Weighted Least Squares Estimation with Missing Data

Tihomir Asparouhov and Bengt Muthén

August 14, 2010

1 Introduction

In this note we describe the Mplus implementation of the weighted least squares estimation in the presence of missing data. This estimation method has been available in Mplus since Version 3. The method yields consistent estimates under some general missing data assumptions, however, those assumptions are somewhat more restrictive than assumptions usually used with the maximum-likelihood estimator. In this note we prove the consistency of the weighted least squares estimates under the correct missing data assumptions and also conduct a simulation study to illustrate the performance of this estimator.

2 Types of Missing Data

Suppose that $Y = (Y_1, ..., Y_p)$ are the p observed dependent variables, $X = (X_1, ..., X_q)$ are the q observed independent variables in the model. In this note we consider the situation when missing data occurs only for the dependent variables, i.e., we assume that missing data for the independent variables does not occur. If this is not the case in a particular application the independent variables with missing values have to be included as dependent variables in the model so that the model can infer the missing values. Let $p_1, p_2, ..., p_L$ are all possible incomplete missing data patterns. Note that we only list the incomplete data patterns, i.e., the full data pattern is not included in this list.

Suppose that in pattern p_l we have the decomposition $Y = (Y_{o,l}, Y_{m,l})$, where $Y_{o,l}$ are the observed variables in that pattern and $Y_{m,l}$ are the missing variables. A missing data mechanism is a model for the missing data pattern, i.e., a model that describes the probability $P(p_l)$ that the pattern p_l occurs. In particular we are interested in the effect of the model variables $Y_{o,l}$, $Y_{m,l}$ and X on the probability $P(p_l)$. There are four different types of missing data mechanisms that we consider in this note. We list these below in order from the most restrictive to the most general.

- MCAR - Missing completely at random This type is defined by the equation

$$P(p_l|X, Y_{o,l}, Y_{m,l}) = f(l),$$

  where f is a function, i.e., none of the three types of variables $X, Y_{o,l}, Y_{m,l}$ have an effect on the missing data patterns.

- MARX - Missing at random with respect to X This type is defined by the equation
$$P(p_l|X, Y_{o,l}, Y_{m,l}) = f(l, X),$$
where f is a function, i.e., only the covariate variables X have an effect on the missing data patterns. Note here that if there are no covariates in the model then MARX is equivalent to MCAR.

- MAR - Missing at random This type is defined by the equation
$$P(p_l|X, Y_{o,l}, Y_{m,l}) = f(l, X, Y_{o,l}),$$
where f is a function, i.e., only the covariate variables X and the observed dependent variables $Y_{o,l}$ have an effect on the missing data patterns. Note here that if there is only one dependent variable Y then there is only one incomplete pattern which has no observed dependent variables in it. Therefore MAR is equivalent to MARX for models with one dependent variable.

- NMAR - Missing at random This type is defined by the equation
$$P(p_l|X, Y_{o,l}, Y_{m,l}) = f(l, X, Y_{o,l}, Y_{m,l}),$$
where f is a function, i.e., all three types of variables have an effect on the missing data patterns.

It is well know how FIML (full information maximum-likelihood) estimation performs under all of these conditions, see Rubin and Little (2002) and Muthen and Brown (2001). FIML yields consistent parameter estimates and standard errors when the missing data is MAR (and also under the more restrictive assumptions MARX and MCAR) when the model is estimated from the entire data sets including observations with missing data. FIML estimates can be biased when under the NMAR assumption. Under the NMAR assumption it is possible to obtain consistent estimates if the missing data mechanism model is estimated as well, see Muthén et at. (2010).

If listwise deletion is applied, i.e., the model is estimated only from observations with full records then the ML estimates are consistent under the MCAR and MARX assumptions but the estimates are less efficient than the FIML estimates based on the entire data set.

In this note we study the performance of the weighted least square estimation in Mplus under the various missing data assumptions. For brevity we denote the weighted least square estimation by WLS, but everything in this note applies also for the remaining weighted least squares estimators WLSMV, WLSM and ULSMV.

3 WLS under MARX

In this section we will show that under the MARX assumption the WLS estimator yields consistent estimates. Suppose that there are N observations in the data. Define the missing variable indicator R_{ir} for $i = 1, ..., P$ and $r = 1, ..., N$ by

$$R_{ir} = \begin{cases} 0 & \text{if } Y_{ir} \text{ is missing} \\ 1 & \text{otherwise.} \end{cases}$$

We follow the description of the WLS estimator given in Muthén and Satorra (1995) for the complete data case and explain how that estimator is modified to accommodate missing data. Denote by σ_1 the first stage parameters (intercepts, thresholds and slopes) and by σ_2 the second stage parameters (correlations and covariances). Denote by $l_{ir} = L(Y_{ir}|X)$ and $l_{ijr} = L(Y_{ir}, Y_{jr}|X)$ be the univariate and the bivariate conditional log-likelihoods for the $r-$th individual. The full information univariate and bivariate conditional log-likelihoods are

$$l_i = \sum_{r=1}^{N} l_{ir} R_{ir} \tag{1}$$

and

$$l_{ij} = \sum_{r=1}^{N} l_{ijr} R_{ir} R_{jr} + \sum_{r=1}^{n} l_{ir} R_{ir}(1 - R_{jr}) + \sum_{r=1}^{n} l_{jr}(1 - R_{ir}) R_{jr}. \tag{2}$$

Under the MARX condition

$$P(R_{ir} = 0) = \sum_{l} f(l, X)$$

where the sum is taken over all missing data patterns p_l for which the $i-$th variable is missing. Therefore the multivariate MARX condition implies univariate MARX condition since in the above formula only X influences the probability of missingness. Similarly one can establish that the multivariate MARX condition leads to bivariate MARX condition for any pair of variables Y_i and Y_j. Since the MARX condition is a special case of the MAR condition we can therefore conclude that both the univariate and bivariate models can be estimated constantly using FIML estimation. The WLS estimator uses the univariate FIML estimates as the first stage estimate $\hat{\sigma}_1$ and therefore these estimates are consistent. The second stage estimates $\hat{\sigma}_2$ are obtained by fixing the σ_1 parameters in (4) to the first stage estimates $\hat{\sigma}_1$ and then maximizing (4) over the second stage parameters. Since the first stage estimates are consistent this estimation is equivalent to the estimation where the first stage parameters are fixed to their true values, i.e., to the FIML estimation of the second stage parameters where the first stage parameters are fixed to their true values. Since the missing data mechanism is MAR this FIML estimation is consistent and therefore the second stage WLS estimates are also consistent.

Note here that

$$l_i = \sum_r l_{ir} \tag{3}$$

where the sum is taken over all observations r for which Y_{ir} is present. Also if we ignore the second and third sums in (4) which do not contain any second stage parameters we get that

$$l_{ij} = c + \sum_r l_{ijr} \tag{4}$$

where the sum is taken over all observations r for which both Y_{ir} and Y_{jr} are present and c is a constant independent of the second stage parameters, i.e., a constant that can be ignored in the second stage optimization. This shows that the first and the second stage WLS estimates are essentially obtained by univariate and bivariate listwise deletion, i.e., by pairwise deletion.

As in Muthén and Satorra (1995) (under the regularity conditions B1-B7) the consistency of the first and the second stage estimates implies the consistency of the third stage estimates. The proof of the asymptotic normality of the parameter estimates is the same as in Muthén and Satorra (1995). The only new assumption that we make is that for all pairs (i,j) as $n \to \infty$, $\sum_{r=1}^N R_{ir}R_{jr} \to \infty$, i.e., the pairs of variables where both variables are present goes to infinity as N goes to infinity. Of course, this is a requirement for the consistency as well.

To obtain the standard errors for the WLS estimates we use the same method as in Muthén and Satorra (1995). Let g_r be the vector of all first derivatives for the r-th observation

$$g_r = \left(R_{1r}\frac{\partial l_{1r}}{\partial \sigma_{1,1}}, ..., R_{pr}\frac{\partial l_{pr}}{\partial \sigma_{1,p}}, R_{1r}R_{2r}\frac{\partial l_{21r}}{\partial \sigma_{2,21}}, ..., R_{p-1\ r}R_{pr}\frac{\partial l_{pp-1\ r}}{\partial \sigma_{2,pp-1}} \right).$$

Let $g = \sum_{r=1}^{N} g_r$. Let $\hat{\sigma}$ be the first and the second stage estimates and let $\bar{\sigma}$ be the true parameter values. For some point σ^* between $\hat{\sigma}$ and $\bar{\sigma}$ $0 = g(\hat{\sigma}) = g(\bar{\sigma}) + \partial g(\sigma^*)/\partial\sigma(\hat{\sigma} - \bar{\sigma})$ and therefore

$$N^{1/2}(\hat{\sigma} - \bar{\sigma}) = \left(\frac{-N^{-1}\partial g(\sigma^*)}{\partial \sigma}\right)^{-1} N^{-1/2} g(\bar{\sigma}).$$

By Liapounov CLT $n^{-1/2}g(\bar{\sigma})$ is asymptotically normal with mean zero and variance $V = N^{-1} lim \sum_{r=1}^{n} E(g^r(\bar{\sigma})g^r(\bar{\sigma})')$. Also if

$$A = plim\left(\frac{-N^{-1}\partial g(\sigma^*)}{\partial \sigma}\right) = plim\left(\frac{-N^{-1}\partial g(\bar{\sigma})}{\partial \sigma}\right)$$

we get that $N^{1/2}(\hat{\sigma} - \bar{\sigma}) \xrightarrow{d} N(0, \Gamma = A^{-1}VA'^{-1})$.

The structural parameters θ are estimated in the third stage by minimizing the objective function

$$F(\theta) = \sum(\sigma(\theta) - \hat{\sigma})W^{-1}(\sigma(\theta) - \hat{\sigma})'$$

where W is chosen to be either Γ or the diagonal of Γ or the identity matrix depending on which weighted least square estimator we use. To get the asymptotic distribution of the structural parameters $\hat{\theta}$ we apply Theorem 4.1.3. in Amemiya (1985) and we get that

$$Var(\hat{\theta}) = N^{-1}(\Delta' W^{-1}\Delta)^{-1}\Delta' W^{-1}\Gamma W^{-1}\Delta(\Delta' W^{-1}\Delta)^{-1}$$

where $\Delta = \partial\sigma/\partial\theta$.

Let's also consider the properties of the listwise deletion WLS estimation. Listwise deletion can be thought of as a two-stage missing data scheme. In the first stage missing data occurs from the true missing data mechanism and in the second stage all variable of an observation are removed if any of the variables for that observation is missing. Under the MARX assumption only X variables can affect the missing data pattern for the true missing data mechanism. This however implies that only X variables affect the two-stage missing data mechanism. Thus under MARX the WLS estimation with listwise deletion is a special case of the WLS estimation with pairwise deletion for a MARX two stage missing data mechanism. Therefore under MARX the WLS estimation with listwise deletion is also consistent however it uses less information that the WLSMV pairwise deletion and therefore it can be expected to be less efficient.

4 Simulation Study

In this section we compare the WLSMV estimator described above which we call WLSMV-PD (pairwise deletion) with the WLSMV estimator using listwise deletion which we call WLSMV-LD. Consider the following simple structural equation model with three dependent binary variables and one independent continuous covariate. Let the binary variable Y_j take values 0 and 1. Let the continuous covariate be X and the latent factor variable be η. The model is described by the following two equations

$$P(Y_j = 0) = \Phi(\tau_j - \lambda_j \eta)$$

$$\eta = \beta X + \xi$$

where Φ is the standard normal distribution function. We generate 100 data sets using the above model and the following parameter values $\tau_j = 0$, $\lambda_j = 1$, $\beta = 1$. Both ξ and X are generated from a standard normal distribution. Missing data is generated using the following missing data mechanism

$$P(Y_j \text{ is missing}) = \frac{1}{1 + Exp(1 - X)},$$

i.e., from a MARX missing data mechanism. This missing data mechanism yields approximately 26% of missing data for each variable and approximately 57% of all observations have at least one missing value. Thus the WLSMV-LD is based on 43% of the data while WLSMV-PD is based on 74% of the data. We analyze each data set with both estimators. The results of the simulation study are presented in Table 1. It is clear from these results that both WLSMV-PD and WLSMV-LD produce unbiased estimates for the parameters and their standard errors, however the WLSMV-PD estimates are much more efficient than the WLSMV-LD estimates.

5 Discussion

In this note we described how the weighted least squares estimators in Mplus handle missing data and we showed that the estimators are consistent under the MARX missing data assumption. We also showed that these estimators are much more efficient than estimators based on listwise deletion. This weighted least squares method has been available in Mplus since Version 3.

Table 1: MARX simulation results

	WLSMV-PD	WLSMV-LD	WLSMV-PD	WLSMV-LS
Parameter	Bias(Coverage)	Bias(Coverage)	MSE	MSE
τ_1	0.01(.96)	0.02(.92)	0.006	0.016
τ_2	0.00(.97)	-0.02(.92)	0.005	0.016
τ_3	-0.01(.95)	-0.02(.96)	0.006	0.014
λ_1	0.02(.99)	0.03(.97)	0.013	0.029
λ_2	0.00(.93)	0.01(.97)	0.017	0.028
λ_3	0.03(.95)	0.05(.96)	0.024	0.050
β	0.00(.99)	0.00(.96)	0.004	0.012

In the latest Version 6 new methods are available that can handle less restrictive missing data assumptions. In Version 6 it is possible to use any of the following three method under the more general MAR assumption: the maximum-likelihood estimator, the Bayes estimator, and the multiple imputation method followed by the weighted least squares estimator. The simulation studies in Section 3 of Asparouhov and Muthén (2010) illustrate the deficiencies of the weighted least squares estimation under the more general MAR condition and show that the above three approaches resolve these deficiencies.

In this note we described the missing data handling of the weighted least squares estimation for single level models however similar arguments and results apply for the weighted least squares estimation of two-level models. Two-level simulation studies are presented in Asparouhov and Muthén (2010).

Note also that the weighted least squares method can not be used for NMAR modeling. Many of the NMAR models include the missing data indicators in the model. Under this modeling approach however the missing data mechanism is not MARX but it is MAR since the missing data indicator variables have an effect in the missing data patterns.

Nevertheless the weighted least squares treatment of missing data should be useful in situations when the amount of missing data is not substantial or when the MARX assumption is considered plausible. Including more covariates in that model that could potentially predict missingness can yield a model where the MARX condition is more plausible.

References

[1] Asparouhov, T. and Muthén, B. (2010). Bayesian analysis of latent variable models using Mplus. Technical Report. http:\\statmodel.com \download \BayesAdvantages6.pdf

[2] Amemiya, T. (1985) Advanced Econometrics. Harvard University Press.

[3] Muthén, B., Asparouhov, T., Hunter, A. and Leuchter, A. (2010). Growth modeling with non-ignorable dropout: Alternative analyses of the STAR*D antidepressant trial. Submitted for publication. http:\\www.statmodel.com

[4] Muthén, B. and Brown, H. (2001) Non-Ignorable Missing Data in a General Latent Variable Modeling Framework. Draft. http:\\gseis.ucla.edu \faculty \muthen \articles \Article_093.pdf

[5] Muthén, B. and Satorra, A. (1995) Technical Aspects of Muthen's Liscomp Approach to Estimation of Latent Variable Relations With a Comprehensive Measurement Model Psychometrika-Vol. 60, No. 4, 489-503.

[6] Rubin, D. and Little, R. (2002). Statistical analysis with missing data (2nd ed.). New York: Wiley