

Implementing Resampling Methods for Design-based Variance Estimation in Multilevel Models: Using HLM6 and SAS together

Fritz Pierre¹ and Abdelnasser Saïdi¹

¹ Statistics Canada, 100 Tunney's Pasture Driveway Ottawa ON K1A 0T6

Abstract

Multilevel models (MLMs) are often used to analyse clustered survey data. Several commercial software packages developed for fitting such models to survey data have implemented a model-based sandwich estimator to estimate the design-based variance (DBV) of the model parameters. However, there is empirical evidence that this estimator underestimates the DBV due to ignoring the stochastic adjustments to the sampling weights (Kovacevic et al. (2006)). Therefore, resampling methods have been proposed to more accurately estimate the DBV. However, none of the software that can currently fit MLMs to survey data has implemented these methods.

In this paper, we present a brief review of resampling methods for the DBV estimation of the estimated parameters in MLMs. We then describe BHLMSAS_V0, a SAS macro that has been developed to use HLM6 (specialized software to fit MLMs) in the estimation of the DBV by resampling methods for a group of commonly-used two-level models. We illustrate its use with data from the Canadian National Longitudinal Survey of Children and Youth (NLSCY).

Key Words: Multilevel modeling, Resampling methods, Design-based variance, SAS, HLM6.

1. Introduction

Multilevel modeling is often used to analyze hierarchical data from surveys with multistage sampling designs. Such designs are complex due to the stratification, clustering and unequal selection probabilities used to select the sample and also due to nonsampling problems such as coverage and nonresponse. A design-based approach is the most commonly used to analyze data from such surveys. Briefly, in contrast to the model-based approach, a design-based approach explicitly accounts for the randomness due to the survey design (a detailed definition of these two concepts is given in Binder and Roberts, 2006). In fact, survey weighting is used to produce estimates of the model parameters and design-based variance (DBV) measures the variability among estimates from possible samples selected by the same design from the same population. In practice, the most-often used methods for obtaining DBV estimates are Taylor linearization (Binder, 1983), and replication methods such as the survey bootstrap (see Rust and Rao, 1996). Over the past few years, several software packages have been designed to enable analysts to use a design-based approach when fitting multilevel models (MLMs) to survey data. However, for the variance estimation of the estimated model parameters most of them, including HLM6² (referred to as HLM in this paper) use a the model-based sandwich estimator which ignores some sample design components, such as stratification, design clustering, and weight adjustments (see Chantala and Suchindran, 2006). There is empirical evidence that the model-based sandwich estimator underestimates the DBV. Also, the design-based sandwich estimator, as implemented in MPlus, (Asparouhov, 2006) cannot still account for all weight adjustments. In addition, a researcher might not have access to detailed sample design information, such as the PSU and stratum indicators, that are necessary for sandwich estimators; instead he may have access to replicate weights. For these reasons, resampling methods such as the bootstrap method have recently been proposed as an alternative for variance estimation. Note that, currently, none of the software that can fit MLMs to survey data has implemented these resampling techniques.

In this paper, we describe a new SAS macro, BHLMSAS_V0, that has been developed to address this issue for a group of commonly-used two-level models. The macro uses HLM software to estimate both the fixed effects and the random effects parameters of the MLM (refer to Raudenbush et al., 2004, p. 9 and 10 for details on the estimation methods used by HLM), while the variances of these estimates is estimated using a resampling method programmed in SAS. We start by presenting a brief review of resampling methods used in BHLMSAS_V0, and then describe the macro. The program

² Scientific Software International, Inc. (c) 2000 techsupport@ssicentral.com, www.ssicentral.com.

has been applied to analysis of data from several surveys (e.g., the Canadian National Longitudinal Survey of Children and Youth (NLSCY), and the Canadian Workplace and Employee Survey (WES)). In this paper we illustrate its use with data from the NLSCY.

2. Resampling methods for MLMs

Software packages³ capable of fitting MLMs to survey data have implemented a model-based sandwich estimator. It considers only the variation among the model clusters (level-2 units in two-level model). Note that the relationship between the model hierarchy and the sample design hierarchy can be one of the following:

- i) Model hierarchy is nested in that of the survey design;
- ii) Model hierarchy coincides with the hierarchy of the survey design;
- iii) Model hierarchy is completely different from the hierarchy of the survey design.

For situations i) and ii) assuming a small sample fraction of clusters, but still a large number of them (more than 50), when the survey weights are non-adjusted probability weights, the design-based sandwich estimator, as well as the usual model-based sandwich estimator perform well (Pfeffermann et al., 1998). However, in general, these estimators tend to underestimate the DBV, partly because the sampling weights incorporated into the calculations, instead of being the inverses of the probabilities of selection, contain nonresponse and calibration adjustments (Kovacevic, et al., 2006).

For the same situations i) and ii) the resampling methods such as the bootstrap methods that are briefly described below seem to have better statistical properties (Kovacevic et al. 2006). Many commonly used resampling methods consist of subsampling the initial or original sample. Then the process is repeated B times, creating B new samples (or replicates). Weights are recalculated for each of the B samples applying the same procedures (e.g., non-response and calibration adjustments) as were applied to the original sample. The B sets of weights are called the resampling or replication weights.

For situations i) and ii) the resampling variance estimates can be solely based on the variability of the units at the highest level of the sample design (called primary sampling units (PSUs) in multi-stage sampling designs) under certain assumptions. Situation i) can occur when analyzing repeated measures data from a longitudinal survey such as the NLSCY when the highest model level is the person, and the repeated measures are observations from different survey cycles, which are nested within the person (see section 4); but the person is nested within the PSU which is the highest design unit. The Workplace and Employment Survey (WES) provides an example of the second situation, where the model hierarchy coincides with that of the survey design, if the highest model level is the workplace/employer (which is the PSU too), and the first model level is the employee. The survey bootstrap weights used are computed using the usual method of Rao, Wu and Yue (1992).

Notice that Grilli and Pratesi (2004) and Kovacevic et al. (2006) proposed bootstrap methods for MLM that are modifications of usual survey bootstrap methods. For all of these methods, the authors assume that the sampling weights are available for every hierarchical level of the model. However, in reality it is seldom the case, and some approximations of the weights, especially at higher levels, are needed (see Kovacevic and Rai, 2003).

It should be noted that the simulation studies using resampling methods developed for relationships i) and ii) show that they generally yield good results (Grilli and Pratesi, 2004, Kovacevic et al, 2006). However, additional research into the statistical properties of the resulting estimates is needed.

A typical example of the third situation for which the two hierarchies are different involves examination of the school effect on the academic performance of students selected by a household survey. Here, the model hierarchy is students within schools. However, the design hierarchy could have geographic regions as PSUs and students selected from these, so that schools are not necessarily nested within PSUs. To our knowledge, in this situation, no design-based variance estimation method has yet been developed at the time of writing of this paper.

With the program presented in this paper, we assume that situation (i) or (ii) is true with respect to the relationship between the hierarchy of the model being fitted to the data and the hierarchy of the survey design. We also assume that

³ MPLUS, LISREL, Stata (GLLAMM), MLWIN and HLM.

the replication method used to produce the replication weights is appropriate for the survey design and that the resampling variance estimator has the form:

$$v_{RS}(\hat{\theta}) = \frac{C}{B} \sum_{b=1}^B (\hat{\theta}_b - \hat{\theta})^2 \quad (1)$$

where $\hat{\theta}$ is an estimate of a fixed parameter or of a random effects parameter obtained using the full-sample weights, $\hat{\theta}_b$ is its estimate obtained by using the b -th set of replicate weights; and B corresponds to the number of replicates. Note that estimator (1) can be used for the Rao-Wu-Yue survey bootstrap and for balanced repeated replication variance estimation provided that the replicate weights are obtained properly, and provided that C is set equal to 1; if mean bootstrap weights are used, then C is the number of bootstrap samples combined to produce each mean bootstrap weight variable (see Chowhan and Buckley, 2005).

3. Description of BHLMSAS_V0

BHLMSAS_V0 is a SAS macro program that can be used to estimate variances under the design-based approach using the resampling methods mentioned in the previous section. Either situation i) or ii) should exist between the hierarchical structure of the survey design and of the model. This macro is based on the HLM software package for estimating the parameters of the multilevel model. The program calls HLM within SAS to calculate $\hat{\theta}$ and $\hat{\theta}_b$ ($b=1$ to B); it then uses SAS to estimate the variance of $\hat{\theta}$ using the formula in equation (1). This macro is easy to use, but it is assumed that the user is familiar with HLM to be able to specify the model for which the estimate of the variance of the estimated parameters has to be determined. As well, users require a minimum knowledge of SAS in order to provide the program's input parameters.

This macro was designed for estimating the variance of estimated parameters in two-level hierarchical models where the distribution of the variable of interest is either normal distribution or Bernoulli.

The program is divided into three sections. The first is used to define directories, files and variables that will be needed. The second section lets users define the parameters of the model to be fitted, and the third contains the key code of the program that the user should not modify. Due to space constraints, only the first two sections are described in detail below. Interested readers may contact the authors (fritz.pierre@statcan.gc.ca or abdelnasser.saidi@statcan.gc.ca) for a full copy of BHLMSAS_V0.

Section 1: Defining the directories, file names and variables

In this section, users must define the following SAS macro variables that are used to identify directories, file names and variables: *HLMPTCH*, *HLMTEMP*, *OUTRST*, *NRST*, *DAFILE*, *DBFILE*, *Afile*, *BfileL1*, *BfileL2*, *IDENT*, *fwgtL1*, *fwgtL2*, *bswL1*, *bswL2*, *BOOT*, *C*, *outc*, *nonlin*, *covL1* and *covL2*.

HLMPTCH: specifies the installation path for HLM. For example, `%let HLMPTCH = C:\Program Files\HLM6`

*HLMTEMP*⁴: is used to indicate the directories where the temporary files are saved. There cannot be any spaces in its path and name. For example `%let HLMTEMP=F:\DARC\GCM\JSM_article\NLSCY1` is not valid. It should instead be `%let HLMTEMP=F:\DARC\GCM\JSM_article\NLSCY1`.

OUTRST and *NRST*: are used, respectively, to specify the directory and name of the file for saving results; that is to say estimates of the parameters, standard errors, p-values, confidence intervals and the number of replications that yielded meaningful results (see *BOOT* for a description of meaningful results).

⁴ Three sub-directories are created here automatically by the program: MDM, MLM and RST. If these sub-inventories exist, the program will erase all of the files in them before generating new ones. Every sub-inventory contains a certain number of files corresponding to the original sample (with the suffix 0) and to every replicate, some of which may be consulted. For example, users may consult the results of the model produced by HLM for a given replicate by opening one of the *.txt* files (e.g. *OUTB_2.txt* contains the results for the second replicate) in the RST sub-inventory.

DAFILE and *DBFILE*: identify directories containing, respectively, the analysis file (*Afile*) and the replication files at levels 1 and 2 (*BfileL1* and *BfileL2*).

Afile: represents the name of the analysis file (must be a SAS file). It must contain all of the variables needed for the analysis, the design weight variables for the full (original) sample for the first (if applicable) and second (if applicable) hierarchical levels of the model, respectively, and the level-2 *IDENT* (described below) unique identifier records in the sample. In order to reduce the program's run time, it is recommended that unnecessary variables not be retained, and to only retain records that are part of the population (e.g. individuals aged 12 or over) to which the analyses are being applied, if applicable.

BfileL1 and *BfileL2*: are used to designate the files (must be SAS files) that contain level-1 and 2 replicate weights, respectively. If the weight is only available for a single level (meaning that it is equal to 1 for all units of the other level, thus, there is no need for a weight variable at that level) then the macro variable for the other level must be left at *none* (for example `%let fwgtl1 = none; %let bswl1 = none;`) if there is no weight at the level-1

IDENT: is used to identify the unique identifier of level-2 records. It is important not to add the identifier of level-1 records here. This variable must be of character type and no longer than 12 characters.

fwgtL1 and *fwgtL2*: are used to designate the design weight variables for the full sample for hierarchical levels 1 and 2 for the model, respectively; they must appear in the *Afile*. The macro variable of a given level is left at *none* if there is no weight for it. The design weight for level-2, which can be indicated by w_j is the inverse of the probability of inclusion of the level-2 unit j , which may have been multiplied by adjustment factors (e.g. non-response and post-stratification). The one for level-1 (w_{ij}) is the inverse of the probability of inclusion of the level-1 i unit selected within the level-2 j unit; this weight may also have been multiplied by adjustment factors. To reduce the bias in the estimates, especially for the smaller samples, it is recommended that the w_{ij} weight be adjusted for scale (standardized); the second standardization method of Pfeffermann et al. 1998 is the one automatically used in HLM (see Raudenbush et al. 2004, pages 56 and 57). Thus, it is not necessary for the design weights (reported at *fwgtL1* and *fwgtL2*) and the associated bootstrap weights (reported at *bswL1* and *bswL2*) to be adjusted for scale.

bswL1 and *bswL2*: designate the prefixes of replicate weight variables for both hierarchical levels; they may not contain more than four alphanumeric characters, the first of which must not be numeric. Thus, the names of these replicate weight variables have to be of the *xxxx#* form, where *xxxx* is the prefix and *#* is the number of the replicate (e.g. *BL1_500* and *BL2_500* could be the name of the level-1 and 2 weight variables, respectively, for replicate number 500). The macro variable for a level is left at *none* if there is no replicate weight for it, as for the main sample weights.

BOOT: indicates the number of replicates that should be used to estimate the variances. It should be noted that the program was designed to eliminate by default the replicates that did not yield useful results. For example, take the case where we chose 100 replicates (`%let BOOT=100`); suppose that the program does not yield useful results for x replicates. There will then be $B=100-x$ replicates that will be used to estimate the variances.

C: indicates the replication method: 1 for the bootstrap or BRR method and >1 to indicate the mean bootstrap weights method. In this case, *C* also indicates the number of bootstrap samples used to produce each mean bootstrap weight variable (e.g., *C* is equal to 50 in the case of the WES).

Outc and *nonlin*: identify the name of the variable of interest and its distribution, respectively. For the normal distribution, we use `%let nonlin=n`; or `%let nonlin=b` in the case of the Bernoulli distribution.

covL1 and *covL2*: are used to report the names of all possible covariates⁵ at levels 1 and 2, respectively, to be retained in the temporary files while running the program. It is also necessary to indicate their number after the / sign (e.g. `%let covL1= covl1_1+covl1_2/2; %let covL2= covl2_2/1;`).

Section 2: Specification of the two-level model

⁵In this paper, we use the terms predictor and covariate interchangeably to designate the independent (or explanatory) variables in the models.

In this section, users define the two-level hierarchical model to be fit by specifying one equation for the level-1 and up to ten equations for the level-2. They must also define two more macro variables, one of which will be used to fix a convergence criterion and the other to establish the confidence level for calculating the confidence intervals. Thus, the following macro variables have to be defined: *L1model*, *L2INTC*, *L2SLOP1* to *L2SLOP9*, *NBITER* and *alpha*.

L1model: helps to define the level-1 equation of the model and the number of covariates included in it. The terms of the equation are separated by the “+” sign, and are separated from the number of level-1 covariates by a “/” (e.g. `%let L1model= INTRCPT1 + cov11_1 + cov11_2/2;`). It should be noted that the random term, i.e. the level-1 error term, is included by default in the program and should not be specified in the *L1model* statement. As well, it is understood that the term *INTRCPT1* represents the intercept for the model, if one is to be included, but does not contribute to the count of the number of covariates. Up to nine level-1 covariates may be included in the level-1 equation.

L2INTC: if an intercept is included in the level-1 equation (e.g. `%let L1model= INTRCPT1 + cov11_1/1;`), the *L2INTC* macro variable can be used to model it as a function of the level-2 covariates. Otherwise, it will be ignored by the program `%let L2INTC=none;`). As with the level-1 equation, up to nine level-2 covariates can be included; to include an intercept in *L2INTC*, insert the term *INTRCPT2* immediately after the equal sign. Unlike the *L1model*, this equation may include or exclude a random term. In other words, users must indicate whether the level-1 intercept is random (e.g. `%let L2INTC= INTRCPT2 + cov2_1+ random;`) or not (e.g. `%let L2INTC= INTRCPT2 + cov2_1;`).

L2SLOP1 to *L2SLOP9*: as with the *L2INTC*, these macro variables can be used to model the slopes of the level-1 covariates included in the level-1 equation. Up to nine level-2 predictors can be included in each, as well as one intercept and one random term. However, for technical reasons, it is important to note that the total number of level-2 random terms, including the intercept (see *L2INTC* defined above), if applicable, must be between one and three. In other words, *BHLMASAS_V0* will not allow a model to run without at least one random (level 1) coefficient (e.g. `%let L1model = INTRCPT1/0; %let L2INTC= INTRCPT2+cov2_1+cov2_2;`). Nor will it allow a model with more than three such terms to run (e.g. `%let L1model=INTRCPT1+cov11_1+cov11_2+cov11_3/3; %let L2INTC= cov2_1+random; %let L2SLOP1= INTRCPT2+cov2_1+random; %let L2SLOP2=cov2_2+random; L2SLOP3=cov2_3+random; L2SLOP4=none; L2SLOP5=none; ..., L2SLOP9=none;`). If one or more of these macro variables are not used they will be set to none as in this previous example.

NBITER: indicates in the program the maximum number of iterations and what to do if the model fails to converge before this number (because the estimation techniques for the MLM parameters are iterative procedures). For example, when specifying `%let NBITER=10;`, if there is no convergence after 10 iterations, the program will ask whether it should continue or stop iterating (enter y for “yes” and n for “no”). (See comments in the program for the other options).

Alpha: used to modify the level of significance for the confidence levels: the level of significance used by default in the outputs produced is 5% (`%let Alpha=0.05;`). To modify this level to calculate the confidence intervals, just change the default value of *Alpha* by specifying `%let Alpha=desired;` level. For example, to produce confidence intervals of 90%, `%let Alpha=0.10;` would be specified.

4. Example

In this section, we illustrate the use of *BHLMASAS_V0* with an example based on data from the NLSCY.

Growth curve model:

This example was motivated by the article by Letourneau et al. (2006) entitled “Longitudinal study of postpartum depression, maternal-child relationships, and children’s behaviour at 8 years of age”. In this study, the authors address, among other things, the following research question: Do children of mothers who experienced postpartum depression have higher levels of anxiety, which then increases over time, than those whose mothers did not experience this condition?

To answer this question, we used the data from the first four biennial cycles of the NLSCY. More specifically, we followed to a maximum of 72-96 months of age the children who were aged 0 to 24 months in cycle 1. This produced a sample of 3,533 children with a total of 9,403 observations after a few additional exclusions. This yields an average of

almost three observations per child. For additional details on the data used and the definition of the variables, refer to Letourneau et al. (2006). The question of interest stated above can be translated into the growth curve model shown below:

Growth curve model

$$\begin{aligned} \text{Level 1 equation: } & \text{Anxiety}_{ij} = \beta_{0j} + \beta_{1j} \text{AGEY}_{ij} + \varepsilon_{ij}, \\ \text{Level 2 equations: } & \begin{cases} \beta_{0j} = \gamma_{00} + \gamma_{01} \text{DICHDEPR}_j + u_{0j}, \\ \beta_{1j} = \gamma_{10} + \gamma_{11} \text{DICHDEPR}_j + u_{1j}, \end{cases} \end{aligned} \quad (2)$$

where i represents a cycle of the NLSCY ($= 1, 2, 3, \text{ or } 4$) and j refers to the child,

$$\varepsilon_{ij} \sim N(0, \sigma^2), (u_{0j}, u_{1j})' \sim N\left((0, 0), \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{bmatrix}\right), \text{ and } \varepsilon_{ij} \perp u_{0j} \text{ and } u_{1j},$$

Notes: Anxiety is a continuous score variable taking values in the range 1-3

age (AGEY) is in years and centred at 2,

$$\text{DICHDEPR} = \begin{cases} 1 & \text{if the mother had postpartum depression (derived from a score)} \\ 0 & \text{if not} \end{cases}$$

As indicated in section 2, in order to apply a replicate method in estimating the variances for the estimated parameters of this model, we need design weights for both hierarchical levels. In this case, since the cycles or “ i ” occasions are not sampled for the “ j ” children, we assume that the design weight for the level-1 units is one ($w_{ij}=1$); for the level-2 unit design weights, we use the longitudinal weight provided for cycle 1 of the survey ($w_j = w_{1j}$)⁶. We use the corresponding bootstrap weight file, which will be merged with the *Afile* in the program, using the unique identifier for the children (PERSRUK); we use the 1,000 available bootstrap replicates. It should be noted that the underlying replicate method for this model is that of Rao, Wu and Yue (1992) because this is the one that was used in the NLSCY to produce the bootstrap weights.

The input parameters for the program are:

Section 1

```
%let HLMPTCH = C:\Program Files\HLM 6;
%let HLMTEMP = F:\DARC\GCM\RDC_ITBarticle\NLSCY;
%let OTRST = F:\DARC\GCM\RDC_ITBarticle\NLSCY;
%let NRST = Example1;
%let DAfile = "F:\DARC\GCM\RDC_ITBarticle\NLSCY ";
%let DBfile = "F:\DARC\GCM\RDC_ITBarticle\NLSCY ";
%let Afile = Afile;
%let Bfile1 = none;
%let Bfile2 = BVC1_11a;
%let IDENT = PERSRUK;
%let fwgt1 = none; * this is equivalent to a weight of one;
%let fwgt2 = SAMPWT;
%let bsw1 = none;
%let bsw2 = BSW;
%let BOOT = 1000;
%let C = 1;
%let outc = ANXIETY;
%let nonlin = n;
%let covL1 = AgeY/1;
%let covL2 = DICHDEPR/1;
```

⁶The NLSCY also provides a transversal weight, which is different from the longitudinal weight that has to be used in these kinds of analyses.

Section 2

```
%let L1model = INTRCPT1+AgeY/1;
%let L2INTC = INTRCPT2+DICHDEPR+random;
%let L2SLOP1 = INTRCPT2+DICHDEPR+random;
%let L2SLOP2 = none;
...
%let L2SLOP9 = none;
%let NBITER = 100;
%let alpha = 0.05;
```

The results are presented in Table 1, below. The first two columns contain the parameters and their estimates, respectively. The following two columns contain the standard errors and the p-values (based on the Wald test)⁷. The confidence intervals are in columns five and six. Finally, the number of replicates used in the variance calculations appears in the last column. Thus, we can conclude that children whose mothers suffered from postpartum depression have a higher level of anxiety ($\hat{\gamma}_{01} = 0.0811$ and p-value = 0) than those whose mothers did not. However, there is no significant difference in the rate of increase of the anxiety as a child grows older, that is to say the slope does not differ ($\hat{\gamma}_{11} = 0.0001$ and p-value = 0.9834) between the two groups.

Table 1: Results

<i>Parameter</i>	<i>Estimate</i>	<i>Standard error</i>	<i>p value (Z-tests)</i>	<i>Lower limit conf. int. 95%</i>	<i>Upper limit conf. int. 95%</i>	<i>Number of replicates used</i>
<i>Fixed Effect</i>						
<i>For INTRCPT1, B0</i>						
<i>INTRCPT2:G00</i>	1.1705	0.0085	0.0000	1.1538	1.1871	1000
<i>DICHDEPR:G01</i>	0.0811	0.0186	0.0000	0.0446	0.1177	1000
<i>For AGEY slope, B1</i>						
<i>INTRCPT2:G10</i>	0.0263	0.0026	0.0000	0.0211	0.0315	1000
<i>DICHDEPR:G11</i>	0.0001	0.0059	0.9834	-0.0114	0.0117	1000
<i>Random Effect</i>						
<i>var(INTRCPT1)</i>	0.0165	0.0045	0.0002	0.0078	0.0253	1000
<i>cov(AGEY,INTRCPT1)</i>	-0.0015	0.0012	0.2204	-0.0038	0.0009	1000
<i>var(AGEY)</i>	0.0018	0.0004	0.0000	0.0009	0.0026	1000
<i>Residual variance</i>	0.0459	0.0022	0.0000	0.0416	0.0502	1000

5. Conclusion

In this paper we presented a program that produces design-based variance estimates for some commonly-used two-level MLMs. These DBV estimates were calculated using resampling methods that are an alternative to the model-based sandwich estimator implemented in several software packages that fit MLMs. The program is based on a joint use of the SAS and HLM software programs. It has been used to fit MLMs to data from several Statistic Canada surveys such as NLSCY and WES. While the statistical properties of these variance estimates still need to be investigated, this program is very promising as a tool for using resampling methods with MLMs fitted to survey data. Future improvements of the program includes: allow more covariates at level-1 and level-1 and more random effects; the fitting of three-level MLLs, the DBV to be estimated by jackknife method; and more accurate tests for the dispersion parameters.

Acknowledgments

We are grateful to Dr. M. Kovacevic and Dr. G. Roberts for their valuable comments, and helpful suggestions in writing this paper.

⁷It is generally not recommended to opt for the Wald tests for dispersion parameters ($\sigma^2, \tau_{00}, \tau_{01}, \tau_{11}$).

References

- Asparouhov, T. (2006). General Multi-Level Modelling with Sampling Weights. *Communications in Statistics –Theory and methods*, 35, 439-460.
- Asparouhov, T. (2004). Weighting for unequal probability of selection in multilevel modeling, Mplus Web Notes No. 8 available from <http://www.statmodel.com/>.
- Binder, D.A. (1983). On the variance of asymptotically normal estimators from complex surveys. *International Statistical Review*, 51, 279-291.
- Binder, D. and Roberts, G. (2006). Approaches for Analyzing Survey Data: a Discussion. *ASA Proceedings of Survey Research Methods Section*, pp. 2711-2778.
- Chantala, K. and Suchindran, C. (2006). Adjusting for Unequal Selection Probability in Multilevel Models: A Comparison of Software Packages. Proceedings of the American Statistical Association, Seattle, WA: American Statistical Association.
- Chowhan, J. and Buckley, N. J. (2005). Using Mean Bootstrap Weights in Stata: a BSWREG Revision. The Research Data Centres Information and Technical Bulletin. Statistics Canada, Catalogue No. 12-002-XIE, vol. 1, no 1.
- Grilli, L. and Pratesi, M. (2004). Weighted estimation in multilevel ordinal and binary models in the presence of Informative sampling designs. *Survey Methodology*, vol. 30, no 1, pp. 104-114.
- Kovacevic, M., Huang, R. You, Y. (2006). Bootstrapping For Variance Estimation in Multi-Level Models Fitted To Survey Data. ASA Proceedings of the Survey Research Methods Section, pp. 3260-3269.
- Kovacevic, M. S. and Rai, S. N. (2003). A Pseudo Maximum Likelihood Approach to Multilevel Modelling of Survey Data. *Communications in Statistics: Theory and Methods*, 32 p. 103-121.
- Letourneau, N.L., Fedick, C.B., Willms, J.D., Dennis, C.L., Hegadoren, K., & Stewart, M.J. (2006). Longitudinal study of postpartum depression, maternal-child relationships, and children's behaviour at 8 years of age. Devore (Ed.). *Parent-child relations: New research*, Pp. 1-19. Hauppauge, NY: Nova Science Publishers.
- Rao, J. N. K, Wu, C.F.J. and Yue, K. (1992). "Some recent work on resampling methods for complex surveys." *Survey Methodology*, Vol. 18, no 2, pp. 225-234.
- Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of Royal Statistical Society B* 60:23–40.
- Raudenbush, S., Bryk, A, Cheong, Y. F., Congdon R. and Du Toit, M. (2004). *HLM 6: Hierarchical Linear & Nonlinear Modeling*. Scientific Software International, Inc. 7383 North Lincoln Avenue, Suite 100.Lincolnwood, IL 60712-1747.
- Rust, K.F. and Rao, J. N. K (1996). Variance estimation for complex surveys using replicate techniques. *Statistical Methods in Medical Research* 5, 283-310.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1-25