Extended Bootstrap Bias Correction with Application to Multilevel Modelling of Survey Data under Informative Sampling

Solange Corrêa
Department of Methods and Quality, Brazilian Institute of Geography and Statistics, Rio de Janeiro, 20031-170, Brazil

Abstract
Data available to analysts are often obtained from complex sample surveys in which population units are selected by stratified multi-stage designs with unequal selection probabilities. Unweighted estimators of the model parameters may be severely biased in such cases if the selection probabilities are related to the outcome values even after conditioning on the model covariates (informative sampling). Probability weighting reduces the bias very significantly but does not eliminate it, unless the sample sizes at the various levels of the model hierarchy are very large. In this paper we propose a general approach for bias correction based on resampling procedure. We assess the performance of our proposed approach by an extensive simulation study using probability weighted estimators of two-level model parameters when fitting complex survey data under informative sampling designs. The proposed method showed to be effective in bias reduction in all the cases considered.
Key Words: Bias correction, Bootstrap, Sample distribution, Probability weighting.

1. Introduction

Unweighted multilevel analysis (Goldstein, 2003) of complex survey data may lead to severely biased estimates (Korn and Graubard, 1995) if the inclusion probabilities are related to the model response variable even after conditioning on the design variables, known in the sampling literature as *informative sampling design* (Pfeffermann, Krieger and Rinott, 1998). Under such schemes, the model holding for the population values is likely to be different from the model holding for the sample data, defined as *sample model* by Pfeffermann *et al.* (1998a). Therefore, the sample model needs to be estimated from the sample data in order to perform inferential statistical analyses based on the sample values.

Another important issue when fitting multilevel models to sample survey data is how to account for the sampling weights in multilevel analysis estimation. A large number of studies on how to do this have been proposed lately in the literature (Pfeffermann, Skinner, Holmes, Goldstein and Rabash, 1998); Korn and Graubard, 2003; Grilli and Pratesi, 2004; Rabe-Hesketh and Skrondal, 2006). Most of them are based on incorporating the sampling weights in the likelihood function and maximising it via numerical integration since closed expression for the estimators are not available. Pfeffermann *et al*. (1998b) propose a probability weighted iterative generalised least squares approach (PWIGLS), which is an adaptation of the iterative generalised least squares (IGLS) method (Goldstein, 1986) by analogy to the pseudo maximum likelihood principle (Binder, 1983; Skinner, 1989; Chambers, 2003). The PWIGLS approach basically consists of probability weighting of first and higher level units with weights equal to the reciprocal of the corresponding sampling inclusion probabilities. However, as shown in that article, the use of this approach, although reducing the bias of unweighted parameter estimators very substantially, does not eliminate it completely, unless in large samples.

Classical bootstrap bias corrections (Efron, 1979) involve estimating the bias of an estimator by apriorily chosing a function that depends only on the original and bootstrap estimates of the parameter of interest. In analyses that involve more than one parameter, it could well be that the bias of an estimator in estimating one parameter may depend on the value of that parameter and on the bias in estimating the other parameters.
This article proposes a general approach for bias correction, entitled the Extended Bootstrap Bias Correction (EBS), based on the bootstrap resampling procedure and on a parametric model.

In Section 2 we describe the classical bootstrap bias correction methods. The EBS approach is presented in Section 3. In Section 4 we compare the performance of the EBS method to classical bootstrap bias corrections via an extensive Monte Carlo study. In this experiment, we consider unweighted (thereafter *naïve*) and PWIGLS estimators when fitting two-level models to survey data under informative sampling of first units with small sample sizes at both levels. Conclusions and Remarks are presented in Section 5.

2. Classical Bootstrap Bias Corrections

Efron (1979) proposed to estimate the bias by use of bootstrap samples as obtained by drawing units with replacement from the original sample. These resampling methods have become very popular in statistical inference and are applied in many diverse applications in order to obtain estimates of standard errors, confidence intervals, biases, etc. (Shao and Tu, 1995). In what follows, parametric and nonprametric bootstrap are reviewed along with the classical bootstrap bias corrections.

Let $z_1,...,z_n$ be the outcomes of independent and identically distributed (i.i.d.) random variables $Z_1, Z_2,..., Z_n$ having distribution F. Denoting the observed data by $\underline{z} = (z_1,..., z_n)$, the objective is to assess the accuracy of a statistic $\hat{\psi} = t(\underline{z})$ in estimating the unknown parameter of interest $\psi = t(F)$.Let $\underline{z}_1^*,..., \underline{z}_B^*$ be B independent (parametric or nonparamentric) bootstrap samples and $\hat{\psi}_1^*,...,\hat{\psi}_B^*$ the corresponding bootstrap replications of the statistic $\hat{\psi}$, where $\hat{\psi}_i^* = t(\underline{z}_i^*)$. Thus, measures of accuracy of the statistic of interest are inferred from the observed values of the bootstrap replications $\hat{\psi}_1^*,...,\hat{\psi}_B^*$. In particular, the bootstrap estimation of bias is straightforward, as shown in Efron and Tibshirani (1986) and described as folows. The bias of the statistic $\hat{\psi} = t(\underline{z})$ in estimating the true value $\psi = t(F)$ is

$$bias_F = bias_F(\hat{\psi}, \psi) = E_F[t(\underline{z})] - t(F) \qquad (1)$$

where $E_F[\cdot]$ is the expectation under the distribution F. Replacing F by the estimated distribution $\hat{F}$ in equation (1), we find the bootstrap estimate of bias:

$$bias_{\hat{F}} = E_{\hat{F}}\left[t(\underline{z}^*)\right] - t(\hat{F}). \qquad (2)$$

In practice, $E_{\hat{F}}\left[t(\underline{z}^*)\right] = E_{\hat{F}}\left[\psi^*\right]$ is approximated by averaging $\hat{\psi}_1^*,...,\hat{\psi}_B^*$ over a large number B of bootstrap replications yielding

$$\hat{b}ias_B = \overline{\hat{\psi}}^* - t\left(\hat{F}\right) \tag{3}$$

where $\overline{\hat{\psi}}^* = B^{-1}\sum_{b=1}^{B}\hat{\psi}_b^*$. As B tends to infinity, $bias_B$ tends to $bias_{\hat{F}}$ (Efron and Tibshirani, 1986).

Once an estimate of the bias is available, one can correct the original estimate by subtracting the estimated bias from it. Hence, the bootstrap bias-corrected estimate of the parameter of interest ψ, also known as additive correction, is given by

$$\hat{\psi}^{BC} = \hat{\psi} - \hat{b}ias_B = \hat{\psi} - \left[\overline{\hat{\psi}}^* - t\left(\hat{F}\right)\right] = 2\hat{\psi} - \overline{\hat{\psi}}^* . \tag{4}$$

Similarly, the multiplicative bias correction (Hall and Maiti, 2006) is given by $\hat{\psi}^{BC} = \hat{\psi}^2 / \overline{\hat{\psi}}^*$.

3. The Extended Bootstrap Bias Correction (EBS) Approach

The main idea of the EBS approach is to use data generated under an assumed model and a plausible parameter space to identify the relationship between the true parameter value and its estimates from the original and bootstrap samples. Hence, the functional relationship between the error of the estimator under study and its original and bootstrap estimates is extracted from the data themselves, rather than arbitrarily chosen. Besides, not only original and bootstrap estimates of the target parameter are included in that relationship but corresponding estimates of other model parameters can possibly be included in the function as well. To allow for the fact that the bias may depend on the true value of the parameter, the procedure explicitly takes into account a set of plausible parameter values in the process of identification of the function.

The EBS approach is motivated by the classical bootstrap bias corrections and has two main advantages. The first one is that it provides not only a bias-corrected estimator of the target parameter but also the bootstrap distribution of the bias-corrected estimator, allowing estimation of its measures of accuracy. The second advantage is that the EBS approach is not restricted to a particular bias correction formula, permitting to express the bias of the target estimator as a function of the biases of other estimators involved in the analysis.

Let ξ denote a superpopulation model with density function $f_\xi(\mathbf{z};\boldsymbol{\psi})$ for an *original sample* $\mathbf{z} = \left(z_1,..., z_n\right)$ and a K-dimensional parameter of interest $\boldsymbol{\psi} = (\varphi_1,...,\varphi_K)$. Let $\hat{\boldsymbol{\psi}}$ denote the 'original' estimate of $\boldsymbol{\psi}$ from the original sample. Assume that parametric (or nonparametric) bootstrap samples are drawn from the original sample yielding the average over the bootstrap estimates $\overline{\hat{\boldsymbol{\psi}}}^*$, also referred to as 'bootstrap mean'.

Consider a single component φ_k of $\boldsymbol{\psi}$ for now, $k = 1,..., K$. The aim is to estimate the error of $\hat{\varphi}_k$, in estimating φ_k, i.e. $\hat{\varphi}_k - \varphi_k$, by identifying the relationship between $\hat{\varphi}_k - \varphi_k$ and other potential factors that may be related to it, such as the original and bootstrap estimates of *all* parameters involved in the analysis, i.e., $\hat{\boldsymbol{\psi}}$ and $\overline{\hat{\boldsymbol{\psi}}}^*$. Once this functional relationship is identified, a bias-corrected estimator of φ_k is obtained by applying this function to the original sample. Since the bias corrections holding for estimators of different parameters do not have to be the same, the EBS approach is applied to each target parameter separately.

To identify the functional relationship, a large number of values of $\boldsymbol{\psi}$, $\hat{\boldsymbol{\psi}}$ and $\overline{\hat{\boldsymbol{\psi}}}^*$ is necessary, corresponding to the plausible parameter values and respective original and bootstrap estimates. In practice, however, the original sample produces only one value of $\hat{\boldsymbol{\psi}}$ and $\overline{\hat{\boldsymbol{\psi}}}^*$, and there is only one true (unknown) parameter value $\boldsymbol{\psi}$. The range of values is generated by mimicking the random process that originated the original estimate and bootstrap mean, $\hat{\boldsymbol{\psi}}$ and $\overline{\hat{\boldsymbol{\psi}}}^*$, from the original sample.

The idea of the EBS method is to generate a set of plausible parameter values $\boldsymbol{\psi}_1,...,\boldsymbol{\psi}_g,...,\boldsymbol{\psi}_G$ based on the original sample estimate and, for each of those, to generate one *pseudo original sample* from $f_\xi(\mathbf{z};\boldsymbol{\psi}_g)$. Bootstrap samples are then generated (parametrically or nonparametrically) from this pseudo original sample. As a result, each *generated parameter value* generates one *pseudo original estimate* and *corresponding bootstrap estimates*. A mathematical relationship for the bias of the estimator under study can then be identified. A bias-corrected estimator for the target parameter is obtained by applying this function to the original and bootstrap estimates obtained from the original sample. The EBS method is described in six steps as follows:

Step 0 - Obtaining Original and Bootstrap Estimates

Assume that the original sample yields original and bootstrap estimates, $\hat{\boldsymbol{\psi}}$ and $\overline{\hat{\boldsymbol{\psi}}}^*$, of the parameter $\boldsymbol{\psi}$. These values are used later in step 5.

A single component of the vector of parameters $\boldsymbol{\psi}=(\varphi_1,...,\varphi_K)$, say φ_1, is assumed to be the target for bias-corrected estimation. The approach is applied to the other model parameters in an identical manner.

Step 1 – Generation of Plausible Parameter Values

A set of G plausible parameter values $\underline{\varphi}_1=(\varphi_{1,1},...,\varphi_{1,G})$ for φ_1 is built according to a process that will be specified in section (4.2.), item II.iii). Similarly for φ_2, obtaining $\underline{\varphi}_2=(\varphi_{2,1},...,\varphi_{2,G})$, and so on. Define $\boldsymbol{\psi}_1,...,\boldsymbol{\psi}_G$ K-dimensional vectors of plausible parameter values for $\boldsymbol{\psi}$, where $\boldsymbol{\psi}_1=(\varphi_{1,1},\varphi_{2,1},...,\varphi_{K,1}),\ldots,\boldsymbol{\psi}_g=(\varphi_{1,g},\varphi_{2,g},...,\varphi_{K,g}),\ldots,\ \boldsymbol{\psi}_G=(\varphi_{1,G},\varphi_{2,G},...,\varphi_{K,G})$.

Remark 1: This step could have a Bayesian interpretation but this is not considered in this research since the objective here is to obtain a range of plausible values for the parameter under study and not to make inference regarding the posterior distribution of $\boldsymbol{\psi}$.

Step 2 – Calculating the Estimation Error

For each vector of parameters $\boldsymbol{\psi}_g$, one pseudo original sample is generated from the population model $f_\xi(\mathbf{z};\boldsymbol{\psi}_g)$ and pseudo original estimates $\hat{\boldsymbol{\psi}}_g=(\hat{\varphi}_{1,g},\hat{\varphi}_{2,g},...,\hat{\varphi}_{K,g})$ are obtained, $g=1,...,G$. The error of $\hat{\varphi}_1$ in estimating φ_1 is then computed for each pseudo original sample g as $Error_g(\hat{\varphi}_1)=\hat{\varphi}_{1,g}-\varphi_{1,g}$, $g=1,...,G$.

Step 3 - Estimating the Estimation Error

For each pseudo original sample obtained in step 2 a large number B of parametric bootstrap samples are generated from $f_\xi(\mathbf{z};\hat{\boldsymbol{\psi}}_g)$ and bootstrap estimates $\hat{\boldsymbol{\psi}}^*_{g,b}=(\hat{\varphi}^*_{1,g,b},\hat{\varphi}^*_{2,g,b},...,\hat{\varphi}^*_{K,g,b})$, $g=1,...,G$; $b=1,...,B$ are computed. Let $\overline{\hat{\boldsymbol{\psi}}}^*_g=(\overline{\hat{\varphi}}^*_{1,g},\overline{\hat{\varphi}}^*_{2,g},...,\overline{\hat{\varphi}}^*_{K,g})$ denote the bootstrap means from the pseudo original sample

g , $g = 1,...,G$.

An estimate of $Error_g(\hat{\varphi}_1) = \hat{\varphi}_{1,g} - \varphi_{1,g}$ for each pseudo original sample g is then given by $\hat{E}rror_g(\hat{\varphi}_1) = \overline{\hat{\varphi}}^*_{1,g} - \hat{\varphi}_{1,g}$, $g = 1,...,G$, which is also viewed as an estimate of the bias of $\hat{\varphi}_1$ (Efron and Tibshirani, 1986).

Step 4 – Identifying Bias Correction Functions

The idea is to model the error $E = \hat{\varphi}_1 - \varphi_1$ as a function of $\hat{\boldsymbol{\psi}}$ and $\overline{\hat{\boldsymbol{\psi}}}^*$ using the 'bias correction data' defined by the values $\boldsymbol{\psi}_g$, $\hat{\boldsymbol{\psi}}_g$ and $\overline{\hat{\boldsymbol{\psi}}}^*_g$, $g = 1,...,G$, from steps 1-3. A possibility is to fit a standard linear regression model (or function $h(\cdot)$) to this data, i.e.

$$E_g = \mathbf{x}^t_g \mathbf{a} + \varsigma_g \ , \ g = 1,...,G \ , \tag{5}$$

where, for each observation g , $E_g = \hat{\varphi}_{1,g} - \varphi_{1,g}$, $\mathbf{x}_g = (1, \hat{\boldsymbol{\psi}}_g , \overline{\hat{\boldsymbol{\psi}}}^*_g)^t$, $\mathbf{a}$ is the vector of regression coefficients and ς_g is the model random error.

In fact, the analyst can apply sophisticated modelling procedures and goodness of fit tools to identify the 'best' model $h^*(\cdot)$ for the data at hands by modelling $\hat{\varphi}_{1,g} - \varphi_{1,g} = h(\hat{\boldsymbol{\psi}}_g , \overline{\hat{\boldsymbol{\psi}}}^*_g)$, $g = 1,...,G$.

A possible way to choose the model $h^*(\cdot)$ is by using a validation method in which the G parameter values are split into two groups: *modelling group*, with $G - v$ values, and *validation group*, with v values. Candidate functions $h(\cdot)$ are identified based on ($G - v$) values and then validated in the validation group. Bias-corrected estimates of φ_1 and corresponding errors are then obtained for each of the v parameter values and candidate function $h(\cdot)$ considered, i.e.

$$\hat{\varphi}^{EBS}_{1,g} = \hat{\varphi}_{1,g} - h_H(\hat{\boldsymbol{\psi}}_g , \overline{\hat{\boldsymbol{\psi}}}^*_g) \quad \text{and} \quad Error_g(\hat{\varphi}_1) = \hat{\varphi}^{EBS}_{1,g} - \varphi_{1,g} \ , \ g = 1,...,v \ ;$$
$$H = 1,...,H_0 \ , \tag{6}$$

where $h_H(\cdot)$ is the H -th candidate function identified in the modelling group and H_0 is the number of candidate functions considered.

The 'best' bias correction function $h^*(\cdot)$ is the one that originates the smallest estimated bias for $\hat{\varphi}^{EBS}_1$ based on the *validation values*, i.e., $h^*(\cdot)$ is chosen such that

$$\hat{B}ias(\hat{\varphi}^{EBS}_1) = v^{-1} \sum_{g=1}^{v} | \hat{\varphi}^{EBS}_{1,g} - \varphi_{1,g} | , \tag{7}$$

an estimate of $Bias(\hat{\varphi}^{EBS}_1) = E(\hat{\varphi}^{EBS}_1 - \varphi_1)$, is minimum.

Note that the function for bias correction is completely extracted from the bias correction data generated according to the features of the original sample available for analysis. Therefore, for each original sample found in practice there will be a different bias correction function that holds for that specific data.

Step 5 - Obtaining Bias Correction for the Original Estimate

Estimates $\hat{\boldsymbol{\psi}}$ and $\overline{\hat{\boldsymbol{\psi}}}^*$ obtained in step 0 are plugged in the bias correction function $h^*(\cdot)$ identified in step 4, yielding an estimate of the bias of $\hat{\varphi}_1$ in estimating φ_1 and, therefore, a bias-corrected estimate of the target parameter φ_1 , i.e.,

$$\hat{\varphi}_1^{EBS} = \hat{\varphi}_1 - \hat{B}ias(\hat{\varphi}_1) = \hat{\varphi}_1 - h^*(\hat{\boldsymbol{\psi}}, \overline{\hat{\boldsymbol{\psi}}}^*). \tag{8}$$

It is important to emphasise that the classical bootstrap bias corrections are implicitly included in the EBS approach by appropriately choosing the vector of explanatory variables $\mathbf{x}$ in expression (5) (this is illustrated in section 4.2).

Unlike the classical bootstrap bias corrections, the EBS approach allows estimation of measures of accuracy of $\hat{\varphi}_1^{EBS}$ as an estimator of φ_1, such as mean squared error (MSE) and confidence interval (C.I.). Once the best bias correction function $h^*(\cdot)$ is chosen, the *bootstrap distribution* of the error $\hat{\varphi}_1^{EBS} - \varphi_1$ can be obtained by applying that function to the bias correction data in step 4, yielding G values of the estimation error, i.e. $\hat{\varphi}_{1,g}^{EBS} - \varphi_{1,g}$, $g = 1,...,G$ (similarly to expression (6)). Thus, for example, an estimate of the variability of the estimator $\hat{\varphi}_1^{EBS}$ is given by

$$M\hat{S}E(\hat{\varphi}_1^{EBS}) = G^{-1}\sum_{g=1}^{G}(\hat{\varphi}_{1,g}^{EBS} - \varphi_{1,g})^2 . \tag{9}$$

Equation (9) measures the variability of the EBS estimator $\hat{\varphi}_1^{EBS}$ over all plausible values of φ_1. There are several methods available for obtaining *bootstrap confidence intervals* (Davison and Hinkley, 1997). A direct method (DiCiccio and Efron, 1996) consists of obtaining a $(1-\alpha)\%$ bootstrap C.I. for the error $\hat{\varphi}_1^{EBS} - \varphi_1$ by taking the $(\alpha/2)\%$ and $(1-\alpha/2)\%$ percentiles, $p_{\alpha/2}$ and $p_{1-\alpha/2}$, respectively, of the bootstrap distribution (based on G values) of the difference $\hat{\varphi}_1^{EBS} - \varphi_1$ as the respective lower and upper bounds, i.e. $C.I.(\hat{\varphi}_1^{EBS} - \varphi_1; 1-\alpha) = (p_{\alpha/2}; p_{1-\alpha/2})$. Hence, a $(1-\alpha)\%$ bootstrap C.I. for φ_1 is given by

$$C.I.(\varphi_1; 1-\alpha) = (\hat{\varphi}_1^{EBS} - p_{1-\alpha/2}; \hat{\varphi}_1^{EBS} - p_{\alpha/2}). \tag{10}$$

In the particular case of informative sampling designs (Pfeffermann *et al.*, 1998a), application of the EBS approach requires estimation of the sample model, say $f_s(\mathbf{z}; \boldsymbol{\psi})$, from the *original sample* in order to generate the *pseudo original sample* for each plausible parameter value generated in step 2 of the method. Similar remark holds for the step 3 of the EBS approach, with the bootstrap samples from each pseudo original sample being generated from the sample model in case of an informative sampling design.

Hence, the pseudo original samples (step 2) and the corresponding bootstrap samples (step 3) can be generated either from the population model, say $f_\xi(\mathbf{z}; \boldsymbol{\psi}_g)$ (in the case of noninformative sampling, i.e. $f_s(\mathbf{z}; \boldsymbol{\psi}_g) = f_\xi(\mathbf{z}; \boldsymbol{\psi}_g)$) or from the sample model $f_s(\mathbf{z}; \boldsymbol{\psi}_g)$ (in the case of informative sampling). Nonparametric bootstrap is an alternative for step 3 (but not for step 2) of the EBS approach for both informative and noninformative sampling designs.

4. Bias Corrections of Unweighted and PWIGLS Estimators of a Two-level model Under an Informative Sampling Design

In this section, the EBS approach is applied in order to reduce the bias of unweighted ('naïve') and PWIGLS (Pfeffermann *et al.*, 1998b) estimators of two-level model parameters under an informative sampling scheme with small sample sizes of the upper level units. The study is based on simulated data adopting the same sampling scheme considered by Pfeffermann *et al*. (1998b) and under which the PWIGLS estimators showed large biases.

Since the sampling scheme adopted in this section is informative, the discussion at the end of the previous section holds for the two-level model studied here. The only particularity is that the sample model needs to be estimated for each level of the hierarchy.

In section 4.1, the population model and sampling design considered in this application are described along with the developments for estimation of the sample models. The performance of the EBS approach using parametric and nonparametric bootstrap is assessed by an extensive Monte Carlo study described in section 4.2. The results of this study are presented in section 4.3.

4.1 Population Model, Sampling Design and Estimation of Sample Models

Consider the following two-level random intercept model (denoted ξ):

$$y_{ij} \mid \beta_j = \beta_j + \varepsilon_{ij},\ i=1,...,N_j \text{ (level 1 model)}$$
$$\beta_j = \beta + u_j,\ j=1,...,M \text{ (level 2 model)}, \tag{11}$$

where u_j and ε_{ij} are independent random errors such that $u_j \overset{i.i.d.}{\sim} N(0,\sigma_u^2)$, $\varepsilon_{ij} \overset{i.i.d.}{\sim} N(0,\sigma_\varepsilon^2)$. Let $\boldsymbol{\psi} = (\beta, \sigma_\varepsilon^2, \sigma_u^2)^t$ be the vector of population parameters.

Consider a two-stage disproportional stratified clustered sampling design with informative sampling only at level 1 (elementary units). At the first stage, m level 2 units are selected by a probability proportional to size without replacement design. The measures of size are the level 2 sizes N_j, which are assumed to be uncorrelated with the random intercepts β_j, such that the sampling design is noninformative at level 2. At the second stage, level 1 units in selected level 2 unit j are partitioned into 2 strata according to whether $\varepsilon_{ij} > 0$ or $\varepsilon_{ij} \leq 0$ and simple random sampling without replacement of sizes $n_{j.1}$ and $n_{j.2}$ (assuming $n_{j.1} \neq n_{j.2}$) are drawn from stratum 1 and 2 of level 2 unit j, respectively. In this case, the first level inclusion probability is related to the level 1 random error ε_{ij} and, consequently, to the outcome y_{ij}, featuring an informative sampling design at level 1.

Pfeffermann *et al*. (1998b) considered the same first level sampling design for assessing the performance of the PWIGLS estimators, finding nonnegligible biases when the assumption of noninformativeness at level 1 was violated. Therefore, a noninformative sampling design at level 2 and an informative sampling design at level 1 are assumed in this study in order to assess the EBS approach for bias correction of the PWIGLS estimators under the worst scenario considered by Pfeffermann *et al*. (1998b). The authors did not consider the sample model in their work. Estimation of the sample model under the sampling design described above is presented below.

Let s denote the sample of level 2 units and s_j the sample of level 1 units from the selected level 2 unit j. In what follows, the sample model holding for the level 1 units in the sample s_j is estimated for the present sampling design and population model ξ.Let O_{ij} be the stratum membership indicator taking the value 1 if the level 1 unit i in level 2 unit j belongs to stratum 1 ($\varepsilon_{ij} > 0$) and zero, otherwise. Since the random errors ε_{ij} are assumed to be normally distributed with zero mean, for sufficiently large N_j the expected number of units in each stratum is $N_j / 2$. This is due to the fact that $N^*_{j.1}$, the number of units in the population of stratum 1, is such that $N^*_{1.j} \sim Binomial(N_j, 0.5)$ since each ε_{ij} is larger (or smaller) than zero with probability 0.5. Following Pfeffermann, Krieger and Rinott (1998) and Pfeffermann, Moura and Silva (2006), the sample model $f_{s_j}(\cdot)$ for the distribution of the level 1 dependent variable y_{ij} given the random intercept β_j and inclusion in the sample, is given by

$$\begin{aligned} f_{s_j}(y_{ij} \mid \beta_j, \boldsymbol{\psi}) &\overset{def.}{=} f(y_{ij} \mid i \in s_j, \beta_j, \boldsymbol{\psi}) \\ &= \Pr(i \in s_j \mid y_{ij}, \beta_j, \boldsymbol{\psi}) f_\xi(y_{ij} \mid \beta_j, \boldsymbol{\psi}) / \Pr(i \in s_j \mid \beta_j, \boldsymbol{\psi}), \end{aligned} \tag{12}$$

Note that the distribution of y_{ij} given β_j under the model ξ, $f_\xi(y_{ij} \mid \beta_j, \boldsymbol{\psi})$, follows a normal distribution with mean β_j and variance σ^2_ε.
The probability in the numerator of (12) is given by

$$\begin{aligned} &\Pr(i \in s_j \mid y_{ij}, \beta_j, \boldsymbol{\psi}) = 2n_{j.1} / N_j, \text{ if } y_{ij} - \beta_j > 0 \text{ and} \\ &\Pr(i \in s_j \mid y_{ij}, \beta_j, \boldsymbol{\psi}) = 2n_{j.2} / N_j, \text{ otherwise,} \end{aligned} \tag{12A}$$

using the fact that once both y_{ij} and β_j are known, the ε_{ij} is also known and, therefore, $\Pr(i \in s_j \mid y_{ij}, \beta_j, \boldsymbol{\psi})$ is simply the inclusion probability in the corresponding stratum. The term $\Pr(i \in s_j \mid \beta_j, \boldsymbol{\psi})$ in the denominator of (12) is given by

$$\begin{aligned} \Pr(i \in s_j \mid \beta_j, \boldsymbol{\psi}) &= \int_{-\infty}^{\infty} \Pr(i \in s_j \mid y_{ij}, \beta_j, \boldsymbol{\psi}) f_\xi(y_{ij} \mid \beta_j, \boldsymbol{\psi}) dy_{ij} \\ &= \int_{-\infty}^{\beta_j} \Pr(i \in s_j \mid y_{ij}, \beta_j, \boldsymbol{\psi}) f_\xi(y_{ij} \mid \beta_j, \boldsymbol{\psi}) dy_{ij} + \int_{\beta_j}^{\infty} \Pr(i \in s_j \mid y_{ij}, \beta_j, \boldsymbol{\psi}) f_\xi(y_{ij} \mid \beta_j, \boldsymbol{\psi}) dy_{ij} \\ &= \frac{2n_{j.2}}{N_j} \int_{-\infty}^{\beta_j} \phi(y_{ij}, \beta_j, \sigma^2_\varepsilon) dy_{ij} + \frac{2n_{j.1}}{N_j} \int_{\beta_j}^{\infty} \phi(y_{ij}, \beta_j, \sigma^2_\varepsilon) dy_{ij}, \end{aligned} \tag{12B}$$

where $\phi(y_{ij}; \beta_j, \sigma^2_\varepsilon)$ denotes the normal probability density function with mean β_j and variance σ^2_ε evaluated at point y_{ij}; also $\Pr(i \in s_j \mid y_{ij}, \beta_j, \boldsymbol{\psi})$ is equal to the sampling fraction in stratum 2 of level 2 unit j, i.e. $\frac{2n_{j.2}}{N_j}$, for $y_{ij} \le \beta_j$ (or $\varepsilon_{ij} \le 0$). Similarly for stratum 1, with corresponding probability equal to $\frac{2n_{j.1}}{N_j}$. Since each integral in (12B) is equal to 0.5, the denominator of (12) is written as

$$\Pr(i \in s_j \mid \beta_j, \boldsymbol{\psi}) = \frac{2n_{j.2}}{N_j} \int_{-\infty}^{\beta_j} \phi(y_{ij}, \beta_j, \sigma_\varepsilon^2) dy_{ij} + \frac{2n_{j.1}}{N_j} \int_{\beta_j}^{\infty} \phi(y_{ij}, \beta_j, \sigma_\varepsilon^2) dy_{ij}$$
$$= \frac{2n_{j.2}}{N_j} \cdot \frac{1}{2} + \frac{2n_{j.1}}{N_j} \cdot \frac{1}{2} = \frac{n_{j.1} + n_{j.2}}{N_j} = \frac{n_j}{N_j}, \qquad (12C)$$

which is equal to the sampling fraction at the level 2 unit j.

Replacing (12A) and (12C) in (12), the sample distribution for the level 1 measurements y_{ij} given the random intercept β_j and inclusion in the sample is

$$f_{s_j}(y_{ij} \mid \beta_j, \boldsymbol{\psi}) = f(y_{ij} \mid i \in s_j, \beta_j, \boldsymbol{\psi}) = \begin{cases} \dfrac{(2n_{j.1}/N_j)\phi(y_{ij};\beta_j,\sigma_\varepsilon^2)}{n_j/N_j}, & if\ y_{ij} > \beta_j \\ \dfrac{(2n_{j.2}/N_j)\phi(y_{ij};\beta_j,\sigma_\varepsilon^2)}{n_j/N_j}, & if\ y_{ij} \le \beta_j \end{cases} \qquad (13)$$
$$= \begin{cases} 2n_{j.1} n_j^{-1} \phi(y_{ij};\beta_j,\sigma_\varepsilon^2), & if\ y_{ij} > \beta_j \\ 2n_{j.2} n_j^{-1} \phi(y_{ij};\beta_j,\sigma_\varepsilon^2), & if\ y_{ij} \le \beta_j. \end{cases}$$

Since the sampling design is noninformative at level 2, the sample distribution for the level 2 measurements β_j given inclusion in the sample is the same as the distribution of β_j in the population, i.e., $f_s(\beta_j \mid \boldsymbol{\psi}) = f(\beta_j \mid j \in s, \boldsymbol{\psi}) = \phi(\beta_j; \beta, \sigma_u^2)$. (14)

4.2 Monte Carlo Study

In this section, the performance of the EBS approach is assessed via a Monte Carlo study for the population model, sampling scheme and sample model described in the previous section. The experiment mimics the simulation study performed by Pfeffermann *et al*. (1998b) with an additional step for adjusting the bias of unweighted and PWIGLS estimators by applying the EBS approach. The scaled 2 PWIGLS estimators proposed by Pfeffermann *et al*. (1998b) is adopted in this study.

Let $\hat{\boldsymbol{\psi}} = (\hat{\beta}, \hat{\sigma}_\varepsilon^2, \hat{\sigma}_u^2)^t$ and $\bar{\hat{\boldsymbol{\psi}}}^* = (\bar{\hat{\beta}}^*, \bar{\hat{\sigma}}_\varepsilon^{2,*}, \bar{\hat{\sigma}}_u^{2,*})^t$ be the respective vectors of original estimates and bootstrap means of the vector of population parameters $\boldsymbol{\psi} = (\beta, \sigma_\varepsilon^2, \sigma_u^2)^t$.

The experiment involves generating populations from the model in (11) with parameters $\beta = 1$, $\sigma_u^2 = 0.2$, $\sigma_\varepsilon^2 = 0.5$ and $M = 300$ second level units. The second level sizes N_j were determined by $N_j = 75\exp(\tilde{u}_j)$, where $\tilde{u}_j \sim N(0, \sigma_u^2)$ truncated below by $-1.5\sigma_u$ and above by $1.5\sigma_u$. The values of N_j lie in the interval [38;147] with average around 80. The sample size at the first stage is $m = 35$ level 2 units. At the second stage, simple random samples of level 1 units of sizes $n_{j.1} = 2$ and $n_{j.2} = 7$ are drawn from strata 1 and 2, respectively. Therefore, the total sample sizes n_j are fixed ($n_j = 9$) for all level 2 units j.

Generation of bootstrap samples from the original and the pseudo original samples (step 0 and step 3 in section 3, respectively) is carried out by both parametric and nonparametric bootstrap. The latter involves selecting the second level units by a simple random sampling design, with all the first level units (individual level) from the sampled level 2 units being included in the sample. The sampling weights used in the replicated values (bootstrap samples) are identical to the original sampling weights. For simplicity, generation of the plausible parameter values (step 1, section 3) was performed from a uniform distribution between two boundaries determined by the original estimates (see step II.iii) below). The experiment consists in replicating the six steps of the EBS approach (section 3) a large number of times in order to assess the bias and mean squared error of the EBS estimators. The method is described below for the PWIGLS estimator $\hat{\boldsymbol{\psi}}^{pw}$, with an identical process applied to the naïve estimator $\hat{\boldsymbol{\psi}}^{naive}$. When the selection (or generation) of the sample does not depend on the type of the estimator (i.e. in steps II) and ii.a) below), both estimators $\hat{\boldsymbol{\psi}}^{pw}$ and $\hat{\boldsymbol{\psi}}^{naive}$ were obtained from the same sample. The EBS approach was implemented as follows:

I) Generate a population according to the population model in (11);

II) From the population in I), select one original sample by an informative sampling scheme described in the previous sub-section. For this sample, perform the following:

- i) Obtain PWIGLS original estimates, $\hat{\boldsymbol{\psi}}^{pw}$, of $\boldsymbol{\psi}$ with corresponding PWIGLS standard errors $se(\hat{\boldsymbol{\psi}}^{pw})$;
- ii) Obtain bootstrap means $\overline{\hat{\boldsymbol{\psi}}}^{pw,*}$ from $B = 250$ bootstrap samples generated by
  - ii.a) nonparametric bootstrap;
  - ii.b) parametric bootstrap via the sample model based on $\hat{\boldsymbol{\psi}}^{pw}$, i.e., $f_{s_j}(y_{ij} \mid \beta_j, \hat{\boldsymbol{\psi}}^{pw})$ and $f_s(\beta_j \mid \hat{\boldsymbol{\psi}}^{pw})$;
- iii) Obtain a set of $G = 400$ parameter values for each component of $\boldsymbol{\psi}$. For an original estimate $\hat{\beta}^{pw}$ of β, for example, the values are generated by randomly drawing G values from $uniform(a,b)$, where $a = \hat{\beta}^{pw} - 3se(\hat{\beta}^{pw})$ and $b = \hat{\beta}^{pw} + 3se(\hat{\beta}^{pw})$, yielding $\boldsymbol{\psi}_1^{pw},...,\boldsymbol{\psi}_G^{pw}$.
- iv) For each $\boldsymbol{\psi}_g^{pw}$, $g = 1,...,400$, in iii):
  - iv.a) Generate one pseudo original sample from $f_{s_j}(y_{ij} \mid \beta_j, \boldsymbol{\psi}_g^{pw})$ and $f_s(\beta_j \mid \boldsymbol{\psi}_g^{pw})$, obtaining $\hat{\boldsymbol{\psi}}_g^{pw}$;
  - iv.b) Obtain bootstrap means $\overline{\hat{\boldsymbol{\psi}}}_g^{pw,*}$ from $B = 250$ bootstrap samples generated by:
    - iv.b.1) nonparametric bootstrap;
    - iv.b.2) parametric bootstrap via the sample model based on $\hat{\boldsymbol{\psi}}_g^{pw}$, i.e., $f_{s_j}(y_{ij} \mid \beta_j, \hat{\boldsymbol{\psi}}_g^{pw})$ and $f_s(\beta_j \mid \hat{\boldsymbol{\psi}}_g^{pw})$;
- v) Allocate $v = 50$ parameter values to the validation group and $G - v = 350$ values to the modelling group (step 4 in section 3). For correcting the bias of, say $\hat{\beta}^{pw}$, identify functions $h^*(\cdot)$ by modelling $\hat{\beta}_g^{pw} - \beta_g^{pw} = h(\hat{\boldsymbol{\psi}}_g^{pw}, \overline{\hat{\boldsymbol{\psi}}}_g^{pw,*})$, $g = 1,...,350$. After validating them in the validation group, the 'best' function, say $h_1^*(\cdot)$, for correcting the bias of $\hat{\beta}^{pw}$ is chosen according to expression (7) in section 3. Similarly, functions $h_2^*(\cdot)$ and $h_3^*(\cdot)$ are chosen for correcting the bias of $\hat{\sigma}_u^{2,pw}$ and $\hat{\sigma}_\varepsilon^{2,pw}$, respectively;

vi) Obtain EBS bias-corrected estimator of $\hat{\beta}^{pw}$, $\hat{\sigma}_u^{2,pw}$ and $\hat{\sigma}_\varepsilon^{2,pw}$ such that $\hat{\beta}^{EBS} = \hat{\beta}^{pw} - h_1^*(\hat{\psi}^{pw}, \overline{\hat{\psi}}^{pw,*})$; $\hat{\sigma}_u^{2,EBS} = \hat{\sigma}_u^{2,pw} - h_2^*(\hat{\psi}^{pw}, \overline{\hat{\psi}}^{pw,*})$ and $\hat{\sigma}_\varepsilon^{2,EBS} = \hat{\sigma}_\varepsilon^{2,pw} - h_3^*(\hat{\psi}^{pw}, \overline{\hat{\psi}}^{pw,*})$. Note that the original estimates from steps i) and ii) are plugged in the functions.

Mean squared error estimates of the EBS estimators are obtained by using the estimates from step v). For $\hat{\beta}^{EBS}$, for example, let $\hat{\beta}_g^{EBS} = \hat{\beta}_g^{pw} - h_1^*(\hat{\psi}_g^{pw}, \overline{\hat{\psi}}_g^{pw,*})$, $g = 1,...,G = 400$, and obtain $M\hat{S}E(\hat{\beta}^{EBS}) = G^{-1}\sum_{g=1}^{G}(\hat{\beta}_g^{EBS} - \beta_g^{pw})^2$.

Steps I-II above were repeated $R = 100$ times. The number of plausible parameter values $G = 400$ was chosen in order to allow a reasonable number of parameter values for estimation (350 values) and validation (50 values) of the candidate bias correction functions (step 4 in section 3). In practical situations, however, only one original sample is available for analysis and usually only one type of estimator and type of bootstrap is considered in the bias corrections, which make the EBS approach much easier to be implemented with a reasonable execution time (a few hours in this application). The experiments conducted in this research were all implemented in the statistical software R (R Development Core Team, 2008) and in the statistical software SAS (SAS Institute Inc., 1999), including a specific routine for the PWIGLS approach not available in standard statistical software. The naïve parameter estimates were obtained by using the function lme of the statistical software R. In practice, the modelling of the bias correction data in step v) of the experiment needs to be done just once. However, in order to undertake a Monte Carlo study of the performance of the EBS bias-corrected estimators with $R = 100$ original samples, an automatic procedure for searching for the 'best' function for bias correction of each estimator is needed. This procedure was implemented as follows. Having generated the $G = 400$ values for a specific parameter (step iii) of the experiment), $v = 50$ values were randomly selected as the validation group (step v)) and $G - v = 350$ values as the modelling group. Concentrate on the modelling group for now. For a specific response variable (error of the estimator to be corrected), a full model is fitted to the 'raw data' (model 0) and extreme values were excluded from the analysis based on the Cook's distance (Fox, 1991), which is a well-known outlier indicator available in most of the standard statistical software. Using the 'clean data', the full model is re-fitted (model 1). Another model is obtained (model 2) by applying a backward elimination regression method on the clean data. Model 2 without the intercept (model 3) was also assessed in situations where although that estimate was not significantly different from zero the intercept was kept in the model by the backward elimination procedure. Hence, for each parameter and bias correction function identified, the three models (models 1, 2 and 3 above) were validated on the validation data, obtaining realisations of bias of the EBS estimator. Six different functions were considered in the modelling process and models 1, 2 and 3 were fitted for each function. The chosen model, among the 18 models assessed, was the one producing the smallest estimated bias for the EBS estimator based on the validation group (a discussion on this issue is presented in section 5). Six functions were used in the process of choosing the best function for bias correction for each estimator: σ_u^2, β and σ_ε^2. These functions were identified by firstly fitting the model manually for a few original samples. Then, they were included in the list of candidate functions to be evaluated by the automatic procedure described in the previous paragraph. The classical additive and multiplicative bootstrap bias corrections are special case, for example, of functions 1 and 2, respectively, shown in Table 1, taking $a_1 = 1$ and $a_0 = a_2 = a_3 = 0$.

Table 1: Examples of candidate bias correction functions used in the EBS approach for $\hat{\sigma}_u^2$.

Function Code	Formulae of function $h(\cdot)$, $g = 1,\dots,G - v = 350$
1	$\frac{\hat{\sigma}_{u,g}^2}{\sigma_{u,g}^2} = a_0 + a_1 \cdot (\frac{\bar{\hat{\sigma}}_{u,g}^{2,*}}{\hat{\sigma}_{u,g}^2}) + a_2 \cdot (\frac{\bar{\hat{\sigma}}_{\varepsilon,g}^{2,*}}{\hat{\sigma}_{\varepsilon,g}^2}) + a_3 \cdot (\frac{\bar{\hat{\beta}}_g^*}{\hat{\beta}_g})$
2	$\hat{\sigma}_{u,g}^2 - \sigma_{u,g}^2 = a_0 + a_1 \cdot (\bar{\hat{\sigma}}_{u,g}^{2,*} - \hat{\sigma}_{u,g}^2) + a_2 \cdot (\bar{\hat{\sigma}}_{\varepsilon,g}^{2,*} - \hat{\sigma}_{\varepsilon,g}^2) + a_3 \cdot (\bar{\hat{\beta}}_g^* - \hat{\beta}_g)$

4.3 Results

Tables 2-4 show summary statistics for the naïve, PWIGLS and respective bias-corrected estimators of the three parameters β, σ_u^2 and σ_ε^2 under the classical corrections and the EBS approach. Results are reported for nonparametric bootstrap only. Similar conclusions are valid for the parametric case. The number of original samples considered for naïve and PWIGLS estimators is $R = 72$ and $R = 95$, respectively.

Let $\hat{\psi}_r$ denote an estimator (naïve, PWIGLS, EBS or classical bias-corrected estimator) of the parameter ψ (known in the simulation study) for the original sample $r = 1,\dots,R$. The following summary statistics were computed for $\hat{\psi}_r$: simulation mean (Mean): $\bar{\hat{\psi}} = R^{-1}\sum_{r=1}^{R}\hat{\psi}_r$; simulation standard deviation (SD): $\sqrt{(R-1)^{-1}\sum_{r=1}^{R}(\hat{\psi}_r - \bar{\hat{\psi}})^2}$; empirical bias (Bias): $R^{-1}\sum_{r=1}^{R}(\hat{\psi}_r - \psi)$; empirical relative bias (RB): $R^{-1}\sum_{r=1}^{R}(\frac{\hat{\psi}_r - \psi}{\psi})$; empirical root mean squared error (RMSE$^{\text{emp}}$): $\sqrt{R^{-1}\sum_{r=1}^{R}(\hat{\psi}_r - \psi)^2}$.

As anticipated, the naïve and PWIGLS estimators based on the original sample are highly biased in the present scenario, especially for the variance component parameters. The EBS bias-corrected estimators show very good performance for all model parameters, including the variance component estimators, which are expected to be most problematic to estimate due to their high sensitivity to small sample sizes. Classical bias corrections, however, perform poorly with small or no reduction in the biases.

As anticipated, the naïve and PWIGLS estimators based on the original sample are highly biased in the present scenario, especially for the variance component parameters. The EBS bias-corrected estimators show very good performance for all model parameters, including the variance component estimators, which are expected to be most problematic to estimate due to their high sensitivity to small sample sizes. Classical bias corrections, however, perform poorly with small or no reduction in the biases.

It is worth emphasising that the bias-corrected PWIGLS estimators obtained by the EBS approach perform well even in the case where the non-corrected estimator is practically unbiased, which is a desirable characteristic of a bias correction procedure. This is the case of the PWIGLS estimator of the intercept β (Pfeffermann *et al.*, 1998b). In addition, the trade-off bias-variance does not seem to be an issue for the EBS bias-corrected estimators, since the mean squared errors ($RMSE^{emp}$) show minimum increase or even reduction when compared to the non-corrected estimators.

5. Conclusions and Final Remarks

In this article a bias correction approach – the extended bootstrap bias correction (EBS) – was proposed. The method was applied to bias adjustment of unweighted and PWIGLS estimators of linear two level model parameters under informative sampling of level 1 units with small sample sizes at both levels. The EBS procedure was assessed by Monte Carlo study, evaluating the behaviour of the EBS estimators through mean squared errors and biases estimates.

The main finding of the Monte Carlo study conducted in this study is that the EBS bias correction approach performs very well for all the scenarios considered (naïve and PWIGLS estimators under nonparametric bootstrap – tables 2-4).

When the sampling design is informative and the parametric bootstrap is applied, the estimation of the sample model is necessary in the EBS approach (for generation of pseudo original samples and bootstrap samples) and also in the classical corrections (for generation of bootstrap samples). If the nonparametric bootstrap is adopted, the sample model is still needed for generation of the pseudo original samples in the EBS approach, although it is not necessary when using the classical corrections (but the results revealed poor performance in this situation – tables 2-4).

In the case of a noninformative sampling scheme, knowledge of the population model is required for both parametric and nonparametric bootstrap if the EBS approach is used. For classical corrections, however, assumptions about the population model can be relaxed if the nonparametric bootstrap is adopted.

The EBS approach is more computing intensive when compared to the classical corrections, but shows good performance for all the scenarios studied. In addition, unlike the classical corrections, it allows mean squared error and confidence interval estimation for the bias-corrected estimators in practical situations where only one original sample is available for analysis.

A range of factors can be changed to improve the EBS approach proposed in this article. First, the best function can be chosen such that the *estimated MSE* (not the bias) of the EBS estimator is minimum (Corrêa, 2008, Chapters 4 and 5).

Another issue to be explored is the generation of the parameter values. Adopting wider intervals and distributions other than the uniform is an important issue to be considered. For example, considering a wider interval $[a,b]$, $a=\hat{\psi}-4\cdot se(\hat{\psi})$ and $b=\hat{\psi}+4\cdot se(\hat{\psi})$, for the generation of the plausible parameter values (step iii) in section 4.2) may improve the fitting of the model and, consequently, the bias correction results, since the true parameter value will almost surely be covered by an wider interval. On the other hand, it may be true that the further the plausible parameter value is from the true value the more different the relationship is, suggesting considering a narrower interval. Besides, it is important to perform a sensitivity analysis to assess the effect of the choice of the distribution used to generate the plausible parameter values on the performance of the EBS approach.

The number of plausible parameter values G, and also the number of bootstrap samples, can be increased in practice since only one original sample is actually available for analysis. In addition, more sophisticated regression analysis tools can be used on the identification of the bias correction function, yielding better EBS estimators.

The extended bootstrap bias correction procedure proposed and assessed in this article adopted Monte Carlo methods in an attempt to mimic common situations found in practice, where biased estimates are produced as a result of analysing survey data. We conclude therefore that the EBS approach is applicable to real survey data.

Acknowledgements

This work was conducted under the joint supervision of Professor Danny Pfeffermann and Professor Chris Skinner. This work was partially supported by the Brazilian Research Council CAPES grant BEX2333/03-1 and the Brazilian Institute of Geography and Statistics (IBGE).

References

Chambers, R.L. (2003). Introduction to Part A. In *Analysis of Survey Data* (eds. R.L. Chambers and C. J. Skinner). 13-27, Chichester: Wiley.

Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Satistical Review*, **51**, 279-292.

Corrêa, S. T. (2008). *Bias Corrections in Multilevel Modelling of Survey Data with Applications to Small Area Estimation*. Ph.D. Thesis. University of Southampton, England.

Davidson, A. C. and Hinckley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.

DiCiccio, T.J. and Efron B. (1996). Bootstrap confidence intervals (with discussion). *Statistical Science*, **11**, 189-228.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The annals of Statistics*, **7**, 1-26.

Efron, B. and Tibshirani, R.J. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, **1**, 54-75.

Fox, J. (1991). *Regression Diagnostics*. Thousand Oaks, CA: Sage Publications. Quantitative Applications in the Social Sciences Series No. 79.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalised least squares. *Biometrika*, **73**, 43-56.

Goldstein, H. (2003). *Multilevel statistical models.* 3rd edn. London: Hodder Arnold.

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.

Hall, P. and Maiti, T. (2006). On bootstrap methods for small area prediction. *Journal of Royal Statistical Society B*, **68**, 221-238.

Korn, E.L. and Graubard, B.I. (1995). Examples of differing weighted and unweighted estimates from a sample survey. *American Statistician*, **49**, 291-295.

_____, (2003). Estimating variance components by using survey data. *Journal of the Royal Statistical Society B*, **65**, 175-190.

Pfeffermann, D., Skinner, C.J., Holmes, D.J., Goldstein, H. and Rabash, J. (1998b). Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society B*, **60**, 23-40.

Pfeffermann, D., Krieger, A. and Rinott, Y. (1998a). Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica*, **8**, 1087-1114.

Pfeffermann, D., Moura, F. and Silva, P. (2006). Multi-level modelling under informative probability sampling. *Biometrika*, **93**, 943-959.

R Development Core Team. (2008). R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL http://www.R-project.org.

Rabe-Hesketh, S. and Skrondal, A. (2006). Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society A*, **169**, part 4, 805-827.
SAS Institute Inc. (1999). *SAS®/STAT user's guide version 8*. NC: SAS Institute Inc.
Shao, J. and Tu, D. (1995). *The Jackknife and the Bootstrap*. New York: Springer-Verlag.
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), 59-87. Chichester: Wiley.

Table 2: Naïve and PWIGLS estimators. Nonparametric bootstrap. True value $\beta = 1$.

Estimator	Mean	SD	Bias	RB	RMSE^{emp}
$\hat{\beta}^{naive}$	0.691	0.086	-0.309	-30.9%	0.321
Additive correction	0.691	0.086	-0.309	-30.9%	0.321
Multiplicative correction	0.691	0.086	-0.309	-30.9%	0.321
EBS	1.002	0.151	0.002	0.2%	0.150
$\hat{\beta}^{pw}$	0.994	0.092	-0.006	-0.6%	0.091
Additive correction	0.995	0.092	-0.005	-0.5%	0.092
Multiplicative correction	0.995	0.092	-0.005	-0.5%	0.092
EBS	0.996	0.089	-0.004	-0.4%	0.088

Table 3: Naïve and PWIGLS estimators. Nonparametric bootstrap. True value $\sigma_u^2 = 0.2$.

Estimator	Mean	SD	Bias	RB	RMSE^{emp}
$\hat{\sigma}_{\varepsilon}^{2,naive}$	0.169	0.046	-0.031	15%	0.055
Additive correction	0.165	0.049	-0.035	17%	0.060
Multiplicative correction	0.165	0.049	-0.035	17%	0.060
EBS	0.187	0.071	-0.013	6%	0.071
$\hat{\sigma}_{u}^{2,pw}$	0.158	0.052	-0.042	-20%	0.065
Additive correction	0.165	0.054	-0.035	-15%	0.064
Multiplicative correction	0.165	0.054	-0.035	-15%	0.064
EBS	0.196	0.051	-0.004	-2%	0.050

Table 4: Naïve and PWIGLS estimators. Nonparametric bootstrap. True value $\sigma_{\varepsilon}^2 = 0.5$.

Estimator	Mean	SD	Bias	RB	RMSE^{emp}
$\hat{\sigma}_{\varepsilon}^{2,naive}$	0.429	0.041	-0.071	-14.2%	0.082
Additive correction	0.448	0.043	-0.052	-10.4%	0.067
Multiplicative correction	0.449	0.043	-0.051	-10.2%	0.066
EBS	0.493	0.050	-0.007	-1.4%	0.050
$\hat{\sigma}_{\varepsilon}^{2,pw}$	0.522	0.056	0.022	4.4%	0.060
Additive correction	0.522	0.056	0.022	4.4%	0.060
Multiplicative correction	0.522	0.057	0.022	4.4%	0.060
EBS	0.503	0.053	0.003	0.2%	0.054