Parametric Bootstrap Confidence Intervals for Survey-Weighted Small Area Proportions

Benmei Liu[1] and Mamadou Diallo[2]
1. Division of Cancer Control and Population Sciences, National Cancer Institute, 9609 Medical Center Drive, Rockville, Maryland 20850
2. Westat, 1600 Research Blvd, Rockville, MD 20850

Abstract

In this paper, we apply the recently developed parametric bootstrap method in constructing confidence intervals for the well-known Fay-Herriot model in estimating survey-weighted small area proportions. Through design-based simulation studies from a real finite population and extensive purely model-based simulation studies, we examine the coverage properties of the parametric bootstrap confidence intervals for the Fay-Herriot model. We also compare them with those obtained from other competing methods for the same model.

Key Words: Survey-weighted proportions, Fay-Herriot model, parametric bootstrap, coverage properties

1. Introduction

Small area estimation (SAE) methods are often used to estimate the proportions of units with a given characteristic for small areas. For example, Census Bureau has been using SAE methods to estimate poverty rates for states, counties, and school districts in its Small Area Income and Poverty Estimates (SAIPE) program since 1990s (Citro and Kalton, 2000; Maples and Bell, 2005); Substance Abuse and Mental Health Services Administration (SAMHSA) has been applying SAE techniques to estimate substance rates for states with data from the National Survey on Drug Use and Health (NSDUH) (Wright et al., 2007); the National Center for Education Statistics used SAE techniques to estimate proportions at the lowest level of literacy for states and counties with data from the National Assessment of Adult Literacy (NAAL) (Mohadjer et al., 2012). In each case, the survey's sample sizes in the small areas are not large enough to support direct estimates (survey weighted estimates) of adequate precision. A wide variety of methods have been developed to address such small area estimation problems. See Rao (2003) and Jiang and Lahiri (2006) for reviews, and Chattopadhyay et al. (1999), Farrell et al. (1997), and Malec et al. (1997, 1999) for methods specifically for estimating small area proportions. The range of methods includes both empirical best prediction (EBP) and hierarchical Bayes (HB) approaches and models developed at both the area and unit levels.

In this paper, we apply the recently developed parametric bootstrap method by Chatterjee, Lahiri and Li (2008) in constructing confidence intervals for small area proportions using the well-known Fay-Herriot model and evaluate its coverage property.

In section 2, we describe some commonly used models. Section 3 lays out how to construct parametric bootstrap confidence intervals. The coverage property of the parametric bootstrap confidence intervals are evaluated by means of a design-based Monte Carlo simulation study in which stratified simple random samples are selected from a fixed finite population as well as a purely model-based simulation. The simulation study and the results are described in Section 4. Section 5 provides some concluding remarks.

2. Commonly used small area models for estimating small area proportions

Let N_i denote the population size in area i $(i = 1, \dots, m)$ of the target finite population. Let y_{ik} be the binary response for the characteristic of interest for unit k $(k = 1, \dots, N_i)$ in area i. The parameters to be estimated are the small area proportions $P_i = \Sigma_k y_{ik} / N_i$.

Let n_i denote the sample size in area i and w_{ik} denote the sampling weight for sampling unit k in area i. The standard direct survey estimator for P_i is:

$$p_{iw} = \frac{\sum_{k=1}^{n_i} w_{ik} y_{ik}}{\sum_{k=1}^{n_i} w_{ik}},\ i = 1, \dots, m. \tag{1}$$

The associated variance of p_{iw} can be expressed as

$$Var(p_{iw}) = \frac{P_i(1-P_i)}{n_i} DEFF_i, \tag{2}$$

where $DEFF_i$ is the design effect reflecting the impact of the complex sample design (Kish, 1965).

The problem is that p_{iw} is very imprecise when the sample size n_i is small or even cannot be computed if the sample size is zero. Small area estimation procedures can be used to address this problem.

There are various statistical models and methods of estimation for developing small area estimates (SAEs) using survey data. The most prominent fundamental approach is the Fay-Herriot model (Fay and Herriot, 1979) originally developed to estimate per-capita income for U.S. areas with populations of less than 1000. It has two components: the sampling model and the linking model. The sampling model is a model for the sampling error of the direct survey estimates conditionally on the parameters of interest. The linking model relates the population value for an area to area-specific auxiliary variables $x_i = (x_{i1}, \dots, x_{ip})'$. A simple form of the *Fay-Herriot model* to estimate P_i can be written as:

Sampling model: $$p_{iw} \mid P_i \overset{ind}{\sim} N(P_i,\ D_i); \tag{3}$$

Linking model: $P_i \mid \beta, A \overset{ind}{\sim} N(x_i'\beta, A)$; (4)

where D_i is the sampling variance of the direct estimate p_{iw} and is assumed known. In practice the sampling variances are estimated and due to the small sample sizes their estimates can be very unstable.

Carter and Rolph (1974) applied an arcsine transformation function [$\hat{\theta}_i = \arcsin\left(\sqrt{p_i}\right)$] in their false alarm probability estimation example. For the model-based study presented in this paper, both the Fay-Herriot model without any transformation and the Fay-Herriot model after the arcsin transformation following Carter and Rolph (1974) were examined.

There have been many developments of SAE models beyond the basic Fay-Herriot approach. For a full range review of different small area models, we refer to Rao (2003) and Jiang and Lahiri (2006).

3. Parametric bootstrap confidence intervals based on the Fay-Herriot model

Given the Fay-Herriot model (3)-(4), the empirical Bayes (EB) estimator for P_i is:

$\hat{P}_i^{EB} = (1 - \hat{B}_i)p_{iw} + \hat{B}_i x_i'\hat{\beta}$, where $\hat{B}_i = D_i/(\hat{A} + D_i)$, $i = 1, \dots, m$.

The estimate $\hat{\beta}$ and $\hat{A}$ are often obtained using maximum likelihood estimation (MLE) method and restricted maximum likelihood (REML) method respectively. One disadvantage of REML method for estimating A is that it could produce zero and even negative estimates. To produce a strictly positive estimate of A, Li and Lahiri (2010) introduced an adjustment to the maximum (profile or residual) likelihood estimator of A for small m. We apply their method to estimate A when producing parametric bootstrap confidence intervals.

Following Chatterjee, Lahiri and Li (2008), we consider the five steps below to construct the parametric bootstrap prediction interval for the i th small area:

Step 1: Generate B bootstrap samples (say B=1000) using the following distributions:

$P_i^* \sim N(x_i'\hat{\beta}, \hat{A})$ and $p_{iw}^* | P_i^* \sim N(P_i^*, D_i)$, $i = 1, \dots, m$, where $\hat{\beta}$ and $\hat{A}$ are estimates based on the original sample.

Step 2: For each of the B bootstrap samples, obtain $\hat{A}^*$, $\hat{B}_i^*$, and $\hat{\beta}^*$, where $\hat{B}_i^* = D_i/(\hat{A}^* + D_i)$.

Step 3: For each bootstrap sample, compute: $t_i = (P_i^* - \hat{P}_i^{EB*})/\sqrt{D_i(1 - \hat{B}_i^*)}$, where $\hat{P}_i^{EB*} = (1 - \hat{B}_i^*)p_{iw}^* + \hat{B}_i^* x_i'\hat{\beta}^*$, the EB estimator based on the bootstrap sample.

Step 4: Locate the two equal-tail $\alpha/2$ cut-off points (t_1, t_2) using the B (=1000) pivot values computed from step 3.

Step 5: Construct the parametric bootstrap prediction interval for the small area i:

$$PI_i = \left\{ \hat{P}_i^{EB} + t_1 \sqrt{D_i\left(1-\hat{B}_i\right)},\ \ \hat{P}_i^{EB} + t_2 \sqrt{D_i\left(1-\hat{B}_i\right)} \right\}, \quad (5)$$

where $\hat{P}_i^{EB}$ and $\hat{B}_i$ are estimated using the original sample, while t_1 and t_2 are obtained from the bootstrap procedure described in steps 1-4 above.

Chatterjee, Lahiri and Li (2008) showed that the prediction interval defined by (5) has a coverage probability of $1-\alpha$ with marginal error of $O\left(m^{-3/2}\right)$ given the model is correct.

4. Simulation Study

In this research, we conducted extensive simulation studies to evaluate the performance of the bootstrap confidence intervals for the small area proportions through design-based and model-based simulations.

4.1 Design-based simulation

4.1.1 The study population

The sampling frame for the design-based study was the 2002 Natality public-use data file that covered all births occurring within the United States in that calendar year. The file contained data obtained from the certificates filed for births occurring in each state and territory (for details see U.S. National Center for Health Statistics, 2009).

The finite target population studied was restricted to the 4,024,378 records of live births that occurred in 2002 in the 50 states of U.S. and the District of Columbia (DC) and that had birth weights reported. The parameter of interest was the state level percentage of weights at birth below the national median birthweight P_i, $i = 1,...,51$. The national median birthweight for the 2002 target population was equal to 3,345 grams. The true parameters of interest P_i obtained from the sampling frame varied from 40.2% to 58.5%.

We used the same sampling design as the one used in the simulation study described in Liu et al (2007) but with some modifications. Within each state, a stratified SRS design was used to draw samples from the birth records. Mother's race (White, Black, and Other) was used as the stratification variable. In Liu et al (2007), the national sample size was set to be about 1,500 birth records for each race group. A uniform sampling fraction was used across the states for each race group, subjecting to the condition at least two birth records were sampled within each race group in each state. The resultant national sample size turned out to be $n = 4{,}526$ birth records. The state sample sizes n_i ranged from 7 (for small states such as Vermont) to 690 (for California), with a median sample size of 61. For this study, the states with an original sample size less than 50, we increased the sample size to 50 so the total sample size was 5,148. This sampling procedure was repeated $R = 1{,}000$ times, creating 1,000 independent samples. The sampling weights remain the same over different simulation runs.

The reason we chose percentage of below median birthweight for this research was to get a set of true P_i that are normally distributed so the assumption of the linking model (4) of the Fay-Herriot model could hold. We tested this assumption using this set of true P_i . The following five auxiliary variables were selected to fit the model after stepwise model selection process: 1) percent of births with father being White; 2) percent of births with mother being Non-Hispanic; 3) percent of births with being first live child in family; 4) percent of births with mother being native born; and 5) percent of births with no prenatal care. All the five selected auxiliary variables were significant in predicting P_i at the significant level of $\alpha = 0.05$. We ran the regression version of model (4), i.e., $P_i = x_i^{'}\beta + v_i$, where $v_i \sim N(0, A)$, with and without the five selected covariates. Figure 1 below shows the histogram of the residuals v_i, the left histogram is for the case when no covariates were included, the right histogram is for the case when the 5 selected auxiliary variables were included. We further assessed the normality of the residuals using both Shapiro-Wilk normality test and Kolmogorov-Smirnov normality test . The p-values from the Shapiro-Wilk normality test were 0.2283 and 0.1382 for the case without covariates and the case with 5 covariates respectively. The corresponding p-values were 0.7348 and 0.2608 from the Kolmogorov-Smirnov normality test. Both normality tests concluded that the residuals were normally distributed for the two cases. The normality tests suggest that both the model without covariates and the one with the 5 auxiliary variables are good candidate for assumption (4).

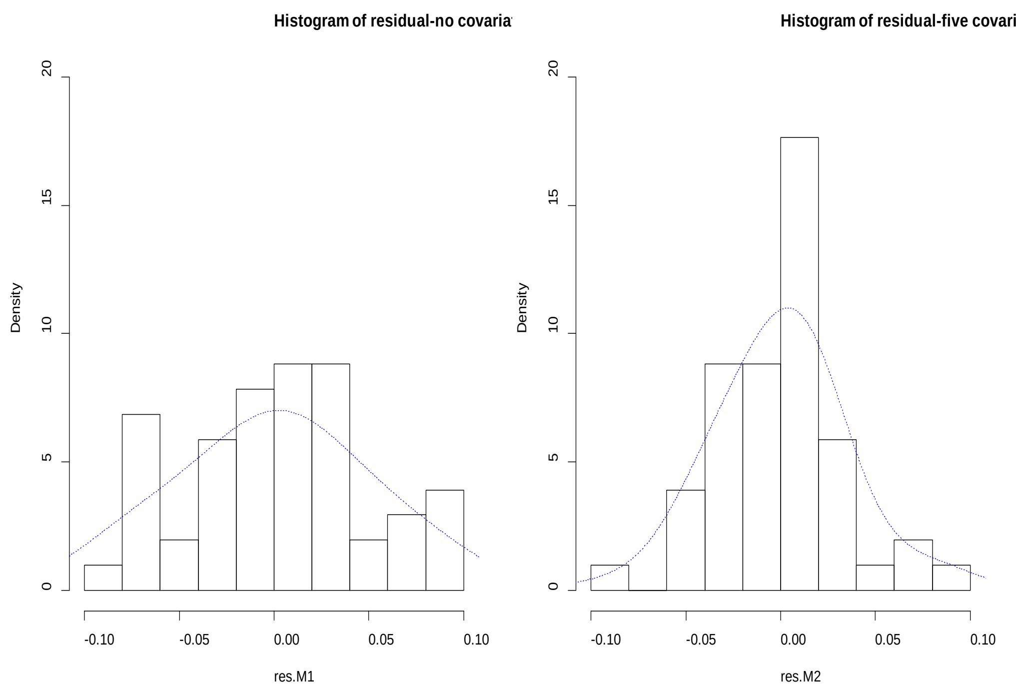

Figure 1: Histograms of the residuals from model (4) without and with covariates

4.1.2 Computation of the model-based Estimates and associated confidence (credible) intervals

For each sample, the first step in the computations was to calculate the state direct sample estimates. Then from both EBP method and HB method obtain the final model-based estimates for each small area. We considered both with and without covariates models in the estimation. We also used several different approaches to compute the 95% confidence interval for the associated point estimates. To avoid extra variability we used the true sampling variances (2) in all the computations.

The EBP approach:

The EBP estimate of P_i based on each sampled data set was computed using formula (5) described in Section 3. To compute the 95% confidence intervals, we applied the following two approaches:
Approach 1: The parametric bootstrap method described in section 3;
Approach 2: The REML-Delta method. This approach consists of using the REML method to estimate A, then apply the Datta and Lahiri (2000) method to estimate the mean square error, and then compute the 95% confidence interval under normality assumption.

The HB approach:

For the HB approach without covariates the following assumptions were made:

1. No auxiliary variables were used, so that $x_i'\beta = \mu$;
2. Flat prior for μ , i.e., $f(\mu) \propto 1$, and uniform prior for A, i.e., $A \sim Unif(0, L)$, where L is a fixed large numerical number. We used $L = 100$ in this study.

For the HB approach with covariates, we included the five selected state level auxiliary variables described in Section 4.1.1. The prior assumptions were the same as the approach without any covariates, except that we assumed a flat prior for β instead of μ.

The direct estimates for each sample were used in turn as input to the WinBUGS software (Lunn et al., 2000), which was used to produce the HB estimates for the Fay-Herriot model. For each WinBUGS run, three independent chains were generated. For each chain, burn-ins of 10,000 samples were produced, with 10,000 samples after burn-in. The samples after burn-in were thinned by a factor of two to reduce auto-correlation of the MCMC samples. The resultant 15,000 MCMC samples from the three chains after burn-in were then used to compute the posterior mean and percentiles for each HB model based on each sample data set. The potential scale reduction factor $\hat{R}$ was used as the primary measure for convergence (see Gelman and Rubin, 1992).

Let P_i^{HB} denote an HB estimator of P_i, the percentage of live births with birthweight below the grand median birthweight in state *i*, and let $P_{i,q}^{HB}$ denote the q^{th} percentile of the posterior distribution of P_i. The 95% credible interval (or confidence interval) for P_i^{HB} was defined as ($P_{i,2.5}^{HB}, P_{i,97.5}^{HB}$).

The direct approach

For comparison purpose, we also computed the 95% confidence interval using the direct estimates defined by (1) and the estimated variances under normality assumption. We computed the associated variances using formula (2) but replacing P_i by p_{iw} and $DEFF_i$ by $deff_{iw}$, the approximate design effect as defined in Liu et al (2007), which does not depend on any unknown parameters.

4.1.3 Simulation results

Based on the results from the 1,000 samples, Tables 1a and 1b present the following for each of the 4 different approaches: the noncoverage probability for the 95 percent credible intervals of the estimate of P_i, i.e., the probability that the interval fails to cover P_i along with the Monte Carlo simulation standard errors; and the mean width of the credible intervals. Table 1a presents the results when no covariates were used in the model, while Table 1b reports the results when the 5 covariates were incorporated in the model.

To examine the effect of state sample size on the simulation results, the 50 states plus DC are placed in 3 groups according to their sample sizes: the 24 states with relatively small sample sizes $(50 \le n_i < 60)$; the 15 states with medium sample sizes $(60 \le n_i < 100)$; and the 12 states with large sample sizes $(n_i \ge 100)$; The results presented in Table 1a and 1b are overall averages across all states and averages for the three groups separately.

Table 1a: Percentage of times that the 95 percent credible intervals fail to cover P_i, mean 95 percent credible interval width, along with the Monte Carlo simulation standard errors based on 1,000 simulations (in percentages) based on 4 different approaches **without any covariates**

	EBP Parametric Bootstrap		EBP REML-Delta		HB		Direct Method	
	Non coverage percentage and Monte Carlo simulation standard error							
State sample size n	estimate	SE	estimate	SE	estimate	SE	estimate	SE
overall	10.4	0.12	19.4	0.15	7.8	0.11	6.0	0.10
50<=n<60 (24 states)	12.9	0.19	25.1	0.24	9.7	0.17	6.3	0.16
60<=n<100 (15 states)	13.4	0.25	25.0	0.31	11.1	0.24	5.9	0.19
100<=n<=690 (12 states)	1.6	0.12	0.8	0.08	1.3	0.10	5.4	0.21
	Mean width of the 95% confidence interval and Monte Carlo simulation standard error							
	estimate	SE	estimate	SE	estimate	SE	estimate	SE
overall	15.2	0.02	13.2	0.02	15.8	0.01	30.3	1.01E-05
50<=n<60 (24 states)	16.1	0.02	13.5	0.03	16.9	0.02	35.4	1.68E-05
60<=n<100 (15 states)	15.6	0.03	13.4	0.04	16.3	0.02	30.8	7.34E-06
100<=n<=690 (12 states)	13.0	0.02	12.6	0.03	13.1	0.02	19.6	3.27E-07

Among the model-based methods, the HB approach gave an overall of 7.8% noncoverage rate using the model without any covariates and 3.8% for the model with the 5 covariates, which were the best among the three model-based methods. The EBP parametric bootstrap approach gave an overall of 10.4% noncoverage rate using the model without any covariates and 7.6% for the model with the 5 covariates which are further away from the 5% nominal value compared to the HB approach. The EBP REML-Delta approach produced the worst noncoverage rates. All three methods ended up with an over coverage for the large group (the noncoverage rates are in the range of 0.8% to 1.9%). Comparison of Table 1a and Table 1b indicates that the use of covariates in the modeling process can help improve the coverage property, though the noncoverage rates were still away from the 5% nominal value. As expected, the direct method has average noncoverage rate close to the 5% nominal value especially for the large states. However, the average widths of the confidence intervals were much larger than those obtained based on model-based approaches.

Table 1b: Percentage of times that the 95 percent credible intervals fail to cover P_i, mean 95 percent credible interval width, along with the Monte Carlo simulation standard errors based on 1,000 simulations (in percentages) based on 4 different approaches **with 5 covariates**

	EBP Parametric Bootstrap		EBP REML-Delta		HB		Direct Method	
	Non coverage percentage and Monte Carlo simulation standard error							
State sample size n	estimate	SE	estimate	SE	estimate	SE	estimate	SE
overall	7.6	0.10	9.6	0.11	3.8	0.08	6.0	0.10
50<=n<60 (24 states)	12.7	0.19	15.2	0.20	6.1	0.14	6.3	0.16
60<=n<100 (15 states)	4.3	0.16	7.3	0.20	1.9	0.11	5.9	0.19
100<=n<=690 (12 states)	1.6	0.11	1.1	0.10	1.9	0.12	5.4	0.21
	Mean width of the 95% confidence interval and Monte Carlo simulation standard error							
	estimate	SE	estimate	SE	estimate	SE	estimate	SE
overall	15.5	0.02	14.4	0.02	16.9	0.01	30.3	1.01E-05
50<=n<60 (24 states)	16.2	0.03	15.2	0.03	18.6	0.02	35.4	1.68E-05
60<=n<100 (15 states)	15.8	0.03	14.4	0.03	17.3	0.02	30.8	7.34E-06
100<=n<=690 (12 states)	13.8	0.02	12.8	0.02	13.3	0.01	19.6	3.27E-07

4.2 Model-based simulation

4.2.1 Data generation

Let P_i and D_i , $\mathrm{i} = 1, \ldots, 51,$ be the state level proportions with birth weight less than the grand median computed from the 2002 Natality data and associated true variances. We generated 1,000 sets of simulated data using the following steps:

Step 1: Obtain β and A by fitting P_i on the five auxiliary variables described in section 4.1.1 using P_i~N($x_i'\beta$, A);

Step 2: Generate θ_i based on θ_i~N($x_i'\beta$, A);

Step 3: Generate 1,000 sets of observed data using different approaches:

Approach A: Generate p_{iw} based on the Level 1 of the Fay-Herriot model: $p_{iw} \sim N(\theta_i, D_i)$;

Approach B: Generate y_i using binomial distribution: $y_i \sim Bin\left(int(\frac{n_i}{DEFF_i}), \theta_i\right)$ and then compute $p_{iw} = \frac{y_i}{int(n_i/DEFF_i)}$.

Approach C: Generate y_i using Poisson distribution: $y_i \sim Pois\left(\frac{\theta_i n_i}{DEFF_i}\right)$ and then compute $p_{iw} = \frac{y_i}{n_i/DEFF_i}$;

4.2.2 Computation of the model-based Estimates and associated confidence (credible) intervals

For each generated data set, we applied the EBP parametric bootstrap approach, the EBP REML-Delta approach and the HB approach with the 5 covariates as described in Section 4.1.2 to obtain the model-based estimates for P_i and the associated 95% confidence (credible) intervals based on the Fay-Herriot model defined by(3)-(4). Using the same simulated data, we also tried the Fay-Herriot model on the data with arcsin transformation. Let $z_i = \arcsin\left(\sqrt{p_{iw}}\right)$ (Carter & Rolph, 1974). The following model was applied on z_i:

Sampling model: $z_i|\theta_i \sim N\left(\theta_i, \frac{DEFF_i}{4n_i}\right)$;
Linking model: $\theta_i = x_i'\beta + v_i$; where $v_i \sim N(0, A)$.

The goal was to estimate $P_i = sin^2(\theta_i)$ and the associated 95% confidence (credible) intervals using the same three approaches (EBP parametric bootstrap, EBP REML-Delta, and HB). For convenience purpose, we labeled the above model as Fay-Herriot model with arcsin transformation. The prior assumptions for the HB approach were the same as those in section 4.1.

4.2.3 Simulation results

Table 2a reports the percentage of times that the 95% confidence interval covers the true value and the mean width of the confidence intervals along with standard Monte Carlo simulation errors for the three approaches (EBP parametric bootstrap, EBP REML-Delta, and HB) when the simulated data was generated using the Fay-Herriot model. The left half of the table reports the estimation results based on the Fay-Herriot model, while the right half of the table reports the estimation results based on the Fay-Herriot model with arcsin transformation. The EBP parametric bootstrap approach based on the Fay-Herriot model gave 4.9% overall average noncoverage rate, which is very close to the nominal 5%. This is pretty consistent with what Chatterjee, Lahiri and Li (2008) showed analytically. The HB approach based on the Fay-Herriot model produced very good average noncoverage rates (4.7%) for the medium and large groups and a little conservative noncoverage rate (3.0%) for the small group. The REML-Delta approach performed worse than the other two approaches in terms of noncoverage rate. When Fay-Herriot model with arcsin transformation was used in the estimation, the average noncoverage rates were quite similar to those based on the Fay-Herriot model for all the three approaches. This is probably due to the fact that the small area proportions were around 50%.

Table 2b reports the percentage of times that the 95% confidence interval covers the true value and the mean width of the confidence intervals along with standard Monte Carlo simulation errors for the three approaches when the simulated data was generated using binomial model. The performances of the three approaches were worse than the case when the data were generating using Fay-Herriot model but better than the case when the data were generated using Poisson regression in terms coverage property. Again, parametric bootstrap approach performs the best among the three methods being compared and the performance of the three approaches did not differ much across the two estimation models (Fay-Herriot model and Fay-Herriot model with arcsin transformation).

Table 2a: Percentage of times that the 95 percent credible intervals fail to cover P_i, mean 95 percent credible interval width, along with the Monte Carlo simulation standard errors based on 1,000 simulations (in percentages) for the 3 different estimation methods - Data generated using level 1 of the Fay-Herriot model (Approach A):

	FH -model			FH-model after arcsin transformation		
	EBP Parametric bootstrap	EBP REML-Delta	HB	EBP Parametric bootstrap	EBP REML-Delta	HB
State sample size n	Non coverage percentage and Monte Carlo simulation standard error					
overall	4.9 (0.09)	6.9 (0.1)	3.9 (0.08)	4.5 (0.09)	6.4 (0.1)	4.1 (0.09)
50<=n<60 (24 states)	4.9 (0.13)	6.2 (0.15)	3 (0.11)	4.6 (0.13)	5.8 (0.14)	3.6 (0.12)
60<=n<100 (15 states)	5.6 (0.18)	8.5 (0.21)	4.7 (0.16)	5.2 (0.17)	7.9 (0.2)	4.5 (0.16)
100<=n<=690 (12 states)	3.9 (0.17)	6.2 (0.2)	4.7 (0.19)	3.6 (0.16)	5.8 (0.2)	4.7 (0.19)
	Mean width of the 95% confidence interval and Monte Carlo simulation standard error					
overall	18.7 (0.02)	17.4 (0.02)	18.7 (0.01)	19 (0.02)	17.6 (0.01)	19 (0.01)
50<=n<60 (24 states)	19.7 (0.03)	18.7 (0.02)	20.6 (0.02)	20.1 (0.02)	19 (0.02)	21 (0.02)
60<=n<100 (15 states)	19.1 (0.03)	17.6 (0.03)	19.1 (0.02)	19.4 (0.03)	17.9 (0.03)	19.4 (0.02)
100<=n<=690 (12 states)	16.1 (0.02)	14.4 (0.02)	14.5 (0.01)	16.2 (0.02)	14.5 (0.02)	14.7 (0.01)

Table 2b: Percentage of times that the 95 percent credible intervals fail to cover P_i, mean 95 percent credible interval width, along with the Monte Carlo simulation standard errors based on 1,000 simulations (in percentages) for the 3 different estimation methods - Data generated using Binomial regression (Approach B):

State sample size n	FH -model			FH-model after arcsin transformation		
	EBP Parametric bootstrap	EBP REML-Delta	HB	EBP Parametric bootstrap	EBP REML-Delta	HB
	Non coverage percentage and Monte Carlo simulation standard error					
overall	4.5 (0.09)	6.3 (0.1)	3.7 (0.08)	4.2 (0.08)	6 (0.1)	3.6 (0.08)
50<=n<60 (24 states)	3.2 (0.11)	3.9 (0.12)	2 (0.09)	3 (0.11)	3.7 (0.12)	2.1 (0.09)
60<=n<100 (15 states)	4.9 (0.17)	7.3 (0.2)	3.6 (0.15)	4.5 (0.16)	6.7 (0.19)	3.2 (0.14)
100<=n<=690 (12 states)	6.5 (0.2)	10.1 (0.24)	7.1 (0.22)	6.1 (0.2)	9.6 (0.24)	7 (0.22)
	Mean width of the 95% confidence interval and Monte Carlo simulation standard error					
overall	18.5 (0.02)	17.2 (0.02)	18.6 (0.01)	18.7 (0.02)	17.4 (0.02)	18.7 (0.01)
50<=n<60 (24 states)	19.5 (0.03)	18.5 (0.03)	20.4 (0.02)	19.8 (0.03)	18.8 (0.02)	20.5 (0.02)
60<=n<100 (15 states)	18.9 (0.03)	17.4 (0.03)	19 (0.02)	19.2 (0.03)	17.7 (0.03)	19 (0.02)
100<=n<=690 (12 states)	15.9 (0.02)	14.3 (0.02)	14.5 (0.01)	16.1 (0.02)	14.4 (0.02)	14.5 (0.01)

Table 2c reports the percentage of times that the 95% confidence interval covers the true value and the mean width of the confidence intervals along with standard Monte Carlo simulation errors for the three approaches when the simulated data was generated using Poisson regression. Again, the left half of the table reports the estimation results based on the Fay-Herriot model, while the right half of the table reports the estimation results based on the Fay-Herriot model with arcsin transformation. The noncoverage rates from the three approaches became worse than those reported in Tables 2a and 2b, though the parametric bootstrap approach still gave the best coverage compared with the other two methods. The noncoverage rates were worse for all the three approaches when the Fay-Herriot model with arcsin transformation was used compared to the case when Fay-Herriot model was used.

Table 2c: Percentage of times that the 95 percent credible intervals fail to cover P_i, mean 95 percent credible interval width, along with the Monte Carlo simulation standard errors based on 1,000 simulations (in percentages) based on 3 different estimation methods - Data generated using Binomial regression (Approach C):

State sample size n	FH -model			FH-model after arcsin transformation		
	EBP Parametric bootstrap	EBP REML-Delta	HB	EBP Parametric bootstrap	EBP REML-Delta	HB
	Non coverage percentage and Monte Carlo simulation standard error					
overall	6.7 (0.11)	8.6 (0.12)	9.3 (0.13)	8.1 (0.12)	10 (0.13)	10.6 (0.14)
50<=n<60 (24 states)	6 (0.15)	7.2 (0.17)	7.7 (0.17)	7.5 (0.17)	8.6 (0.18)	9.2 (0.19)
60<=n<100 (15 states)	6.9 (0.2)	8.5 (0.23)	9.2 (0.23)	8.5 (0.22)	10.2 (0.24)	10.7 (0.25)
100<=n<=690 (12 states)	7.7 (0.24)	11.5 (0.29)	12.5 (0.3)	8.9 (0.26)	12.5 (0.3)	13.3 (0.31)
	Mean width of the 95% confidence interval and Monte Carlo simulation standard error					
overall	24.1 (0.01)	22.9 (0.01)	23.2 (0.01)	24 (0.01)	22.9 (0.01)	23.3 (0.01)
50<=n<60 (24 states)	26.3 (0.02)	25.5 (0.02)	26.1 (0.02)	26.4 (0.02)	25.6 (0.02)	26.2 (0.02)
60<=n<100 (15 states)	24.8 (0.02)	23.4 (0.02)	23.8 (0.02)	24.6 (0.02)	23.4 (0.02)	23.8 (0.02)
100<=n<=690 (12 states)	18.7 (0.01)	16.9 (0.01)	16.9 (0.01)	18.6 (0.01)	16.9 (0.01)	16.9 (0.01)

4 Discussion

In this paper, we report the results of a simulation study from a real finite population and a simulation study where the data were generated based on presumed models to evaluate three different approaches (EBP parametric bootstrap, EBP REML-Delta and HB) in estimating the confidence/credible intervals for small area proportions based on Fay-Herriot type of models.

In the first simulation study, the design-based approach, since the true model was unknown, we chose the outcome of interest (percentage of birth weights below the grand median) which was expected to be normal so Fay-Herriot model is appropriate. We tried both excluding and including auxiliary variables in the estimation model. Although there are some differences in the coverage properties for the state finite population proportions and the use of covariates helped improve the coverage rates, none of the three approaches produced coverage rates close to the nominal rates.
In the second simulation study, the model-based approach, we tried three different models to generate the data and then applied both Fay-Herriot model and Fay-Herriot model with arcsin transformation to produce the confidence intervals using the three different approaches. The EBP parametric bootstrap approach gave the best coverage property across all the cases, though it only produced coverage rates close to the nominal rates when the data was generated using Fay-Herriot model. The HB approach gave the next best coverage property though it tended to produce over conservative credible intervals for most of the cases. For all the three approaches, there was not much difference whether to use Fay-Herriot model or Fay-Herriot model with arcsin transformation in terms coverage property.

In the design-based simulation study, since the true model was unknown, it seems very difficult to obtain confidence intervals with coverage rates close to the nominal value. Similar conclusion was found in Liu et al (2007). Based on our limited results, users of small area estimates need to be cautioned about the interpretation of the credible intervals associated with the estimates.

References

Carter, G.M., and Rolph, J.E. (1974), "Empirical Bayes methods applied to estimating fire alarm probabilities," *Journal of the American Statistical Association*, 69, 880-885.

Chatterjee, S., Lahiri, P., and Li, Huilin (2008). Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models. *The Annals of Statistics*, Vol 36, No 3, 1221-1245.

Chattopadhyay, M., Lahiri, P., Larsen, M., and Reimnitz, J. (1999). Composite estimation of drug prevalence for sub-state areas. *Survey Methodology*, 25, 81-86.

Citro, C., and Kalton, G. (Eds.). (2000). *Small-Area Income and Poverty Estimates: Priorities for 2000 and Beyond*. Washington, DC: National Academy Press.

Datta, G.S., and Lahiri, P. (2000), "A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems," *Statistica Sinica*, 10, 613-627.

Farrell, P.J., MacGibbon, B., and Tomberlin, T.J. (1997). Empirical Bayes estimators of small area proportions in multistage designs. *Statistical Sinica*, 7, 1065-1083.

Fay, R.E., and Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedure to census data. *Journal of the American Statistical Association*, 74, 269-277.

Gelman, A., and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequence. *Statistical Science*, 7, 457-472.

Jiang, J., and Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test*, 15, 111-999.

Li, H. and Lahiri, P. (2010). Adjusted maximum likelihood method in small area estimation problems. *Journal of Multivariate Analysis*, 101 (4): 882.

Liu, B., Lahiri, P., and Kalton, G. (2007). Hierarchical Bayes modeling of survey-weighted small area proportions. *Proceedings of the Survey Research Methods Section of the American Statistical Association* [CD-ROM]. Alexandria, VA: American Statistical Association, 3181-3186.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS – A Bayesian modeling framework: Concepts, structure, and extensibility, Statistics and Computing, Vol.10, No.4, pp.325-337

Malec, D., Davis, W., and Cao, X. (1999). Small area estimates of overweight prevalence using sample selection adjustment. *Statistics in Medicine*, 18, 3189-3200.

Malec, D., Sedransk, J., Moriarity, C.L., and Lecler, F.B. (1997). Small area inference for binary variables in National Health Interview Survey. *Journal of the American Statistical Association*, 92, 815-826.

Mohadjer, L., Rao, J.N.K., Liu, B., Krenzke, T., and Van De Kerckhove, W. (2012). Hierarchical Bayes small area estimates of adult literacy using unmatched sampling and linking models. *Journal of the Indian Society of Agricultural Statistics*, 66 (1), 55-63.

Rao, J.N.K. (2003). *Small area estimation*. New York: John Wiley and Sons.

Wright, D., Sathe, N., and Spagnola, K. (2007). *State Estimates of Substance Use from the 2004-2005 National Surveys on Drug Use and Health.* (DHHS Publication No. SMA 07-4235, NSDUH Series H-31). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.