribution (see Figure 2).

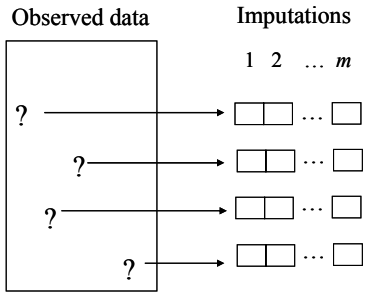

Figure 2: Multiple Imputation

Each of the m data sets is analyzed in the same fashion by a complete-data method, and the results are then combined following Rubin's rules(Rubin, 1987). MI can be highly efficient even for small m, and normally the limited computational power is often not

important as one good set of imputations can be used in many analyses. Under the model-based MI, one must impose a probability model on the complete data (observed and missing values). This imputation model is then used to form posterior predictive distribution of missing data where the imputations are sampled from their underlying posterior predictive distribution.

In general, there are two approaches to specify an imputation: joint imputation models (Schafer, 1997; Little and Rubin, 2002) and sequential imputation models (Raghunathan, Lepkowski, and VanHoewyk 2001). Joint imputation model approach is preferable if the problem consists of small number of items, and it often uses Markov Chain Monte Carlo (MCMC) type algorithms for drawing missing values from the "joint" posterior predictive distribution. However, the joint models may be difficult to impose if there is a large number of incompletely-observed variables, and/or many different types of variables measured with restriction and boundaries are present in the data. An alternative approach considers a variable-by-variable approach where conditional models are used in a sequential fashion to approximate the joint posterior predictive distribution of the missing data. This approach produces the sets of imputations by fitting a sequence of models and drawing the missing values variable-by-variable from their perspective approximate posterior predictive distributions. Because these models can be tailored to approximate the unique features of the data (e.g. clustering, skip patterns or type of variable to be imputed), sequential approach can be advantageous in settings considered in our work (see Figure 1).

Our primary goal in this work is to assess the performance of sequential imputation method in the clustered-data settings. We implement this method using mixed-effects models and employ computational techniques based on MCMC and/or numerical integration to draw imputations. We would then like to evaluate the feasibility of this approach as well as compatibility with the more "coherent" joint modelling methodology to drawing imputations.

The remainder of this article is organized as follows. Section 2 outlines the variable-by-variable imputation approach (called "Sequential Imputation"). Models used to impute missing values in the individual variables are stated and computational methods approximating the posterior predictive distribution under each of these models are given. Section 3 summarizes our findings from the simulation study assessing the performance of the sequential approach in relation to joint distribution. Finally, Section 4 discusses the strengths and limitation of this approach.

2. Sequential Imputation

One of the most difficult step is to postulate a model to be used as a basis for sampling missing data (i.e. imputations) from the implied posterior predictive distribution. Such models, also called imputer's model, should be capable of preserving design features, relationships among variables subject to current/future analyses and should also be rich enough to account for causes of missingness so that the commonly assumed missingness mechanism MAR can be plausible. See Schafer (1997) and Little and Rubin (2002) for more discussion on selecting imputation models. In datasets with structures mentioned above (i.e. clustering, skip patterns/censoring, large number of variables measured at different scales, etc.), postulating one imputation model can be a daunting task. Rather than employing a large imputation model which may not even exist, sequential

imputation method (Raghunathan, Lepkowski, and VanHoewyk, 2001) offers a flexible solution that can potentially possess features of an ideal imputation model. Here we consider proposing a set of hierarchical models that are used sequentially for the purposes of imputations. In each of the sequence, the underlying posterior predictive distribution is simulated to draw to imputation.

2.1 Models

Imputation models are designed to reflect the levels of variation due to the multi-stage sampling. Here we focus on models for clustered sampling where there is only one clustering factor, and it could be extended to higher levels. Suppose U is a random variable subject to missingness and needs to be imputed. It could be continuous, categorical or count variable. Let u_{ij} denote a value of random variable for subject $j = 1,2, \ldots,n_i$ in cluster $i = 1,2, \ldots,m$. We assume that u_{ij} takes the following forms:

$$g(u_{ij} \mid x, \theta_u) = \eta_{ij}^u, \tag{1}$$

where $g(\bullet)$ denotes the link function of response u_{ij} to the linear predictors:

$$\eta_{ij}^u = x_{ij}^T \beta + z_i^T b_i, \tag{2}$$

where parameters β are fixed-effects common to all clusters and parameters b_i are random-effects specific to a cluster i. x_{ij} and z_i are the associated covariates corresponding to individuals within clusters and cluster characteristics. In this work we assume that random-effects follow a multivariate normal distribution $b_i \sim N(0, \sum_b)$.

It is often sufficient to assume a random-intercept model to account for the variation due to clustering sampling. It may be necessary to include more structures such as random-slope to correctly capture variation in other applications (longitudinal studies). Depending on the nature of the variable, the following specific imputation models are used for imputing binary, ordinal, continuous, count and nominal variables, respectively:

$$g(P(u_{ij} = 1 \mid x_{ij}, \theta_u)) = \log \frac{P(u_{ij} = 1 \mid x_{ij}, \theta_u)}{1 - P(u_{ij} = 1 \mid x_{ij}, \theta_u)}, \tag{3}$$

$$g(P(u_{ij} < l_u \mid x_{ij}, \theta_u)) = \log \frac{P(u_{ij} < l_u \mid x_{ij}, \theta_u)}{1 - P(u_{ij} < l_u \mid x_{ij}, \theta_u)}, \tag{4}$$

$$g(u_{ij} \mid x_{ij}, \theta_u)) = u_{ij} \mid x_{ij}, \theta_u, \tag{5}$$

$$g(u_{ij} \mid x_{ij}, \theta_u)) = \log(u_{ij} \mid x_{ij}, \theta_u), \tag{6}$$

$$g(P(u_{ij} = l \mid x_{ij}, \theta_u)) = \log \frac{P(u_{ij} = l \mid x_{ij}, \theta_u)}{P(u_{ij} = L \mid x_{ij}, \theta_u)}, \tag{7}$$

In the above models, x_{ij} means all potential covariates that may carry useful information including observed and imputed variables computed during the cycle of the sequential imputation.

We impose inverted Wishart prior on the variance for general random-effects term (inverted chi-square for the random-intercept models). Gelman (2006) suggested to use half-t family prior on the hierarchical standard deviation, which is another good choice. For fixed-effects we assume improper uniform density as prior. It is desirable to use a vague proper prior distribution on the fixed-effects parameters to avoid unidentifiability or overparameterization problems. Our findings suggest that the role of the priors should be negligible in most settings.

Imputation procedures typically assume that the missing values are missing at random (MAR) in the sense of Rubin (1976). Under MAR, the probability that any data value is missing may depend on quantities that are observed but not on quantities that are missing. Defining the missingness indicator as r_{ij} = 1 if the unit j in cluster i is observed and 0 otherwise, MAR means that $P(r|y,x,\theta) = P(r|\ y_{obs},x,\theta\ )$. The MAR assumption becomes more plausible as the models are enriched to include more information related to nonresponse, which is the main motivation behind the rich imputation models.

2.2 Computations

In the sequential imputation, each step in imputation cycles consists of finding maximum likelihood estimate of the unknown parameters which are used to build approximations to the posterior predictive distribution for the missing values. Below we describe the computation method in each of the specific conditional model.

2.2.1 Linear Mixed-effects Regression Models

Continuous variables are assumed to be related to the other variables in the following linear format which is also known as the mixed-effects models (here we are using a generic y variable to indicate the variable to be imputed):

$$y_{ij} = x_{ij}^T \beta + z_i^T b_i + \varepsilon_{ij}, \tag{8}$$

where x_{ij} and z_i are covariates of fixed effects and random effects for clusters i. Random-effects are assumed to follow a multivariate normal distribution $b_i \sim N(0, \sum_b)$ and errors are assumed to follows normal distribution $\varepsilon_{ij} \sim N(0, \sigma_e^2)$. The unknown parameters in the model are fixed effects β, random effects b_i, variance of random effects $\sum_b$ and variance of error σ_e^2. Conditional posterior distributions for each of these parameters need to be derived so that the Gibbs sampling method can be used to simulate the joint posterior of the unknowns which are $\beta, b_i, \sum_b, \sigma_e^2$ and y_{mis}. We employ Fisher scoring algorithm to get the approximated posterior distributions, and empirical Bayes estimates:

$$\beta \mid x, b_1, ..., b_m, \sum_b, \sigma_e^2 \sim N(\hat{\beta}, V(\hat{\beta})), \tag{9}$$

$$b_i \mid x, \beta, \sum_b, \sigma_e^2 \sim N(\hat{b}_i, V(\hat{b}_i)), \quad i = 1, 2....m, \tag{10}$$

$$\sigma_e^2 \mid x, b_1, ..., b_m, \sum_b, \beta \sim \chi_{DF}^{-2}, \tag{11}$$

$$\sum_b \mid x, b_1, ..., b_m, \beta, \sigma_e^2 \sim (\upsilon + \sum_i b_i^2)\chi_{m+\lambda}^{-2}, \tag{12}$$

The missing values in y are drawn under equation (13) using the draws of the $\beta, \sigma_e^2, \sum_b, b_i$ from their perspective posterior distribution (obtained after the Markov chain defined by (9)-(12) has converged). Let $y_{j\in mis}$ denote the missing values in cluster *i*, then $y_{j\in mis}$ is drawn from:

$$y_{j\in mis} \mid x_{j\in mis}, z_{j\in mis}, \beta, \sigma_e^2, \sum_b, b_i \sim N(x_{j\in mis}^T \beta + z_{j\in mis}^T b_i, z_{j\in mis}^T \sum_b z_{j\in mis} + \sigma_e^2), \quad (13)$$

where $x_{j\in mis}$ and $z_{j\in mis}$ are the associated covariates corresponding to the positions of missing values in *y*.

2.2.2 Generalized Linear Mixed-effect Regression Models

Here we consider binary variable only. Other types of variables such as Poisson, nominal or ordinal variables follow the similar computational techniques. The marginal log-likelihood for a binary variable *U* is

$$l(u \mid \theta = (\beta, b_1, ..., b_m, \sum_b)) = \prod_i \prod_j \left(F(x_{ij}^T \beta + z_i^T b_i) \right)^{u_{ij}} \left(1 - F(x_{ij}^T \beta + z_i^T b_i) \right)^{1-u_{ij}}, \quad (14)$$

where $F(\eta)$ is either $\Phi(\eta)$ or $(1+e^{-\eta})^{-1}$ for probit or logistic regression, respectively. Our interest lies in maximizing the marginal likelihood function $\int_b l(u \mid \theta) f(b) db$, where $f(b)$ represents the probability distribution for the random effects, we normally assume $N(0, \sum_b)$. This likelihood has to be evaluated using numeric method since there is no close-form solution to this maximization problem. There are several methods available to evaluate this integral, such as Gaussian quadrature method and Laplace approximation. We employ Gaussian-Hermite Quadrature method (Stoer, 2002), which can be fast and accurate under random-intercept-only model. Laplace approximation might be preferable if more structure to be introduced in the random-effects. Once we calculate the log-likelihood, we use Fisher scoring algorithm to obtain a proposal distribution of the regression parameters. Then Metropolis-Hastings algorithm is employed to perform the sampling from this proposal distribution. The estimated regression parameters are obtained after the sampler chain reaches its convergence. Finally, the missing values are drawn from the following Bernoulli distribution:

$$u_{j\in mis} \mid x_{j\in mis}, z_{j\in mis}, \beta, \sum_b, b_1, ...b_m \sim Bernoulli(G(x_{j\in mis}^T \beta + z_{j\in mis}^T b_i)), \quad (15)$$

where $G(\eta)$ is either $\Phi(\eta)$ or $\text{logit}^{-1}(\eta)$ for probit or logistic regression, respectively.

Imputation stage occurs in a cyclical manner and consists of a pre-determined *c* cycles. In the first cycle, we regress the variable with least number of missing values, say Y_1, on all completely-observed variables, and impute the missing values under an appropriate model discussed above. When imputation on Y_1 is completed, we continue to impute the variable with second least number of missing values, say Y_2, using the regression model of Y_2 on all other complete variables (observed and imputed Y_1 and the covariates). This

process is repeated until all variables have been imputed. Once the first cycle is completed, second cycle begins with Y_1 using all other variables including most recent imputed values and observed values. The new set of imputed values overwrite on the imputed values from first cycle. This process is repeated for all variables in second cycle, using the most recent imputed values. These cycles are repeated for a pre-determined number of *c* cycles. After all cycles are completed, we take the last several sets of imputed data and conduct multiple imputation inference following MI combining rules (Rubin, 1987).

3. Simulation Study

3.1 Design

3.1.1 Data generation

Our strategy of evaluating the performance of the sequential imputation method consists of the following steps:

(1) Simulate a complete data under general location model with random effects corresponding to unobserved cluster effects. Note that the sequential model is potentially incoherent with this data-generating mechanism. This step can be implemented by the following two steps. We assign i = 1,2,….50, j = 1,2,…,500, and only random-intercept models are considered in both steps:

a. Draw a continuous variable *Y* from the following linear model:

$$y_{ij} = \alpha_0 + u_i + \varepsilon_{ij}, \tag{16}$$

where i indexes the cluster, u_i is the random intercept term for cluster i effect, which is assumed to follow the normal distribution $N(0, \sigma_{by}^2)$, ε_{ij} is the error term assumed to follow the normal distribution $N(0, \sigma_e^2)$. Here we assign $\alpha_0 = 1$, σ_{by} varies from 0.5 to 3 and σ_e varies from 2 to 5. The purpose of assigning different variance values is to obtain different intracluster correlation coefficient (ICC) values ranging from 0.01 to 0.7.

b. Conditioning on *Y*, draw binary variable *X* from following logit mixed-effects model:

$$\text{logit}(\text{P}(x_{ij} = 1)) = \beta_0 + \beta_1 y_{ij} + b_i, \tag{17}$$

where i indexes the cluster, b_i is the random intercept term for cluster i effect, which is assumed to follow the normal distribution $N(0, \sigma_{bx}^2)$. Here we assign $\beta_0 = -1$, $\beta_1 = 0.5$, σ_{bx} varies from 0.1 to 1. Similar to step (a), the purpose of assigning different variance values is to obtain different ICC values ranging from 0.01 to 0.5.

(2) Estimate certain regression parameters of interest (an example is given below) on the data sets generated in the first step, and then combine them as the "true" parameters for the regression. These are "before-deletion" parameter estimates. This step considers the estimation of a regression model that often serves the substantive research objectives, which is also called analyst's model. Suppose the researchers are interested in the following random-intercept analysts models:

$$y_{ij} = \alpha_0 + \alpha_1 x_{ij} + u_i + \varepsilon_{ij}, \quad (18)$$

$$\text{logit}(\text{P}(x_{ij} = 1)) = \beta_0 + \beta_1 y_{ij} + b_i, \quad (19)$$

where $u_i \sim N(0, \sigma_u^2)$, $b_i \sim N(0, \sigma_b^2)$, $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, the parameters of interest in these models are $\alpha = (\alpha_0, \alpha_1)$, $\beta = (\beta_0, \beta_1)$, σ_u^2 and σ_b^2 . These parameters are estimated for each simulated data using complete-data before deletion (BD) and after deletion (AD).

(3) Impose missingness on the complete data under MCAR mechanism.

$$\text{logit}(\text{P}(r_x = 1)) = \gamma_0^x, \quad (20)$$

$$\text{logit}(\text{P}(r_y = 1)) = \gamma_0^y, \quad (21)$$

where r_x and r_y are missingness indicators for *X* and *Y* respectively, which indicate the average rates of missingness.

(4) Use the sequential imputation method to create multiple imputations of the missing values. Compare the results with R PAN (Schafer and Yucel 2002) package. PAN is a multivariate generalization of the linear mixed-effects models, which assumes "normality" on the binary variable. Note that due to lack of software for handling binary and continuous variable imputation via joint model approach, we are using PAN as an approximation for binary data imputation, which generally works satisfactorily (Schafer, 1997, Ch. 5).

(5) Perform MI inference on the regression parameters using these two approaches. In each simulated data, both methods in step (4) are used to create 10 imputed data from which the analyst's model given in (18) and (19) are estimated. The estimates and standard errors are then combined using Rubin's rules (Rubin, 1987).

3.1.2 Assessing the performance

The performance of the sequential imputation method is assessed by evaluating the coverage rates of the parameters of the models given in (18) and (19). We also investigate several factors that may influence the performance of sequential imputation method in multilevel settings, such as ICCs and rates of missingness.

3.2 Summary of the results

Results given in Table 1 suggest that the performance of the sequential imputation method in multilevel settings is reasonable well for continuous variables in all scenarios, with at least 90% coverage rate for 95% nominal confidence interval. For binary variables, the sequential imputation method outperforms PAN in almost all scenarios. The moderate missingness rates do not have much impact on the performance. Similarly, the role of the priors can be negligible in most settings.

Table 1: Simulation Results on Coverage Rates

Scenarios	Methods	μ_x	μ_y	β_0	β_1	α_0	α_1
Prior: flat prior for all parameters	AD*	95.2	95.6	94.5	94.2	95.3	95.8
ρ_x = 20%, ICC_x = 0.01	Sequential	94.3	94.6	94.9	94.0	95.3	95.0
ρ_y = 20%, ICC_y = 0.01	PAN	97.4	94.6	91.0	80.4	93.5	89.4
Prior: flat prior for all parameters	AD	94.4	93.4	91.5	91.4	92.8	92.6
ρ_x = 20%, ICC_x = 0.5	Sequential	93.7	92.7	80.6	39.9	92.1	98.0
ρ_y = 20%, ICC_y = 0.7	PAN	94.6	93.5	92.1	88.0	93.8	94.4
Prior: half-Cauchy on α,β, flat on σ_b, σ_u	AD	94.1	93.8	94.5	90.6	93.3	94.9
ρ_x = 20%, ICC_x = 0.5	Sequential	93.3	91.2	92.0	85.7	90.1	92.3
ρ_y = 20%, ICC_y = 0.2	PAN	95.9	91.0	91.3	62.4	92.5	89.2
Prior: flat prior for all parameters	AD	93.8	94.3	95.1	90.0	94.1	94.4
ρ_x = 20%, ICC_x = 0.5	Sequential	94.9	93.1	93.7	86.4	90.5	93.9
ρ_y = 20%, ICC_y = 0.2	PAN	97.6	92.9	92.0	62.1	93.7	89.0
Prior: flat prior for all parameters	AD	92.4	89.0	95.1	89.5	96.0	95.0
ρ_x = 40%, ICC_x = 0.5	Sequential	90.5	91.7	89.6	76.4	92.1	94.7
ρ_y = 40%, ICC_y = 0.2	PAN	94.1	93.0	88.1	66.0	93.1	84.1
Prior: flat prior for all parameters	AD	92.3	93.1	91.1	91.0	93.4	97.2
ρ_x = 60%, ICC_x = 0.5	Sequential	85.4	93.9	83.8	77.5	91.3	91.6
ρ_y = 60%, ICC_y = 0.2	PAN	94.2	93.8	84.9	67.1	90.5	86.0

* AD: after deletion, it represents the complete case analysis for incomplete data

Figure 3 shows the coverage rates of the regression parameters for different ICC's. We can see that ICC has a great effect on the performance of sequential imputation method for binary variable. Generally, the smaller ICC, the better performance.

4. Discussions

Our primary goal was to assess the performance of the sequential imputation method in the multilevel settings, such as clustered sampling involving many types of variables, restrictions and bounds. Our major finding was that in most scenarios, sequential approach performs as good as its alternatives. In itself, this is a significant findings as the alternatives often fail to be an MI tool in datasets of complexities considered here. Several of the simulation scenarios indicated that when the ICC was high, binary part of the sequential approach failed. This is somewhat unexpected, and is probably due to failure of the models estimating probabilities that are close to the boundaries. In real applications, when this is suspected, practitioners might want to consider pooling the

clusters together to prevent the estimation problems. Alternatively, models specifically designed for such cases (e.g. negative binomial) can be considered for imputation.

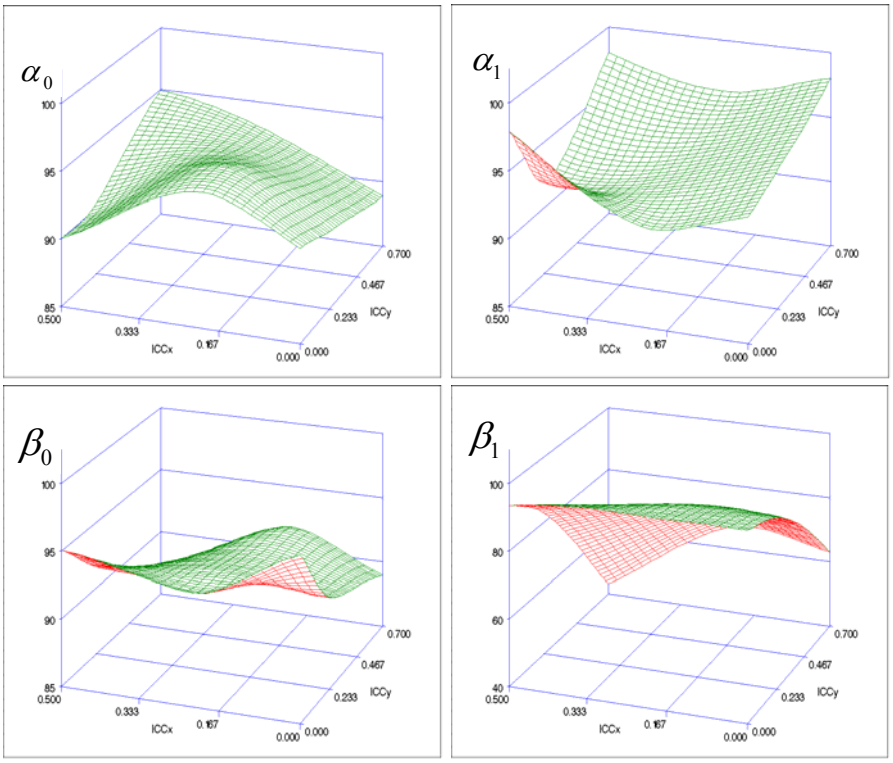

Figure 3: Coverage Rates for Regression Parameters vs. ICC_x and ICC_y

Our current work focuses on several important extensions of this work. First extension pertains to allowing higher number of variables measured in different scales. Second extension is the software development and interactive use of software products for the sequential imputation to take advantage of unique computational advantages offered by each of the products. Third extension consists of considering methods allowing a sequential modeling of missingness mechanism. Assuming one single missingness mechanism may not be realistic as in some application MAR, and MNAR may occur in a conditional fashion (e.g. MNAR given MAR or MCAR). Final extension is to make the methods flexible enough to handle unique data features such as higher levels of nesting.

References

Albert, J. (2007), *Bayesian Computation with R*, New York: Springer.

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

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y. S. (2008), "A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models", *The Annals of Applied Statistics*, 2, 1360-1383.

Gelman, A. (2006), "Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper)", Bayesian Analysis, 1, 515-534.

Little, R. and Rubin, D. (2002), *Statistical Analysis with Missing Data (Second Edition)*, New York: J Wiley & Sons, New York.

Liu, M., Taylor J. M. G., (2002), "Multiple Imputation and Posterior Simulation for Multivariate Missing Data in Longitudinal Studies", *Biometrics*, 56, 1157-1163.

Olsen, M. and Schafer, J. L. (2001), "A Two-Part Random-Effects Model for Semicontinuous Longitudinal Data", *Journal of the American Statistical Association*, 96, 730-745.

Raghunathan, T. E., Lepkowski, J. M., and Van Hoewyk, J. (2001), "A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models", *Survey Methodology*, 27, 85-95.

Raudenbush, S. and Bryke, A. (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods (Second Edition)*, Thousand Oaks, CA: Sage.

Rubin, D. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley and Sons.

Schafer. J. L., and Graham, J. W. (2002), "Missing Data: Our View of the State of the Art", *Psychological Methods*, 7, 147-177.

Schafer, J. L., and Yucel, R. M. (2002), "Computational Strategies for Multivariate Linear Mixed-Effects Models with Missing Values", *Journal of computational and Graphical Statistics*, 11, 421-442.

Schafer, J. L., and Maren K. O. (1998), "Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective", *Multivariate Behavioral Research*, 33, 545-571.

Schafer. J. L. (1999), "Multiple Imputation: a Primer", *Statistical Methods in Medical Research*, 8, 3-15.

Schafer. J. L. (1997), *Analysis of Incomplete Data*, London: Chapman & Hall.

Stoer, J., and Bulirsch, R. (2002), *Introduction to Numerical Analysis (third Edition)*, New York: Springer.

Van Buuren, S. (2007). "Multiple imputation of discrete and continuous data by fully conditional Specification", *Statistical Methods in Medical Research*, 16, 219-242.

Van Buuren, S. and Oudshoorn, C. (2000), *Multiple Imputation by Chained Equations: MICE V1.0 User's Manual*, TNO Preventie en Gezondheid: Report PG/VGZ/00.038.

Yucel, R. M., and Zaslavsky, A. (2004), "Practical Suggestions on Rounding in Multiple Imputation", *In JSM Proceedings, Survey Research Methods Section.* Toronto, ON, Canada: American Statistical Association. 4679-4683.