Weighting for Unequal Probability of Selection in Multilevel Modeling

Tihomir Asparouhov
Muthen & Muthen

Mplus Web Notes: No. 8

Version 3, December 09, 2004

[1]The author is thankful to Bengt Muthen for his guidance, to Linda Muthen for her support and commitment and to Thuy Nguyen for computational assistance. The author is thankful to Rod Little and Laura Stapleton for helpful comments on the earlier draft. This research was supported by SBIR grant R43 AA014564-01 from NIAAA to Muthen & Muthen.

Abstract

In this note we construct an approximately unbiased multilevel pseudo maximum likelihood (MPML) estimation method for weighting in general multilevel models. We conduct a simulation study to determine the effect various factors have on the estimation method. The factors we included in this study are scaling method, size of clusters, invariance of selection, and the informativeness of selection. The scaling method is an indicator of how the weights are normalized on each level. The invariance of the selection is a indicator of whether or not the same selection mechanism is applied across clusters. The informativeness of the selection is an indicator of how biased the selection is. We summarize our findings and recommend a multistage procedure based on the MPML method that can be used in practical applications.

1 Introduction

Multilevel models are frequently used to analyze data from cluster sampling designs. Such sampling designs however often use unequal probability of selection at the cluster level and at the individual level. Sampling weights are assigned at one or both levels to reflect these probabilities. If the sampling weights are ignored at either level the parameter estimates can be substantially biased.

Weighting for unequal probability of selection in single level models is a relatively well established procedure which we discussed in Mplus Web Note #7 [As]. For multilevel models however the situation is completely different. There is currently no well established general multilevel consistent estimation method. Several methods have been proposed in the literature for example in [GK], [KG], [KR], [PSHGR], [PMS] and [St], however the asymptotic properties of these estimators are unknown, especially when the cluster sample sizes are small. These estimators can produce biased parameter estimates and the size of the bias can be evaluated only through a limited simulation studies. General comparison between the methods is not available. Several simulation studies in [GK], [KG], [PSHGR], [PMS] and [St] give some limited information about the direct comparison between some of the methods. The comparisons very often turn out inconclusive to some degree and depend on the particular model and the particular sampling mechanism. These simulation studies have also clearly indicated that as the cluster sample sizes increase the parameter bias can decrease. Such estimation methods are called approximately unbiased. However this property has not been formally established and the exact conditions for which it holds are unclear. In addition the previously proposed methods apply only to certain multilevel models and parameters, and cannot be extended beyond the framework they are defined in.

In this note we introduce the multilevel pseudo maximum likelihood (MPML) estimation method which can be seen as a natural extension of the pseudo maximum likelihood (PML) method defined in [S] for the single level models. We find exact conditions which guarantee that the parameter estimates are approximately unbiased. For certain special models we obtain a closed form expressions for the parameter estimates, which enable us to compare this method with previously proposed methods. The main advantage of this method is that it can be applied to any general multilevel model just as the PML method can be applied to single level models. The method is also flexible and it can be modified for different estimation problems. In Section 2 we provide more background information on unequal probability sampling in multilevel settings. In Section 3 we define the MPML method and describe the variance estimation for the parameter estimates. In Section 4 we clarify the concept of weights scaling, which has been the primary bias reduction tool for example in [PSHGR] and [St], and we show that different scaling methods should be used depending on whether or not the sampling mechanism is of two different types: invariant and non-invariant. In section 5 we develop a measure for the level of informativeness of the selection mechanism which can be used to evaluate the effect of different selection mechanisms on the estimation method and to compare different simulation studies and simulation studies with practical applications. The measure of informativeness can also be used in evaluating the need for incorporating the weights in the analysis. In section 6 and 7 we conduct a simulation study that evaluates the effect of several components on the MPML estimation method. These components are informative index, sample cluster size, invariance of sampling mechanism and scaling method. In Section 8 we conduct a simulation study on a multilevel model that can be estimated only by the MPML method and not by any other previously proposed method. In Section 9 we focus on the tools available in Mplus Version 3 ([MM]; www.statmodel.com) needed to deal with weights in multilevel models. In Section 10 we summarize our findings and outline a 7 step procedure that should be followed for proper usage of weights in multilevel models. This procedure is designed primarily for Mplus users but it is also of general interest and can be useful for such analysis with other statistical packages. In Appendix A we provide theoretical justification for the MPML method and underline the conditions needed for the approximate consistency of the parameter estimates. In Appendix B we obtain a closed form expression for the balanced random intercept model and compare the MPML method to other previously proposed methods.

This note can be regarded as a continuation of Mplus Web Note #7 [As]. The reader is encouraged to review the material in that note first, especially because certain weighted multilevel analyses are discussed only in Web Note #7. These cases are omitted from the discussion here because the nature of the weighting is not multilevel even though the models are.

2 Background

The intricacies of multilevel weighted analysis begin with the choice of weights. Frequently data sets are made available with weights computed to reflect the sampling design. Usually these weights are prepared for the analysis of single level models and for computing means. Unfortunately these weights are not appropriate for multilevel models and the choice of the model, single level or multilevel, affects the computation of the weights. Weights computed for single level analysis can produce erroneous results if used with multilevel models. For example suppose that simple random sampling (SRS) is used at both stages of a two-stage sampling design. If we analyze the data by a single level model we have to construct weights reflecting the relative size of the clusters and the number of units sampled from each cluster. If we analyze the data by estimating a two-level model we should not use weights to reflect the fact that at each level the sampling is done completely at random, i.e., the appropriate weights are 1 at both levels. In the mean estimation framework this is to say that weights used to estimate consistently the average of an observed variable can not be used directly to estimate the average of the cluster means estimated by two-level models.

The usual description of weighted two-level analysis includes weighting for unequal probability of selection at level 2 - the cluster level, weighting for unequal probability of selection for level 1 - the individual level, or both. However, because of the PML method developed in [S], the only case that requires new theoretical considerations is the weighting for unequal probability of selection for level 1 units. When weights are present at level 2 only, that is to say that independent units, namely clusters, have been sampled with unequal probability, we identify this framework as the framework of single level weighted modeling and all methods available for single level weighted analysis can be applied. Indeed a two level model with weights on level 2 can be presented as a multivariate single level model with weights. Thus the single level PML method and any other single level method can be used without any complications in this situation. The model estimates will be consistent regardless of the size of the clusters. It is shown in Section 10 of [As] how a multilevel model with weights on the second level is estimated with the usual single level PML technique developed in [S]. The situation is completely different when weights are present at level 1 because the unequal probability of selection is applied to dependent units and thus the assumptions of the single level methodology are violated. The MPML method defined in Section 3 applies to the general case of weighting on both levels. Our primary interest in multilevel weighted analysis however is in the case when the level 1 units are weighted for unequal probability of selection.

If the selection mechanism is not informative we should exclude the weights from the analysis. The estimates will remain consistent and in fact will be more precise. Including non-informative weights in the analysis can result in a substantial loss of efficiency. It cannot result however in bias of the parameter estimates, when either the PML estimator or the proposed MPML estimator is used, as long as certain regularity conditions on the distribution of the weights and the random variables are satisfied. These regularity conditions are needed by the asymptotic estimation theory (see [A], Chapter 4). General regularity conditions for example guarantee that the effective sample size, as defined in [PWM], approaches infinity as the true sample size approaches infinity. Thus the effective sample size based variance estimation proposed in [PWM], which is unbiased when the weights are constant, is asymptotically equivalent to the PML variance estimate.

In practical applications it may not be possible to easily determine whether the selection is informative or not. A general method for testing the informativeness of the weights is described in [P] and the test can be used for multilevel models as well as single level models. A simpler but incomplete method is proposed in Section 5 based on the informative index. If the weights are determined to be non-informative they should not be used in the analysis. If any informativeness tests are inconclusive, including the weights in the analysis is necessary. From now on we assume that the selection mechanism at level 1 is informative.

The primary tools for evaluating multilevel weighting estimation methods so far have been simulation studies. Simulation studies in general are a weak tool for evaluating the quality of an estimation technique but that is particularly the case when it comes to weighted data analysis because the selection mechanism used in the simulation study has to be sufficiently informative. Otherwise even the unweighted parameter estimates would be approximately unbiased and any reasonable weighted parameter estimates would be so as well. It is not clear however how to choose a sufficiently informative selection, and how to distinguish such a selection from a slightly informative selection. Another danger in choosing a selection mechanism is the possibility that two different subpopulations have been oversampled, which cancel each other out, and essentially result in a non-informative selection. Finally, different estimation techniques can produce better results for different selection mechanisms and different parameters which can lead to very confusing set of recommendations applicable only to the simulation studies these originated in.

One of the key issues in the multilevel weighted estimation literature has been the fact that the parameter estimation are usually only approximately unbiased, i.e., they are unbiased for sufficiently large cluster sample size, but can be severely biased when the cluster sample size is small. Different scaling of the weights has been one of the focal points of the bias reducing techniques for example in [PSHGR] and [St]. There have been no theoretical results to support one scaling method over another. Scaling of the weights comprises of multiplying the weights by a scaling constant so that the sum of the weights is equal to some kind of characteristic of the sample, for example, the total sample size. In multilevel models the scaling modification methods scale the weights differently across clusters so that the total weight of the cluster is equal to some cluster characteristic. In single level modeling the scaling of the weights does not affect the PML estimator at all. In multilevel models that is not the case however and the ratio between the true cluster sample size and the total weight within the cluster is an important quantity since it affects for example the posterior distribution of the level 2 random effects. The different scaling methods however may have different effect on different estimation techniques. If a scaling method performs well with the MPML approach proposed in Section 3, it will not necessarily perform well with other estimation techniques.

Another prospective that exposes the need for scaling the weights is the following. Let the weight for individual i in cluster j be w_{ij}. Typically w_{ij} is computed as $w_{ij} = 1/p_{ij}$ where p_{ij} is the probability that individual i in cluster j is included in the sample. Often however these probabilities are very small. Thus the resulting weights are very large, which is not very practical and the weights are consequently rescaled. Alternatively the size of the population that is being sampled may not be known and consequently the exact probabilities of selection may not be known. Nevertheless unequal probability of selection could be implemented, for example certain ethnicity could be oversampled at a given rate even when the size of the population is not known. In these cases the scale of the weights is undetermined and only the relative value of the weights has practical meaning. Methods that rely on the equations $w = 1/p$ would not be applicable directly. That is why it is desirable that the model estimation technique is scale invariant and there is no dependence on the scale of the weights. In fact we conduct our simulation studies without any specific cluster size assumptions and without generating a specific finite target population. Instead a model defined population is being sampled with unequal probabilities. This approach simplifies the description of the simulation study.

In this note we focus on the maximum-likelihood based estimation only, which has no closed form and it involves an iterative maximization procedure. The estimation techniques proposed in [KG], [GK] and [KR] have a closed form solution however these techniques can be applied only to a limited set of models. Indeed many multilevel models do not allow closed form solutions even without sampling weights and we need to resort to ML based estimation. Several modeling features can not be handled outside of the ML framework. One such feature is missing data with continuous and categorical outcomes when the missing data is MAR. Indeed there is no closed form solution for this estimation problem even for the estimation of the mean of an observed variable. A second feature that confines us to the ML framework is mixture modeling. An example of this type would be a multilevel growth mixture model and the example considered in Section 8. A third feature is random slopes, i.e., modeling individually varying covariance structures. The method developed in [KR] has a closed form solution that applies to linear models with random slopes however that method is less efficient than the ML method even without sampling weights. A forth feature that requires the ML framework is categorical dependent variables for example is multilevel logistic regression.

3 Multilevel Pseudo Maximum Likelihood

In this section we define the MPML estimation method just like the PML method is defined in [S]. The parameter estimates are defined as the solution of a set of score equations, or equivalently as the parameters that maximize the pseudo maximum-likelihood function. This method is more general than previously proposed methods because it is not constrained by a specific multilevel model or particular assumptions about the distribution of the observed or latent variables in the model. The MPML method can be used with any parametric family of distributions and any linear or non-linear multilevel models.

In this general multilevel model the observed variable in cluster $j = 1, ..., M$ of individual $i = 1, ..., n_j$ is y_{ij} and the level 2 random effects in cluster j is η_j. The predictors on the individual level are denoted by x_{ij} and the predictors on the cluster level are denoted by x_j. The density function of y_{ij} is $f(y_{ij}, x_{ij}, \eta_j, \theta_1)$. The density function of η_j is $\phi(\eta_j, x_j, \theta_2)$. The parameters θ_1 and θ_2 are to be estimated. Suppose that the data is sampled with unequal probability of selection on both levels. Suppose that the probability of selection for cluster j is p_j and $w_j = 1/p_j$. Suppose that the probability of selection for individual i in cluster j, given that cluster j is selected, is p_{ij} and the sampling weight is $w_{ij} = 1/p_{ij}$. We define the MPML estimates $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ as the parameters that maximize the weighted pseudo likelihood

$$l(\theta_1, \theta_2) = \prod_j \left(\int \Big(\prod_i f(y_{ij}, x_{ij}, \eta_j, \theta_1)^{w_{ij} s_{1j}} \Big) \phi(\eta_j, x_j, \theta_2) d\eta_j \right)^{w_j s_{2j}},$$

where s_{1j} and s_{2j} are level 1 and level 2 scaling constants. Let $\hat{n}_j = \sum_i w_{ij} s_{1j}$ be sum of the scaled weights within cluster j. The effect of this value on the estimation is very similar to the effect of the cluster sample size on the maximum likelihood estimation without sampling weights. Within the EM maximization algorithm for example $\hat{n}_j$ is an indicator of the extent to which η_j is determined by the observed data and is inversely proportional to the posterior variance of η_j. Thus the scale factor s_{1j} can be used to balance this effect. The second level scaling s_{2j} can be used to counter balance the first level scaling s_{1j}. We will be interested primarily in the following choices $s_{2j} = 1$ or $s_{2j} = 1/s_{1j}$. The MPML estimates are approximately consistent if

1. n_j and $\hat{n}_j$ are sufficiently large,
2. s_{2j} and η_j are independent,
3. $\hat{n}_j/n_j$ and η_j are independent.

The proof of the approximate consistency can be found in the Appendix A. A closed form solution for the MPML estimates for the balanced random intercept model is derived in Appendix B. There is a considerable flexibility in the definition of MPML because the scaling constants are relatively unrestricted. A natural choice for s_{1j} is $s_{1j} = n_j / \sum_i w_{ij}$. In that case $\hat{n}_j = n_j$ and the third condition above is automatically satisfied even when η_j and n_j are not independent.

Denote by $L = log(l)$ the weighted log-likelihood and by $L_j = log(l_j)$ the weighted log-likelihood of the j-th cluster, where l_j is

$$l_j = \int \Big(\prod_i f(y_{ij}, x_{ij}, \eta_j, \theta_1)^{w_{ij} s_{1j}} \Big) \phi(\eta_j, x_j, \theta_2) d\eta_j.$$

The asymptotic covariance matrix given by standard asymptotic theory is

$$(L'')^{-1}\Big(\sum_j (s_{2j}w_j)^2 L'_j L'^T_j\Big)(L'')^{-1},$$

where $'$ and $''$ refer to the first and the second derivative of the log-likelihoods.

The MPML just as the PML is defined as a general estimator not connected to any optimization algorithm. Virtually any ML optimization algorithm can be adapted to include the weights and maximize the MPML objective function. In fact Mplus 3 implements several such algorithms among which are the EM-algorithm, the Quasi-Newton algorithm and the Fisher scoring algorithm.

The MPML as defined above is a generalization of the parameter estimation method used in [GP] for multilevel logistic regression.

4 Scaling

In this section we discuss in detail the concept of scaling of the weights. We assume that the probability of selection has only a relative meaning. If $p_{i_1 j} : p_{i_2 j} = 2$ we interpret this as indication that individuals similar to individual i_1 are oversampled at a rate of 2:1 in comparison to the individuals similar to individual i_2. Therefore the weights w_{ij} should be standardized by the scale factors s_{1j} to some meaningful values. In choosing the proper standardization several considerations should be made. First we need to understand whether or not the ratio between the weights of individuals from different clusters is a meaningful quantity. That is, we have to know whether the $w_{i_1 j_1} : w_{i_2 j_2}$ can be interpreted in the usual sense of oversampling. This would be so if the same sampling mechanism is used across the clusters, i.e., no cluster level variable, a random effect or an observed variable has an effect on the sampling. If different sampling mechanisms have been used, depending on cluster specific information, then the weights within each cluster could be already on different scales and direct comparison between the weights can be detrimental. If the same sample mechanism has been used, i.e., sampling mechanism is invariant across clusters, and the ratio between the weights across clusters is a meaningful quantity we need to choose scaling that takes advantage of that quantity and the relative information is not lost. In practice a good understanding of the selection mechanism would be sufficient to determine its type. For example oversampling adolescents to adults in rates

3:1 is invariant. However oversampling adolescents to adults with a varying rate that can guarantee that a target percentage of the sample population in the cluster are adolescents is not invariant if the outcome variable is not independent of the percentage of adolescents in the cluster. Quota sampling would in general produce a non-invariant selection. Sampling a predetermined number of students with a certain ethnic background within each classroom is not an invariant selection. Another example of a non-invariant selection would be the case when the non-invariance is explicitly used in the sampling design, namely different sampling rates used in different clusters.

Note that we are only discussing scaling for the level 1 weights. In multilevel models the MPML estimates are independent of the scale of the level 2 weights, just as in a single level model the PML estimates are independent on the scale of the weights. In fact the scale constants s_{2j} are not needed to standardize the level 2 weights but to possibly counter the standardization on level 1 and recover any information that may be lost after the level 1 standardization. Of course s_{2j} could not be considered as a scale factor for the level two weights also because they depend on j. The scaling constants s_{2j} also do not need any standardization, that is the MPML estimates depend only on the relative values of s_{2j}. In choosing scaling we also need to make sure that conditions 1, 2 and 3 in Section 3 are satisfied so that the consistency is guaranteed at least for large cluster sizes.

We consider 6 different weighting methods. Since $\hat{n}_j/s_{1j} = \sum_i w_{ij}$ the definition of the scale constants s_{1j} can be expressed as a definition of $\hat{n}_j$. Also define the effective sample size n_{0j} as in [PWM], $n_{0j} = (\sum_j w_{ij})^2/\sum_j w_{ij}^2$. The weighting methods we are interested in are

Method A. $\hat{n}_j = n_j$ and $s_{2j} = 1$

Method AI. $\hat{n}_j = n_j$ and $s_{2j} = 1/s_{1j}$

Method B. $\hat{n}_j = n_{0j}$ and $s_{2j} = 1$

Method BI. $\hat{n}_j = n_{0j}$ and $s_{2j} = 1/s_{1j}$

Method C. $s_{1j} = \sum_j n_j/\sum_{ij} w_{ij}$ and $s_{2j} = 1$

Method D. Unweighted analysis

The scaling methods A and B have been proposed in the literature already in [PWM], [PSHGR] and [St]. The scaling methods AI and BI are based on scaling methods A and B but also include a level two offset scaling. The index letter I in the name of the scaling methods AI and BI is to indicate that the methods are appropriate only for invariant selection mechanisms. Indeed if the selection is not invariant η_j would influence s_{1j} and in turn s_{2j} which would be a violation of condition 2 in Section 3. For invariant selection mechanisms these methods are expected to perform better since they would use all the information including the ratio between weights across clusters. We also demonstrate these facts in the simulation studies in Sections 6 and 7 below. Scaling method C is the scaling method that is traditionally used with mean estimation. This method has constant scaling across clusters and $\sum_j \hat{n}_j = \sum_j n_j$.

5 Informative Index

The quality of the MPML estimates is driven primarily by two factors, the sample size of the clusters and the degree of informativeness of the selection mechanism. While the sample size is an explicit variable, the informativeness is not directly measurable. Such measurement is needed however to study the dependence of the quality of the MPML estimates on the informativeness. Pfeffermann [P] constructs a test statistic that allows us to determine whether the selection mechanism is informative. If $\hat{\theta}_w$ and $\hat{\theta}_0$ are the parameter estimates of the weighted and unweighted analysis respectively, and $\hat{V}(\hat{\theta}_w)$ and $\hat{V}(\hat{\theta}_0)$ are their variance estimates,

$$I = (\hat{\theta}_w - \hat{\theta}_0)^T [\hat{V}(\hat{\theta}_w) - \hat{V}(\hat{\theta}_0)]^{-1} (\hat{\theta}_w - \hat{\theta}_0) \sim \chi_{(p)},$$

has approximately a chi-square distribution with $p = \dim(\theta)$ degrees of freedom. The value of I is a clear measurement of the informativeness of the selection. We now define a similar test statistic for a specific variable Y in the model. Suppose that $\hat{\mu}_w$ is the weighted estimated mean of Y and $\hat{\mu}_0$ the unweighted estimated mean of Y. Let $\hat{\sigma}_w^2$ and $\hat{\sigma}_0^2$ be the estimated variance for $\hat{\mu}_w$ and $\hat{\mu}_0$ respectively. The informativeness of the selection mechanism for variable Y can then be measured by the T statistic

$$I_1(Y) = \frac{\hat{\mu}_w - \hat{\mu}_0}{\sqrt{\hat{\sigma}_w^2 - \hat{\sigma}_0^2}}.$$

Large absolute values would indicate that the selection is informative for Y. The values of $I_1(Y)$ are easy to compute once the weighted and unweighted analysis have been completed. As the sample size increases however all selection mechanisms become highly significant and thus it is not possible to form recommendations based on the value of I_1 that apply to any sample size. Another problem with I_1 is that it is not defined when $\hat{\sigma}_w < \hat{\sigma}_0$. This happens quite often when the sample size is not large or when the weights are not sufficiently different across individuals. An informativeness measure that is independent of the sample size is

$$I_2(Y) = \frac{\hat{\mu}_w - \hat{\mu}_0}{\sqrt{v_0}},$$

where v_0 is the unweighted estimate of the variance of Y. This informativeness measure depends however on the scaling and the cluster sample size since the MPML method depends on the scaling when the cluster size is small. This leads us to define

$$I_3(Y) = \frac{\mu - \hat{\mu}_0}{\sqrt{v_0}},$$

where μ is the true mean of Y. This informativeness measure would be relatively independent of the cluster size and the sample size, however in practice we can only estimate $I_2(Y)$, and approximate $I_3(Y)$ by $I_2(Y)$. This approximation will be sufficient as long as there is no substantial bias in the mean parameter estimate. In a simulation study of course I_3 can be easily computed since the true parameter value μ is known.

The mean parameters are the parameters that usually are the most sensitive to bias selection. Thus evaluating the informative index for all dependent variables will generally be sufficient to determine whether the sampling and the resulting weights are informative. If the sampling weights appear to be non-informative or very slightly informative it is very likely that the reduction in the bias due to the weighting of the analysis will be overwhelmed by the increase in the variance of the parameter estimates. Therefore including the weights in the analysis will in effect only increase the mean squared error. Thus we recommend that in these cases the weights are not used at all. We generally recommend that if the informative index for all dependent variables is less than 0.02 the weights are not included in the analysis. The informative index is a very easy to compute and practical tool, however it is not a universal too of detecting informative selections. For example if the weights are only informative for variance and covariance parameters the informative index would not detect that.

6 Invariant Selection

We conduct a simulation study on the following model

$$y_{ij} = \mu + \eta_j + \varepsilon_{ij},$$

where $\mu = 0.5$, and η_j and ε_{ij} are zero mean normal random variables with variance $\psi = 0.5$ and $\theta = 2$ respectively. The selection model is defined by $P(I = 1) = 1/(1+e^{-\varepsilon_{ij}/\alpha})$, i.e., the probability of inclusion is dependent only on the level 1 residual and is invariant across clusters. We use 3 different values for α to achieve 3 different levels of informativeness of the selection. For $\alpha = 1, 2, 3$ the approximate values for $I_3(Y)$ are 0.5, 0.3 and 0.2 respectively. Within each simulation the cluster size is constant. We use three different values for the cluster size $n_j = 5, 20$, and 100. Each of the analyses is replicated 500 times. In all cases we used 100 cluster units. The results of the simulation are presented in Table 1. The table contains the absolute bias and the percentage of times the true parameter was covered by the 95% confidence intervals in the parenthesis. All estimates for the θ parameter were negatively biased. All estimates for the ψ parameter were positively biased. All estimates for the μ parameter were positively biased except for method BI which had negatively biased μ estimates. We make the following conclusions based on the results in Table 1.

- The unweighted parameter estimates are substantially biased. The bias increase as the informative index I_3 increases but is unaffected by the cluster size n_j.
- The cluster sample size and the informative index I_3 affect the quality of the results for all 5 weighted methods. The most difficult case is when the cluster size n_j is small and the selection mechanism is strongly informative, rows 1, 2 and 4 in each of the subtables qualify to be in this case. In the remaining 6 cases, i.e., the cases when either the cluster size is sufficiently large or the informative index I_3 is small, the asymptotics appear to become valid to a varying extent and all 5 scaled weighting methods appear to perform reasonably well.
- Choosing a scaling method is not an easy task. Consider this, the best method for estimating μ is AI, the best method for θ is BI and the best method for ψ is D in our simulation study. Nevertheless some conclusions are quite clear. Methods AI and BI, which are designed

Table 1: Absolute Parameter Bias (Coverage). Invariant Selection.

μ parameter

n_j	I_3	A	AI	B	BI	C	D
5	0.5	0.27(27)	0.00(95)	0.31(13)	0.13(93)	0.21(50)	0.73(0)
5	0.3	0.12(79)	0.00(94)	0.13(75)	0.02(95)	0.08(89)	0.45(1)
5	0.2	0.07(89)	0.00(96)	0.08(88)	0.00(95)	0.04(92)	0.32(10)
20	0.5	0.10(78)	0.00(93)	0.12(67)	0.15(86)	0.09(82)	0.73(0)
20	0.3	0.03(95)	0.00(96)	0.10(80)	0.02(96)	0.02(96)	0.45(0)
20	0.2	0.02(95)	0.00(96)	0.02(95)	0.00(96)	0.02(95)	0.32(1)
100	0.5	0.03(94)	0.00(94)	0.03(93)	0.09(87)	0.02(93)	0.83(0)
100	0.3	0.00(93)	0.00(92)	0.01(93)	0.01(92)	0.00(93)	0.45(0)
100	0.2	0.01(97)	0.00(96)	0.01(97)	0.00(96)	0.01(97)	0.32(0)

θ parameter

5	0.5	0.62(0)	0.47(3)	0.65(0)	0.30(82)	0.49(0)	0.51(1)
5	0.3	0.22(66)	0.14(78)	0.26(59)	0.09(91)	0.14(73)	0.19(66)
5	0.2	0.09(87)	0.05(91)	0.11(87)	0.03(95)	0.05(90)	0.08(88)
20	0.5	0.30(3)	0.21(36)	0.42(1)	0.15(94)	0.21(22)	0.53(0)
20	0.3	0.07(83)	0.04(90)	0.10(80)	0.03(97)	0.04(90)	0.20(8)
20	0.2	0.03(92)	0.02(95)	0.04(93)	0.01(97)	0.02(94)	0.10(60)
100	0.5	0.09(58)	0.06(78)	0.19(36)	0.04(99)	0.06(74)	0.53(0)
100	0.3	0.01(92)	0.01(93)	0.02(93)	0.00(97)	0.01(93)	0.20(0)
100	0.2	0.00(95)	0.00(95)	0.00(96)	0.00(96)	0.00(95)	0.10(0)

ψ parameter

5	0.5	0.29(58)	0.45(55)	0.17(80)	0.34(67)	0.30(72)	0.02(92)
5	0.3	0.11(92)	0.12(92)	0.08(93)	0.08(93)	0.11(93)	0.02(92)
5	0.2	0.05(94)	0.04(93)	0.03(94)	0.02(93)	0.04(94)	0.02(92)
20	0.5	0.15(77)	0.21(75)	0.09(88)	0.23(78)	0.15(79)	0.00(93)
20	0.3	0.03(95)	0.03(96)	0.02(95)	0.02(95)	0.03(95)	0.01(93)
20	0.2	0.01(95)	0.01(96)	0.01(95)	0.01(95)	0.01(95)	0.01(94)
100	0.5	0.04(94)	0.05(93)	0.02(93)	0.07(91)	0.04(93)	0.01(92)
100	0.3	0.01(94)	0.01(94)	0.00(93)	0.00(93)	0.01(94)	0.01(91)
100	0.2	0.00(93)	0.00(92)	0.00(93)	0.00(92)	0.00(93)	0.00(92)

to take advantage of the fact that the selection is invariant outperform their non-invariant counterparts A and B, except in the first row in the ψ subtable, which however can not outweigh all other cases. As expected also methods A and AI perform somewhat better than B and BI for the mean parameters and worse for variance covariance parameters.

- It is hard to recommend one scaling method for all situations. Nevertheless we single out method AI. This method performs well when the informativeness is not very strong or the cluster sizes are not small. If the cluster sizes are small and the informative index is large however we would recommend that a detailed analysis is undertaken, using several different scaling methods and perhaps a single level model is used as a final model.
- It is hard to determine a specific threshold level for I_3 and n_j that would guarantee unbiased results in general. From what we see in this simulation study, we conclude that if $I_3 < 0.2$, the AI method will work sufficiently well with any sample size. If $0.2 < I_3 < 0.3$ a cluster sample size of at least 10 is needed. If $0.3 < I_3$ we would recommend using the AI method only with cluster size 35-40 and above.
- Our simulation study shows that the intraclass correlation is overestimated in general by all methods to a varying degree when some selection bias remains.
- Method AI seems to produce unbiased estimates for all mean parameters with any cluster size and informative index.

7 Non-invariant Selection

In this section we use the same setup as in the previous section. We only change the selection model to $P(I = 1) = 1/(1 + e^{-y_{ij}/\alpha})$. The informative index I_3 is slightly lower now but approximately unchanged. The selection is clearly non-invariant, because η_j influences the selection. Thus methods AI and BI are not expected to perform well even asymptotically. Table 2 contains the results of this simulation study. The following conclusions can be made from these results.

- Methods AI and BI are not suitable for this situation, the fact that they violate condition 2 from Section 3 results in biased estimates for the μ

Table 2: Absolute Parameter Bias (Coverage). Non-invariant Selection.

μ parameter

n_j	I_3	A	AI	B	BI	C	D
5	0.5	0.23(35)	0.15(87)	0.28(15)	0.31(68)	0.13(74)	0.61(0)
5	0.3	0.10(83)	0.11(83)	0.11(78)	0.13(79)	0.02(94)	0.40(0)
5	0.2	0.07(89)	0.07(89)	0.07(89)	0.08(88)	0.01(94)	0.29(11)
20	0.5	0.08(83)	0.16(70)	0.11(70)	0.39(29)	0.05(89)	0.61(0)
20	0.3	0.03(91)	0.10(77)	0.04(90)	0.13(67)	0.01(92)	0.40(0)
20	0.2	0.01(93)	0.08(83)	0.01(93)	0.09(78)	0.00(93)	0.29(4)
100	0.5	0.02(95)	0.16(52)	0.03(94)	0.39(9)	0.01(95)	0.61(0)
100	0.3	0.01(92)	0.10(72)	0.01(92)	0.13(61)	0.00(92)	0.40(0)
100	0.2	0.01(95)	0.07(85)	0.01(95)	0.08(83)	0.01(95)	0.30(1)

θ parameter

5	0.5	0.52(0)	0.42(8)	0.54(3)	0.27(80)	0.45(2)	0.47(3)
5	0.3	0.19(64)	0.13(81)	0.23(64)	0.08(92)	0.14(75)	0.19(66)
5	0.2	0.10(84)	0.07(90)	0.12(84)	0.04(92)	0.07(88)	0.10(83)
20	0.5	0.24(13)	0.18(47)	0.32(9)	0.12(96)	0.18(35)	0.48(0)
20	0.3	0.06(84)	0.04(89)	0.09(82)	0.03(95)	0.04(88)	0.19(9)
20	0.2	0.03(90)	0.02(91)	0.04(89)	0.01(93)	0.02(91)	0.10(65)
100	0.5	0.07(66)	0.05(78)	0.15(50)	0.04(98)	0.05(74)	0.48(0)
100	0.3	0.01(92)	0.01(92)	0.02(93)	0.00(97)	0.01(91)	0.19(0)
100	0.2	0.00(95)	0.00(95)	0.01(96)	0.00(97)	0.00(95)	0.10(5)

ψ parameter

5	0.5	0.15(91)	0.42(75)	0.01(93)	0.33(68)	0.20(93)	0.21(32)
5	0.3	0.08(93)	0.13(92)	0.04(95)	0.09(93)	0.09(95)	0.09(78)
5	0.2	0.04(94)	0.05(94)	0.02(94)	0.04(94)	0.04(96)	0.06(87)
20	0.5	0.08(93)	0.20(86)	0.00(91)	0.26(73)	0.08(94)	0.22(5)
20	0.3	0.02(94)	0.04(94)	0.01(94)	0.04(93)	0.02(95)	0.10(63)
20	0.2	0.01(94)	0.01(94)	0.00(94)	0.01(94)	0.01(94)	0.05(83)
100	0.5	0.02(93)	0.07(92)	0.00(90)	0.14(84)	0.02(93)	0.21(2)
100	0.3	0.00(92)	0.01(92)	0.00(92)	0.02(93)	0.00(92)	0.10(60)
100	0.2	0.00(93)	0.01(94)	0.00(93)	0.01(93)	0.00(93)	0.05(82)

parameter even when the cluster size is large. The selection bias fails to decrease as n_j increases.

- Methods A, B and C perform well when the cluster size is large or the informative index is small. The case of small cluster size and large informative index is again a difficult one. There is no large difference between the three methods. However method C slightly outperforms method A, which slightly outperforms method B.

8 Multilevel Mixtures

In this section we conduct a simulation study on the following model

$$y_{ij} = \mu + \beta x_{ij} + \eta_j + \varepsilon_{ij},$$

where x_{ij} is a binary covariate taking values 0 and 1 equally likely and where x_{ij} has missing values. As in the previous sections we assume that η_j and ε_{ij} are normally distributed random variables with variance θ and ψ. The parameter values we use in the simulation are as follows $\beta = 1$, $\theta = 2$, $\psi = 0.5$ and $\mu = 0$. The probability that x_{ij} is missing is $1/(1+e^{(y_{ij}-0.75)/4})$. Unequal probability of selection is induced by the following inclusion model $P(I = 1) = 1/(1 + e^{-\varepsilon_{ij}/2})$. We use three different values for the cluster size $n_j = 5, 10$, and 20. Each of the analysis is replicated 500 times. In all cases we used 100 cluster units.

This model is only a slight modification of the model considered in the previous section and usually its estimation would not be very different. That is not the case however if the covariate x_{ij} has missing values. When x_{ij} is missing the conditional distribution of $[y_{ij}|\eta_j]$ is not normal but it is a bimodal distribution which is obtained as a mixture of two normal distributions with equal variance and different means. This is why the estimation of this model is much more complicated than the usual multilevel model. We used the multilevel mixture track in Mplus 3 to estimate the model with the MPML method. Mplus is the only statistical package that can estimate such models as far as we know. Because the probability that x_{ij} is missing depends on y_{ij} the missing data type is not MCAR but it is MAR and therefore listwise deletion method would not only reduce the sample size but would also produce biased estimates. Thus bias in the estimates for this model can arise from not including the incomplete cases or from not including the weights

Table 3: Selection Bias (Coverage) in Multilevel Mixtures with AI Scaling

n_j	μ	β	θ	ψ
5	$-0.01(94)$	$0.01(90)$	$-0.18(85)$	$0.15(91)$
10	$0.00(92)$	$0.02(92)$	$-0.11(92)$	$0.07(92)$
20	$0.00(94)$	$0.01(92)$	$-0.05(96)$	$0.05(93)$

in the analysis. The MPML method however resolves both problems. The selection is invariant across clusters and we therefore can use the AI scaling method. Table 3 contains the bias and coverage of the MPML estimates. The bias of the estimates is quite small except for the two variance estimates when $n_j = 5$. This is in line with the observations and conclusions we made in the previous section. We therefore extend these conclusions to the multilevel mixture models and other multilevel models. The MPML estimate would be approximately unbiased as long as the cluster sizes are not small when the informative index is large.

9 Mplus Implementation

Mplus 3 implements the MPML estimator with scaling method AI. If weights are available for both levels the total weight is entered into the weight variable, that is, if weights are available for both levels the product of the two weight variables is formed and used in the analysis. Sometimes in practice the total weight is the only weight available without specific information about the level 1 weight and the level 2 weight. In that case the total weight variable is used directly.

Mplus 3 can also implement the MPML estimator with scaling method A. For that purpose however the level 1 weights are standardized so that within each cluster the sum of the level 1 weights is equal to the size of the cluster. Once the level 1 weights are standardized, the total weight variable is formed by multiplying the standardized level 1 weight variable and the level 2 weight variable. When only the total weight is available it is not possible to implement scaling method A unless we assume that the total weight is the level 1 weight, i.e., assuming that the level 2 weight is 1. In that case, method A can be implemented. If this assumption is not correct however and part of the weight variable is the level 2 weight the results may be incorrect.

Mplus 3 does not implement scaling methods B, BI and C at this time.

10 Conclusion

In this note we discussed some of the intricacies of weighting for unequal probability of selection in multilevel models. We introduced the multilevel pseudo maximum likelihood estimation method with general scaling options and provided conditions guaranteeing that these estimates are approximately unbiased. Through our simulation studies we demonstrated how sample size, the informative index, the nature of the selection mechanism and the type of scaling affect the quality of the MPML estimates. We summarize our findings in the following 7 steps procedure that can be used to steer away from the pitfalls of weighing in multilevel modeling. This procedure is designed for Mplus users, but it can by used with other statistical packages.

Step 1. Verify that weights are designed for a multilevel analysis. If the weights are designed for a single level analysis, and multilevel weights can not be obtained, instead of attempting multilevel modeling we recommend single level modeling that is designed for cluster sampling designs. Such modeling is available in Mplus with type=complex (see also [As] section 8 and 9).

Step 2. If weights are available only at level 2 there is no need to proceed with the rest of the steps in this procedure. Although the model is multilevel, the nature of the weighting is not. There are no complications in this case. The estimates available in Mplus are consistent and the quality of the estimation is as good as the unweighted case.

Step 3. Determine whether or not the selection mechanism of the level 1 units is invariant across cluster. This information could be extracted from a short description of the sampling design. If such information is not available assume that the selection is non-invariant.

Step 4. If the selection is invariant use scaling method AI. If the selection is non-invariant use scaling method A. Compute the total weight as described in Section 9 taking the scaling method into account.

Step 5. Perform unweighted ML and weighted MPML analysis. Compute the informative index for every variable in the model.

Step 6. If all informative indices are below 0.02 use unweighted analysis and ignore the weights. Incorporating the weights in the analysis can actually decrease the precision of the parameter estimates. If any of the informative indices is above 0.3 and the average cluster size is smaller than 10 then the parameter estimates obtained by the MPML method may have a substantial bias. In this case we recommend a single level analysis as described in step 1. Borderline cases would have to be examined on an individual basis. If the informative index is less than 0.3 for all variables or the average sample size is larger than 10 the MPML estimates are expected to be trustworthy.

Step 7. Avoid using the log-likelihood chi-square test for testing nested models. The test does not take into account the unequal probability of selection design and the test statistic will be overestimated. Using a T-test or a Wald test is a valid alternative since the asymptotic covariance of the estimates accounts for the sampling weights. Alternatively Mplus computes robust chi-square tests with the MPML estimator for many models which provide a correction for the log-likelihood chi-square and take the sampling weights into account. More detailed discussion on this topic is available in [As].

Additional steps which are currently not available in Mplus 3 could involve for example the Pfeffermann test for informativeness of the weights as described in Section 5 and possibly comparing the results from different scaling methods.

The main advantage of the MPML method is its generality. While for some specific estimation problems perhaps a more accurate non-ML based estimator could be constructed, its generality is unparalleled. Cautiously using the MPML method can be a very effective tool in dealing with selection bias in multilevel modeling.

11 Appendix A

For brevity we denote by $F_{0j} = \phi(\eta_j, x_j, \theta_1)$ and by $F_{ij} = f(y_{ij}, x_{ij}, \eta_j, \theta_2)$. Let $\theta = (\theta_1, \theta_2)$ be the total parameter vector. Using the Laplace approximation ([D], [L])

$$l_j \approx e^{g_j(\hat{\theta}_{0j})}\sqrt{2\pi}\hat{\sigma}_{0j}\hat{n}_j^{-0.5},$$

where

$$g_j(\theta) = \log(F_{0j}) + s_{1j}\sum_i w_{ij} log(F_{ij}),$$

$\hat{\theta}_{0j}$ is the mode of $g_j(\theta)$, and $\hat{\sigma}^2_{0j} = -\hat{n}_j/g_j''(\hat{\theta}_{0,j})$. Since $g_j/\hat{n}_j \to E_j(\log(F_{ij}))$ as $n_j \to \infty$, where E_j is the expectation with respect to y_{ij} conditional on individual i being in cluster j, we get that $\hat{\theta}_{0j} \to \hat{\theta}_j$– the mode of $E_j(\log(F_{ij}))$. Similarly $\hat{\sigma}_{0j} \to \hat{\sigma}_j$ as $n_j \to \infty$, where $\hat{\sigma}_j^2 = -1/E_j(\log(F_{ij}))''$. Thus for sufficiently large n_j

$$l_j \approx e^{g_j(\hat{\theta}_j)}\sqrt{2\pi}\hat{\sigma}_j\hat{n}_j^{-0.5}.$$

Maximizing the approximate likelihood amounts to solving the following approximate score equations obtained by omitting the lower level terms and terms that do not depend on the θ parameters

$$\frac{\partial}{\partial\theta_1}\sum_j\sum_i w_j w_{ij} s_{2j} s_{1j} \log(F_{ij}(\hat{\theta}_j)) = 0$$

$$\frac{\partial}{\partial\theta_2}\sum_j w_j s_{2j} log(F_{0j}(\hat{\theta}_j)) = 0.$$

These two equations can be approximated by

$$\frac{\partial}{\partial\theta_1}E(s_{2j}\hat{n}_j E_j(log(F_{ij}(\hat{\theta}_j)))) = 0$$

$$\frac{\partial}{\partial\theta_2}E(s_{2j} log(F_{0j}(\hat{\theta}_j))) = 0.$$

Under the assumptions 2 and 3 in Section 3, these score equations are equivalent to

$$\frac{\partial}{\partial\theta_1}E(n_j E_j(log(F_{ij}(\hat{\theta}_j)))) = 0$$

$$\frac{\partial}{\partial\theta_2}E(log(F_{0j}(\hat{\theta}_j))) = 0.$$

At this point we see that the score equations are independent of the weights and the sampling scheme and thus sampling weights of 1, i.e., using simple random sampling, would yield the same approximate score equations. We conclude that the MPML and the ML estimates are asymptotically equivalent and thus the MPML estimates are approximately unbiased for large enough n_j and $\hat{n}_j$.

12 Appendix B

Here we illustrate the MPML estimator by deriving a closed form expressions for the parameter estimates of a random intercept model. Suppose that

$$y_{ij} = \mu + \eta_j + \varepsilon_{ij}$$

where η_j and ε_{ij} are zero mean normally distributed variables with variances ψ and θ respectively. The weighted likelihood of the j cluster is

$$l_j = \int (2\pi\theta)^{-\hat{n}_j/2}(2\pi\psi)^{-1/2} Exp\bigg(- \frac{s_{1j}}{2\theta}\sum_i w_{ij}(y_{ij}-\mu-\eta_j)^2 - \frac{\eta_j^2}{2\psi}\bigg) d\eta_j =$$

$$(2\pi\theta)^{-\hat{n}_j/2}(2\pi\psi)^{-1/2} Exp\Big(-\frac{s_{1j}}{2\theta}\sum_i w_{ij}(y_{ij}-\bar{y_j})^2\Big) \int Exp\bigg(-\frac{\hat{n}_j}{2\theta}(\bar{y_j}-\mu-\eta_j)^2-\frac{\eta_j^2}{2\psi}\bigg) d\eta_j$$

where $\bar{y_j}$ is the weighted mean $\bar{y_j} = \sum w_{ij}y_{ij}/\sum w_{ij}$. Using the Laplace approximation ([D], [L]) formula which is exact when the function in the exponent is quadratic we get that

$$l_j = (2\pi\theta)^{-(\hat{n}_j-1)/2}(2\pi(\theta+\hat{n}_j\psi))^{-1/2} Exp\bigg(-\frac{s_{1j}}{2\theta}\sum_i w_{ij}(y_{ij}-\bar{y_j})^2-\frac{\hat{n}_j(\bar{y_j}-\mu)^2}{2(\theta+\hat{n}_j\psi)}\bigg).$$

Explicit maximization of the pseudo log-likelihood is possible only when $\hat{n}_j$ is constant across all clusters. In that case the parameter estimates are

$$\hat{\mu} = \frac{\sum_j s_{2j} w_j \bar{y_j}}{\sum_j s_{2j} w_j}$$

$$\hat{\theta} = \frac{\sum_j s_{2j}s_{1j}w_j \sum_i w_{ij}(y_{ij}-\bar{y_j})^2}{\sum_j s_{2j}w_j(\hat{n}_j-1)}$$

$$\hat{\psi} = \frac{\sum_j s_{2j}w_j(\bar{y_j}-\hat{\mu})^2}{\sum_j s_{2j}w_j} - \frac{\hat{\theta}}{\hat{n}_j}.$$

Weighting methods A and AI would allow explicit maximization when all cluster sample sizes n_j are equal. In that case when implementing scaling method A we get

$$\hat{\mu}_A = \frac{\sum_j w_j \bar{y_j}}{\sum_j w_j}$$

$$\hat{\theta}_A = \frac{1}{\sum_j w_j(n_j - 1)} \sum_j n_j w_j \frac{\sum_i w_{ij}(y_{ij} - \bar{y}_j)^2}{\sum_i w_{ij}}$$

$$\hat{\psi}_A = \frac{\sum_j w_j(\bar{y}_j - \hat{\mu})^2}{\sum_j w_j} - \frac{\hat{\theta}}{n_j}$$

and implementing scaling method AI we get

$$\hat{\mu}_{AI} = \frac{\sum_{ij} w_j w_{ij} y_{ij}}{\sum_{ij} w_j w_{ij}}$$

$$\hat{\theta}_{AI} = \frac{\sum_{ij} w_j w_{ij}(y_{ij} - \bar{y}_j)^2}{\sum_{ij} w_j w_{ij}(1 - 1/n_j)}$$

$$\hat{\psi}_{AI} = \frac{\sum_{ij} w_j w_{ij}(\bar{y}_j - \hat{\mu})^2}{\sum_{ij} w_j w_{ij}} - \frac{\hat{\theta}}{n_j}.$$

The parameter estimate for with scaling method A are asymptotically equivalent to Method 2 in [St], note however that the asymptotic covariance of the parameter estimates is not the same. The asymptotic covariance estimates for Method 2 in [St] are negatively biased while the asymptotic covariance estimates of the MPML method are generally consistent. Explicit maximization with scaling method C is possible if $\sum_i w_{ij}$ is constant across clusters. In that case we get that

$$\hat{\mu}_C = \frac{\sum_j w_j \bar{y}_j}{\sum_j w_j}$$

$$\hat{\theta}_C = \frac{\sum_{ij} w_j w_{ij}(y_{ij} - \bar{y}_j)^2}{\sum_j w_j(\sum_i w_{ij} - \bar{w})}$$

$$\hat{\psi}_C = \frac{\sum_j w_j(\bar{y}_j - \hat{\mu})^2}{\sum_j w_j} - \frac{\hat{\theta}}{\hat{n}_j}$$

where $\bar{w} = \sum_{ij} w_{ij} / \sum_j n_j$ and $\hat{n}_j = \sum_i w_{ij}/\bar{w}$. This method is somewhat similar to method C described in [GK] and [KG], the only difference for example in the θ estimation is that $\bar{w}$ is replaced by 1 however that estimator becomes biased for small cluster sample sizes even with SRS as noted in [KG], where as the MPML estimator is consistent even for small cluster sample sizes. In fact $\hat{\theta}_C$ could be used even in unbalanced cases as a moment based estimator that avoids the pitfalls of estimator C of [GK] and [KG].

Scaling methods B and BI have a closed form solution when the effective sample size $(\sum_i w_{ij})^2 / \sum_i w_{ij}$ is constant across clusters. The exact formulas are derived similarly. In that case scaling method B produces the same parameter estimates for θ as Method 3 in [St], while the parameter estimate for ψ is approximately the same especially for large cluster sample sizes, because in [St], $\hat{n}_j$ is replaced by n_j in the computation of the average cluster size.

13 References

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

[As] Asparouhov, T. (2004). Sampling weights in latent variable modeling. Mplus Web Note #7. Forthcoming in Structural Equation Modeling.

[D] De Bruijn, N. G. (1961). Asymptotic Methods in Analysis, Amsterdam: North-Holland.

[GK] Graubard, B. and Korn, E. (1996). Modeling the Sampling Design in the Analysis of Health Surveys. Statistical Methods in Medical Research 5, pp 263-281.

[GP] Grilli, L. and Pratesi, M. (2004). Weighted estimation in multilevel ordinal and binary models in the presence of informative sampling designs. Survey Methodology, 30, 93-103.

[KG] Korn, E. and Graubard, B. (2003). Estimating the variance components by using survey data. Journal of the Royal Statistical Society, Series B, 65, Part 1, 175-190.

[KR] Kovacevic, M. S., and Rai, S. N. (2003). A pseudo maximum likelihood approach to multilevel modeling of survey data. Communications in Statistics, Theory and Methods, 32, 103-121.

[MM] Muthen, L.K. and Muthen, B.O. (1998-2004). Mplus User's Guide. Third Edition. Los Angeles, CA: Muthen & Muthen

[L] Lindley, D.V. (1980). Approximate Bayesian Methods. In Bayesian Statistics (eds. J.M. Bernardo, M.H. Degroot, D.V. Lindley, and A.N.F. Smith). Valencia, Spain: University Press.

[P] Pfeffermann, D. (1993). The Role of Sampling Weights when Modeling Survey Data. International Statistical Review, 61, 2, pp 317-337.

[PMS] Pfeffermann, D., Moura, F., and Silva, P. (2004) Multi-Level Modeling Under Informative Sampling. Submitted.

[PSHGR] Pfeffermann, D., Skinner, C.J., Holmes, D.J., Goldstein, H. and Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. Journal of the Royal Statistical Society, Series B, 60, 23-56.

[PWM] Potthoff, R.F., Woodbury, M.A., Manton, K.G. (1992). "Equivalent Sample Size" and "Equivalent Degrees of Freedom" Refinements for Inference Using Survey Weights Under Superpopulation Models. Journal of the American Statistical Association 87, 383-396.

[S] Skinner, C. J. (1989). Domain means, regression and multivariate analysis. In Analysis of Complex Surveys (eds. C.J.Skinner, D.Holt and T.M.F. Smith) pp. 59-87, Wiley.

[St] Stapleton, L. (2002). The Incorporation of Sample Weights Into Multilevel Structural Equation Models. Structural Equation Modeling 9(4) pp. 475-502.