Further Comparisons of Unit- and Area-Level Small Area Estimators

Robert E. Fay[1]
[1]Westat, Inc., 1600 Research Blvd., Rockville, MD 20850

Abstract
Small area estimation (SAE) uses explicit or implicit statistical models to estimate characteristics for geographic areas or other domains where the available survey data are insufficient to produce acceptably reliable direct estimates. In practice, most SAE procedures are now based on small area models that explicitly account for the error in predicting the target characteristic given the auxiliary data. These SAE procedures can be classified as to whether the model is expressed at the area level using direct survey estimates for the areas being modeled or at the unit level, typically at the level of the individual survey response. In a 2016 paper, Hidiroglou and You examined the performance of some unit- and area-level SAE procedures for simulated samples from a known population, finding that unit-level estimators had a distinct advantage over area-level ones. This paper expands their simulation to a broader set of circumstances and estimators in order to assess the generality of their findings.

Key Words: Pseudo-EBLUP, EBLUP, Fay-Herriot model, Survey regression estimator

1. Introduction

The small area estimation problem is widely recognized in the statistical profession. The problem arises when one or more surveys produce direct estimates of acceptable reliability for the total population and perhaps a limited set of areas or domains, but interest extends to a larger set of smaller areas where the reliability of the direct survey estimates is limited. Indeed, in some applications, some of the small areas may lack any representation in the available sample.

Small area estimation models interpret the observed sample data with statistical models that explicitly account for error in predicting the target characteristic given available auxiliary data. The literature on small area models continues to grow vigorously. In the year in which his book appeared, Rao (2003) provided an extensive review of the underlying theory of small area models and discussed many specific applications; Rao and Molina (2015) more recently updated and expanded this work. The range of models and their associated estimation approaches can be divided along two primary dimensions, which is reflected in the organization of these books. One of these is the inferential basis, where frequentist, empirical Bayes, and hierarchical Bayes alternatives are all possible. The other is the distinction between unit-level and area-level models. Unit-level models are formulated at the individual level and build up the small area estimates from unit-level predictions. Area-level models are instead formulated only at the area level. Of course, small area models can vary along several other dimensions.

In a paper titled "Comparison of Unit Level and Area Level Small Area Estimators," Hidiroglou and You (2016) used simulation to compare the performance of some frequentist unit-level and area-level small area models. At the unit level, they considered both the model introduced by Battese, Harter, and Fuller (1988), which is a form of empirical best linear unbiased estimation (EBLUP), and the pseudo-EBLUP of You and Rao (2002). In contrast to the EBLUP, in which the survey weights play no role, the pseudo-EBLUP incorporates the survey weights in the estimator. Hidiroglou and You (2016) represented area-level small area estimation by three forms of the approach of Fay and Herriot (1979). The best performing of the three, denoted FH-HA, applied an area level model to the weighted Hájek estimator, that is, to the familiar ratio of the weighted sum of the variable divided by the sum of the weights. The Hájek estimator is more typically called the direct estimator.

Hidiroglou and You presented simulation results in detail for two scenarios, both for 30 areas. For each scenario, they simulated one fixed finite population as the standard, and then measured the performance of each estimator over 3,000 repetitions of the sampling. In each scenario, they considered sample sizes of 10 and 30 per area. They investigated the characteristics of the estimates for individual areas, the average overall performance, the estimation of mean square errors, and the coverage properties of the confidence intervals derived from the estimated mean square errors.

In terms of the potential implications for practice, Hidiroglou and You's most striking result is the consistently strong superiority of the pseudo-EBLUP over the area-level model in terms of average overall performance. Table 1 extracts their key findings for the average relative root mean square error, where both EBLUP and pseudo-EBLUP outperform FH-HA. The comparison between EBLUP and pseudo-EBLUP is mixed, where the gain from incorporating survey weights is evident only under Scenario II.

Table 1: Average Percent Relative Root Mean Square Error Reported by Hidiroglou and You (2016)

	Estimator	n=10	n=30
		Scenario I	
Unit level	EBLUP	4.98	3.01
Unit level	Pseudo-EBLUP	5.49	3.58
Area level	FH-HA	9.68	6.51
		Scenario II	
Unit level	EBLUP	6.78	5.62
Unit level	Pseudo-EBLUP	5.42	3.21
Area level	FH-HA	11.21	6.79

For the average percent absolute relative bias shown in Table 2, the unit-level EBLUP performed the best of the three under scenario I, but the pseudo-EBLUP was best under scenario II. Indeed, under scenario II, the average percent relative bias of EBLUP is the worst of the three.

This paper will examine the generality of these findings by considering a broader set of populations, sampling situations, and estimators. The "Further" in the title of this paper signals the goal of building on the results of Hidiroglou and You. The next section details the estimators considered by Hidiroglou and You and then introduces additional estimators based on incorporating survey regression estimation as a first step in the small area model. This step is not new: Bijlsma et al. (2016) previously applied it, and there may be several other such precedents. The third section describes the simulation design. The fourth section presents the findings from the simulation. The discussion section summarizes the conclusions and identifies remaining open questions.

Table 2: Average Percent Absolute Relative Bias Reported by Hidiroglou and You (2016)

	Estimator	*n*=10	*n*=30
		Scenario I	
Unit level	EBLUP	1.71	0.75
Unit level	Pseudo-EBLUP	2.14	0.86
Area level	FH-HA	4.33	2.59
		Scenario II	
Unit level	EBLUP	4.31	4.52
Unit level	Pseudo-EBLUP	0.25	0.12
Area level	FH-HA	3.48	1.47

2. Unit- and Area-Level Small Area Estimators

2.1 Unit-Level Estimation

Hidiroglou and You employed the sampling setup in Rao (2003). A universe U of size N is divided into m non-overlapping small areas U_i of size N_i, for $i = 1,\ldots,m$. Probability-based samples s_i of size n_i are drawn. Hidiroglou and You specifically considered sampling with replacement with n_i fixed, for which survey weights $w_{ij} = n_i^{-1} p_{ij}^{-1}$ were based on selection probabilities p_{ij} for unit j at each sample draw. The Hájek estimator for the population mean $\bar{Y}_i = N_i^{-1} \sum_{U_i} y_{ij}$ is $\hat{\bar{Y}}_{iw} = \sum_{s_i} w_{ij}\, y_{ij} / \sum_{s_i} w_{ij}$.

Battesse, Harter, and Fuller (1988) introduced the basic unit level model,

$$y_{ij} = \boldsymbol{x'}_{ij}\boldsymbol{\beta} + v_i + e_{ij} \tag{2.1}$$

for $j = 1, \ldots, N_i, i = 1, \ldots, m$, where $\boldsymbol{x'}_{ij} = (x_{ij1}, \ldots, x_{ijp})$ is a column vector of auxiliary variables with $x_{ij1} = 1$, and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{p-1})'$ is a column vector of regression parameters. The random effects $v_i \sim_{iid} N(0, \sigma_v^2)$ are assumed independent of the unit errors $e_{ij} \sim_{iid} N(0, \sigma_e^2)$.

When N_i is much larger than n_i, the estimation problem for $\bar{Y}_i$ is essentially equivalent to estimating

$$\theta_i = \bar{\boldsymbol{X}}_i'\boldsymbol{\beta} + v_i \tag{2.2}$$

where $\bar{X}_i = N_i^{-1} \sum_{j=1}^{N_i} x_{ij}$. The best linear unbiased prediction (BLUP) of θ_i under (2.1) is

$$\tilde{\theta}_i = r_i \bar{y}_i + (\bar{\boldsymbol{X}}_i - r_i \bar{\boldsymbol{x}}_i)'\, \tilde{\boldsymbol{\beta}} \tag{2.3}$$

where $\bar{y}_i = n_i^{-1} \sum_{s_i} y_{ij}$, $\bar{\boldsymbol{x}}_i = n_i^{-1} \sum_{s_i} x_{ij}$, $r_i = \sigma_v^2/(\sigma_v^2 + \sigma_e^2/n_i)$, and

$$\tilde{\beta} = \left(\sum_{i=1}^{m} \mathbf{x}_i' \mathbf{V}_i^{-1} \mathbf{x}_i\right)^{-1} \left(\sum_{i=1}^{m} \mathbf{x}_i' \mathbf{V}_i^{-1} y_i\right) \tag{2.4}$$

with $\mathbf{x}_i{}' = (x_{i1}, \dots, x_{in_i})$, $V_i = \sigma_v^2 1_{n_i} 1'_{n_i} + \sigma_e^2 \mathbf{I}_{n_i}$, $y_i = (y_{i1}, \dots, y_{in_i})'$, $i = 1, \dots m$, and where 1_{n_i} is a column vector of ones, and $\mathbf{I}_{n_i}$ is the identity matrix of rank n_i. An empirical best linear unbiased predictor (EBLUP) results from estimating σ_v^2 and σ_e^2, then using them to estimate $\hat{\beta}$ from (2.4), $\hat{r}_i$, and $\hat{\theta}_i^{EBLUP}$ from (2.3).

Maximum likelihood or restricted maximum likelihood are typical choices to estimate σ_v^2 and σ_e^2, but Hidiroglou and You used the method of fitting of constants, which is of interest here because of its role in the pseudo-EBLUP method. The method involves estimating σ_v^2 and σ_e^2 from two unweighted regressions. The regression to estimate σ_e^2 involves centering about the area-level sample averages, thereby eliminating the area level random effect. Specifically, the regression is of $y_{ij} - \bar{y}_i$ on $\mathbf{x}_{ij} - \bar{\mathbf{x}}_i$ giving

$$\dot{\beta} = \left(\sum_{i=1}^{m} \sum_{j \in s_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'\right)^{-1} \left(\sum_{i=1}^{m} \sum_{j \in s_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(y_{ij} - \bar{y}_i)\right) \tag{2.5}$$

Residuals $\hat{\varepsilon}_{ij} = y_{ij} - \bar{y}_i - (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)\dot{\beta}$ are then used to estimate $\hat{\sigma}_e^2 = (n - m - p + 1)^{-1} \sum_{i=1}^{m} \sum_j \hat{\varepsilon}_{ij}^2$. (Adjustments are required if the matrix inverted in (2.5) is of rank less than $(p-1)$). The second regression is of y_{ij} on $\mathbf{x}_{ij}$, the residuals of which, $\hat{u}_{ij}$, can be used to estimate $\hat{\sigma}_v^2 = \max(0, n_*^{-1} \sum_{i=1}^{m} \sum_{j \epsilon s_i} \hat{u}_{ij}^2 - (n-p)\hat{\sigma}_e^2)$, where $n_* = \sum_{i=1}^{m} n_i - tr\left[(\mathbf{X}'\mathbf{X})^{-1} \sum_{i=1}^{m} n_i^2 \bar{x}_i \, \bar{x}_i{}'\right]$ for $\mathbf{X}' = (x'_i, \dots, x'_m)$.

You and Rao (2002) introduced a pseudo-EBLUP unit-level model that incorporates the survey weights. The same population model, (2.1), is assumed as for EBLUP estimator, but if the survey weights are informative, the unweighted estimator (2.4) may not consistently estimate the population parameter. The variance impact of incorporating survey weights w_{ij} on estimating Y_i with $\bar{y}_{iw} = \sum_{j \epsilon s_i} w_{ij} y_{ij} / \sum_{j \epsilon s_i} w_{ij}$ is assumed to be $\delta_i^2 = \sum_{s_i} \tilde{w}_{ij}^2$ where $\tilde{w}_{ij} = w_{ij} / \sum_{s_i} w_{ij}$. You and Rao showed that the BLUP in this situation is

$$\tilde{\theta}_{iw} = r_{iw} \bar{y}_{iw} + (\bar{\mathbf{X}}_i - r_{iw} \bar{\boldsymbol{x}}_{\boldsymbol{iw}})' \, \tilde{\boldsymbol{\beta}}_{\boldsymbol{w}} \tag{2.6}$$

where $r_{iw} = \sigma_v^2 / (\sigma_v^2 + \delta_i^2 \sigma_e^2)$ and

$$\tilde{\boldsymbol{\beta}}_w = \left(\sum_{i=1}^{m} \sum_{s_i} w_{ij} \, \boldsymbol{x}_{ij} \, (\boldsymbol{x}_{ij} - r_{iw} \bar{\mathbf{x}}_{iw})\right)^{-1} \left(\sum_{i=1}^{m} \sum_{s_i} w_{ij} \, \boldsymbol{y}_{ij} \, (\boldsymbol{x}_{ij} - r_{iw} \bar{\mathbf{x}}_{iw})\right). \tag{2.7}$$

They used the fitting of constants method to estimate σ_v^2 and σ_e^2 without incorporating weights, but then estimated $\hat{r}_{iw}$ and substituted it into (2.7) and (2.6).

2.2 Area-Level Estimation

Fay and Herriot (1979) introduced an area-level model

$$\bar{y}_{iw} = \bar{\boldsymbol{X}}_i' \boldsymbol{\beta} + v_i + e_i \tag{2.8}$$

where the term $e_i \sim_{iid} N(0, \sigma_{ei}^2)$ reflects the effect of sampling error on the estimate $\bar{y}_{iw}$, and $v_i \sim N(0, \sigma_v^2)$. The BLUP of $\theta_i = \bar{\boldsymbol{X}}_i' \boldsymbol{\beta} + v_i$ is

$$\tilde{\theta}_{iw} = r_i \bar{y}_{iw} + (1 - r_i)\bar{\boldsymbol{X}}_i{}'\tilde{\boldsymbol{\beta}}_{\boldsymbol{WLS}} \tag{2.9}$$

where $r_i = \sigma_v^2/(\sigma_v^2 + \sigma_{ei}^2)$ and

$$\tilde{\boldsymbol{\beta}}_{\boldsymbol{WLS}} = \left(\sum_{i=1}^{m} \bar{\boldsymbol{X}}_i{}'V_i^{-1}\bar{\boldsymbol{X}}_i\right)^{-1} \left(\sum_{i=1}^{m} \bar{\boldsymbol{X}}_i \boldsymbol{V}_i^{-1}\bar{y}_{iw}\right) \tag{2.10}$$

with $V_i = \sigma_v^2 + \sigma_{ei}^2$. If the σ_{ei}^2 are assumed known, then an EBLUP can be formed by estimating σ_v^2 through maximum likelihood, restricted maximum likelihood (REML), or the method of moments.

As previously noted, Hidiroglou and You considered three alternative estimators for Y_i, but only the best performing of the three, the Hájek estimator, will be considered here. Two forms of the survey regression estimator for Y_i will also be included in the comparison. The usual form of the survey regression estimator (SRE) is based on a weighted unit-level regression

$$\tilde{\beta} = \left(\sum_{i=1}^{m} \mathbf{x}_i{}'\boldsymbol{W}_{\boldsymbol{i}}\mathbf{x}_i\right)^{-1} \left(\sum_{i=1}^{m} \mathbf{x}_i{}'\boldsymbol{W}_{\boldsymbol{i}}y_i\right) \tag{2.11}$$

where, as in the setup for the unit-level model (2.4), $\mathbf{x}_i{}' = (x_{i1}, \dots, x_{in_i})$, $y_i = (y_{i1}, \dots, y_{in_i})'$, $i = 1, \dots m$, and $W_i = diag(w_{i1}, \dots, w_{in_i})$ is a diagonal matrix of weights. The survey regression estimator is

$$\tilde{y}_i = \bar{y}_{iw} + (\bar{\boldsymbol{X}}_i - \bar{\boldsymbol{x}}_{\boldsymbol{iw}})'\,\tilde{\boldsymbol{\beta}} \tag{2.12}$$

The standard variance estimator for (2.12) is based on the residuals from the regression (2.11). The area level model can then be based on $\tilde{y}_i$ and its estimated variance instead of the Hájek estimator, $\bar{y}_{iw}$ and an estimate of σ_{ei}^2.

A third alternative is to adopt the approach to estimating the regression in (2.5)

$$\dot{\beta}_w = \left(\sum_{i=1}^{m}\sum_{j\in s_i}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)w_{ij}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'\right)^{-1} \left(\sum_{i=1}^{m}\sum_{j\in s_i}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)\,w_{ij}(y_{ij} - \bar{y}_i)\right) \tag{2.13}$$

for use in (2.12). The rationale for attempting this approach is that it removes area-level variation from the estimation of the regression, similar to the approach in the fitting of constants method. In this paper this variant will be termed the *modified survey regression estimator* (MSRE).

3. Simulation Design

Hidiroglou and You (2016) created a single population for each of their two scenarios they presented in detail. They reported

> Each finite population had $m = 30$ small areas, and each area consisted of $N_i = 200$ population units. Each finite population was generated using the unit level model $y_{ij} = \beta_0 +$

> $x_{ij}\beta_1 + v_i + e_{ij}$. The auxiliary variable x_{ij} was generated from an exponential distribution with mean 4 and variance 8, and the random components were generated from a normal distribution with $v_i \sim N(0, \sigma_v^2)$, $e_{ij} \sim N(0, \sigma_e^2)$, where $\sigma_v^2 = 100$ and $\sigma_e^2 = 225$. For the first population, the regression fixed effects were set as β_0 =50, $\beta_1 = 10$ for all 30 areas. For the second population, different fixed effects values were used: β_0 =50, $\beta_1 = 10$ for areas $m = 1,\ldots,10$; $\beta_0 = 75$, $\beta_1 = 15$ for areas $m = 11,\ldots,20$; $\beta_0 = 100$, $\beta_1 = 20$ for areas m =21,…,30. We had three different means for the fixed effects $\beta_0 + x_{ij}\beta_1$ in the second population, whereas we only had one in the first population. PPSWR samples within each area were drawn independently from each constructed population. PPSWR sampling was implemented as follows: We first defined a size measure z_{ij} for a given unit (i, j). Using these z_{ij} values, we computed selection probabilities $p_{ij} = z_{ij}/\sum_j z_{ij}$ for each unit (i, j) and used them to select PPSWR samples of equal size $n_i = n$. Within each generated population, we selected samples of size n =10 and 30. The basic design weight is given by $w_{ij} = n_i^{-1}p_{ij}^{-1}$, so that the standardized weight is $\widetilde{w}_{ij} = p_{ij}^{-1}/\sum_j p_{ij}^{-1}$.

Except for a difficulty of replicating the exponential distribution, the preceding description is clear. An exponential distribution with mean 4 and variance 16 was used below instead. The next part of their account is less clear.

> We chose the size measure z_{ij} as a linear combination of the auxiliary variable x_{ij} and data generated from an exponential distribution with mean 4 and variance 16. The correlation coefficient ρ between y_{ij} and the selection probability in each area varied between 0.02 and 0.95. The range of the p_{ij}'s corresponds to non-informative selection ($\rho = 0.02$) to strongly informative selection ($\rho = 0.95$) of the PPSWR samples.

The description does not specify the linear combinations except in the form of a resulting correlation. But it is not clear how the correlation is determined; for example, is it computed theoretically or from the sample? The simulation reported here considered four sampling designs, described below, with two of them attempting to approximate the low and high correlation scenarios considered by Hidiroglou and You.

First, six finite populations were created, each with 30 areas of size $N_i = 200$. For each population a single covariate x_{ij} and its area level mean $\bar{x}_i$ were generated, then y_{ij} was generated conditional on x_{ij}. Estimators were assessed for their predictions of the finite population means $\bar{Y}_i$. Samples of size n_i = 10 and 30 were drawn with replacement from each according to four different sample designs. For each simulation, the sampling was repeated 10,000 times.

Six populations were considered:

Population 1

- x_{ij} from exponential distribution mean 4, variance 16
- $y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + v_i + e_{ij}$ where $v_i \sim N(0,\ 100\ )$, $e_{ij} \sim N(0,\ 225)$ $\beta_0 = 50,\ \beta_1 = 10$

Population 2

- Similar to population 1, except
  - $\beta_0 = 50,\ \beta_1 = 10$ for i=1,…,10;
  - $\beta_0 = 75,\ \beta_1 = 15$ for i=11,…,20;
  - $\beta_0 = 100,\ \beta_1 = 20$ for i=21,…,30;

Population 3

- x_{ij} from exponential distribution mean 3, variance 9 *plus* an area-level effect from exponential mean 1, variance 1
- $y_{ij} = \boldsymbol{x'}_{ij}\boldsymbol{\beta} + \beta_2\bar{X}_i + \nu_i + e_{ij}$ where $\nu_i \sim N(0,\ 25)$, $e_{ij} \sim N(0,\ 225)$, $\beta_0 = 50$, $\beta_1 = 5, \beta_2 = 5$

Population 4

- Similar to population 3, except
  - $\beta_0 = 50,\ \beta_1 = 5, \beta_2 = 5$, for i=1,…,10;
  - $\beta_0 = 75,\ \beta_1 = 7.5, \beta_2 = 7.5$ for i=11,…,20;
  - $\beta_0 = 100,\ \beta_1 = 10, \beta_2 = 10$ for i=21,…,30;

Population 5

- x_{ij} from exponential distribution mean 3, variance 9 *plus* an area-level effect from exponential mean 1, variance 1
- $y_{ij} = \boldsymbol{x'}_{ij}\boldsymbol{\beta} + \beta_2\bar{X}_i + \nu_i + e_{ij}$ where $\nu_i \sim N(0,\ 25)$, $e_{ij} \sim N(0,\ 225)$, $\beta_0 = 50$, $\beta_1 = 0, \beta_2 = 10$

Population 6

- Similar to population 5, except
  - $\beta_0 = 50,\ \beta_1 = 0, \beta_2 = 10$, for i=1,…,10;
  - $\beta_0 = 75,\ \beta_1 = 0, \beta_2 = 15$ for i=11,…,20;
  - $\beta_0 = 100,\ \beta_1 = 0, \beta_2 = 20$ for i=21,…,30;

Populations 3 through 6 include area-level effects either in addition to or instead of the unit-level effects assumed by the unit-level models.

Four different sampling designs were implemented:

1. SRSWR, with equal probabilities
2. PPSWR proportional to a random variable z_{ij} with mean 4 variance 16
3. PPSWR proportional to x_{ij}
4. PPSWR proportional to $|y_{ij}|$

The combination of Population 1 with sample design 2 approximates the Scenario I studied by Hidiroglou and You, and the combination of Population 2 with sample design 3 approximates Scenario II.

4. Results

The results of the simulations will be presented graphically. For comparison, Figure 1 summarizes the Hidiroglou and You results presented in Table 1. As noted previously, the area-level FH-HA consistently had the highest relative root mean square errors of the three competitors. The pseudo-EBLUP shows a definite advantage over EBLUP under Scenario II. Note that for Scenario I, the relative ranking of the three estimators does not change going from a sample size of 10 to 30. Similarly, the change in sample size does not alter the relative ranking within Scenario II.

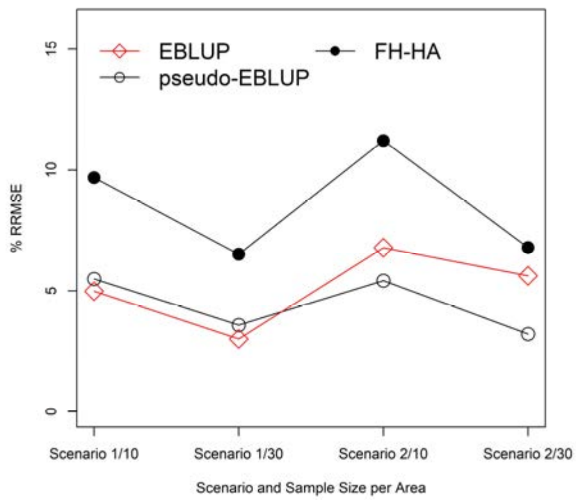


Figure 1: Percent relative root mean square errors reported by Hidiroglou and You (2016)

Figure 2 presents the results for Population 1 over the four sample designs and two sample sizes. In general, the findings are typically slightly better for the modified survey regression estimator (MSRE) than the survey regression estimator (SRE), so only the MSRE and the area-level FH-MSRE are shown in the graphs. Consistent with Hidiroglou and You's results, FH-HA is not competitive with the other estimators for Population 1, but application of an area-level model to MSRE results in an estimator, FH-MSRE, virtually tied with pseudo-EBLUP. Notably, the model-assisted estimator MSRE performs almost as well as those two, indicating that it is almost competitive with the model-based estimators for this population. Generally, EBLUP does well for the first three of the four sampling schemes. That it should do well for the first two sample designs is entirely expected, because either the weights are constant or entirely uninformative, in the second case putting methods that use weights at a disadvantage.

As remarked previously, the combination of Population 1 with sample design 2 approximates Scenario I of Hidiroglou and You, and the ordering of EBLUP, pseudo-EBLUP, and FH-HA agrees with their findings.

For Population 2 in Figure 3, the performances of MSRE, FH-MSRE, and the pseudo-EBLUP are virtually indistinguishable, so the model-based estimators are barely improving upon MSRE, the model-assisted alternative. On the right-hand side of the figure, the unit-level EBLUP is at a disadvantage relative to the alternatives incorporating the sample weights. The third sample design for Population 2 approximates Scenario II, and the ordering of pseudo-EBLUP, EBLUP, and FH-HA is similar to Hidiroglou and You's findings, except that FH-HA outperforms EBLUP for a sample size of 30 units per area.

In Populations 1 and 2, the large size of σ_v^2 =100 relative to $\sigma_e^2 = 225$ leads to variances for the area-level means of 22.5 or 7.5. Consequently, between area variance dominates the estimation problem, so the smoothing estimators barely improve upon the model-assisted alternative. The reduction of σ_v^2 to 25 in Populations 3 through 6 was chosen to encourage more separation of the model-based and model-assisted estimators.

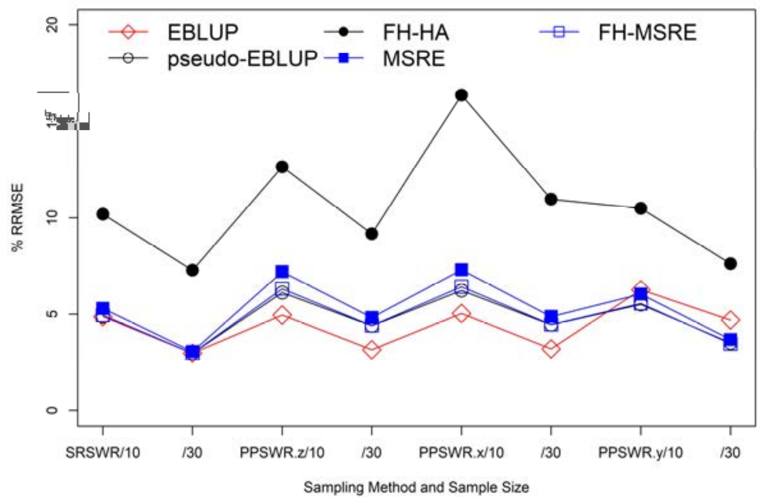


Figure 2: Percent relative root mean square errors for Population 1

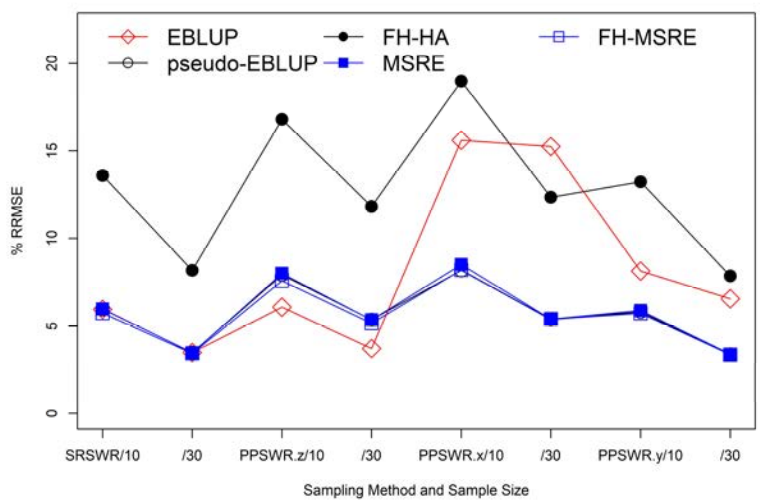


Figure 3: Percent relative root mean square errors for Population 2

In Figure 4, a clear advantage of FH-MSRE over MSRE is evident for Population 3. Including an area effect as well as a unit effect when generating the population resulted in a slight but distinct advantage for FH-MSRE over pseudo-EBLUP. EBLUP does well except under the fourth sampling scheme.

In Figure 5, the advantage of FH-MSRE over MSRE again narrows for Population 4. There is a barely discernable advantage of FH-MSRE over pseudo-EBLUP. EBLUP is adversely affected under both sample designs 3 and 4.

In Figure 6, Population 5 has only area-level effects, so both FH-HA and FH-MSRE perform essentially identically and better than pseudo-EBLUP. The unit-level regression fails to pick up the area-level effect correctly.

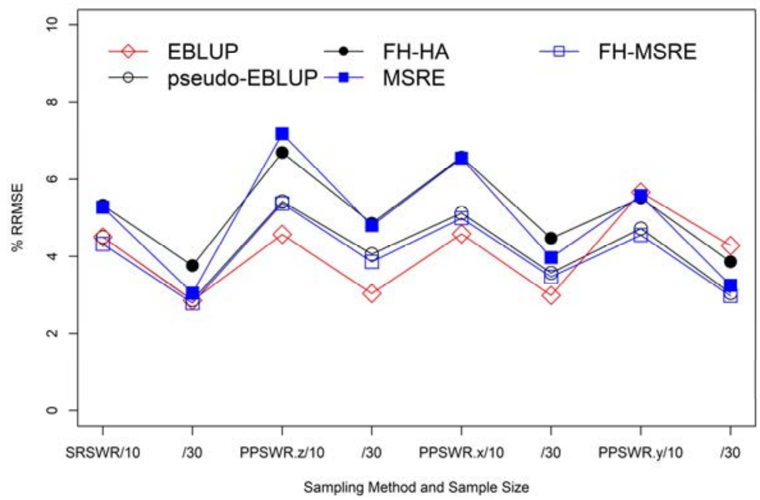


Figure 4: Percent relative root mean square errors for Population 3

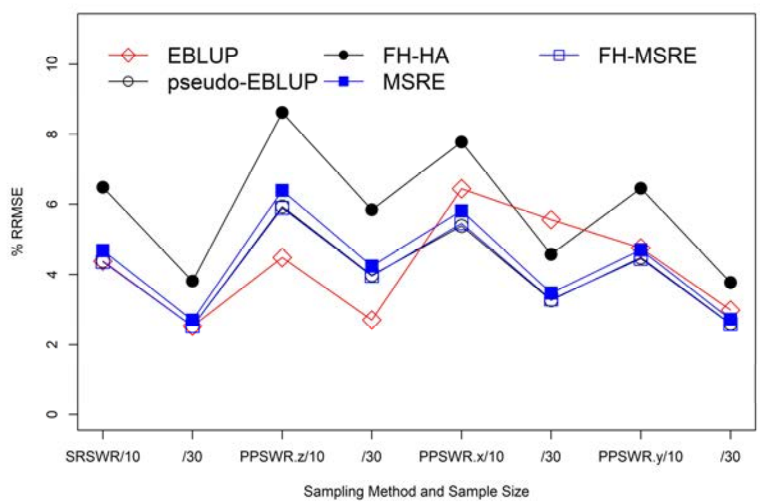


Figure 5: Percent relative root mean square errors for Population 4

In Figure 7, pseudo-EBLUP, FH-HA, and FH-MSRE perform identically in terms of relative root mean square error.

To avoid crowding the figures further, FH-SRE was omitted from Figures 2 through 7, but the Appendix presents comparisons of pseudo-EBLUP, FH-MSRE, and FH-SRE.

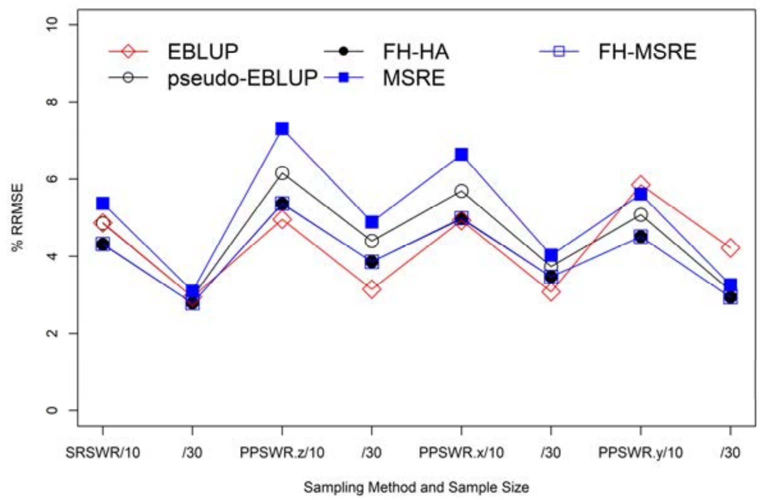


Figure 6: Percent relative root mean square errors for Population 5

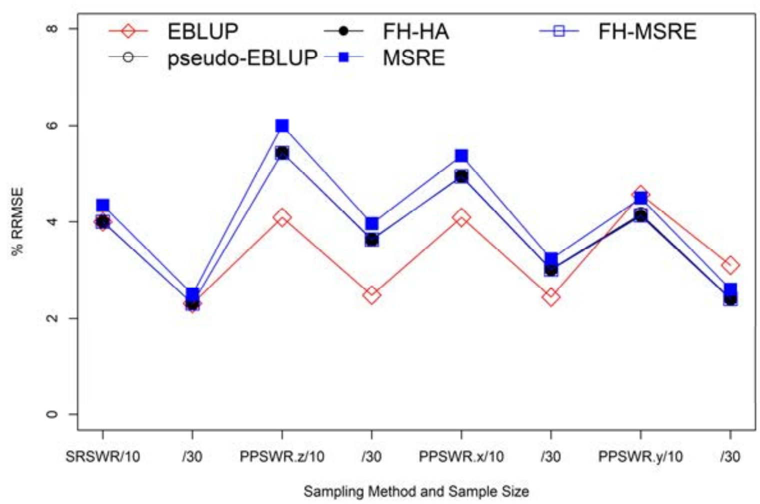


Figure 7: Percent relative root mean square errors for Population 6

The principal focus of the simulation study was on estimators that incorporate weights, and EBLUP was included for comparison. Few of the simulations show EBLUP to be dramatically worse than the weighted versions. As noted earlier, the first two sample designs were favorable to EBLUP, so only the third and fourth sampling designs provided situations to illustrate possible advantages of incorporating the survey weights. In fact, EBLUP often performed well under the third sample design. As a possible explanation, the third sampling design based the sampling probability on x_{ij}, which was also strongly predictive of y_{ij}, so use of x_{ij} in the EBLUP model may have captured most of the information in the weights.

5. Discussion

The Hidiroglou and You paper examined topics such as mean square error estimation and confidence coverage not considered here. The focus here was more narrowly on estimator performance based on mean square error. An apparent conclusion from Hidiroglou and Your (2016) that the unit-level model reliably outperforms an area level version disappears with the option of applying area level models to a survey regression estimator.

The unit-level model and the survey regression estimator have similar requirements for their use; in particular, the unit-level observations x_{ij} and area-level aggregates X_i must come from the same or equivalent sources.

Further research will be needed to investigate issues such as MSE estimation under the simulation conditions studied here.

Acknowledgements

The views expressed in this paper are solely those of the author and not of Westat or any of its clients.

References

Bijlsma, I.; van den Brakel, J.; van der Velden, R.; and Allen, J. (2017), "Estimating Literacy Levels at a Detailed Regional Level: An Application Using Dutch Data," ROA Research Memorandum, ROA-RM-2017/6, Research Center for Education and the Labor Market, June 2017.

Fay, R.E. and Herriot, R.A. (1979), "Estimation of Income for Small Places: An Application of James-Stein Procedures to Census Data," *Journal of the American Statistical Association*, *74*, 268-277.

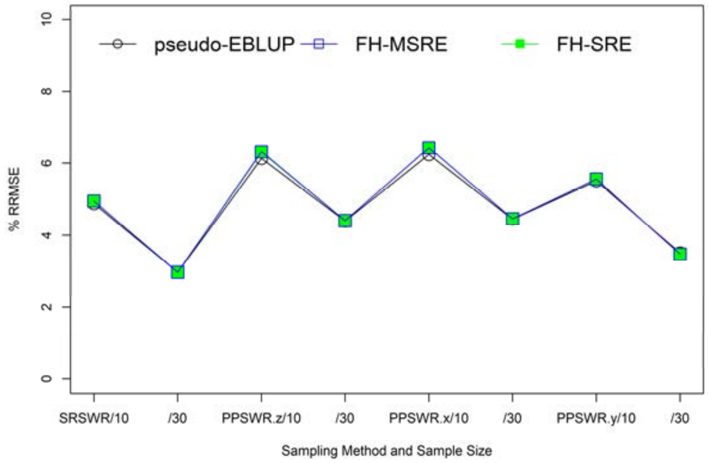
Hidiroglou, M.A. and You, Y. (2016), "Comparison of Unit Level and Area Level Small Area Estimators," *Survey Methodology*, *42*, 41-61.

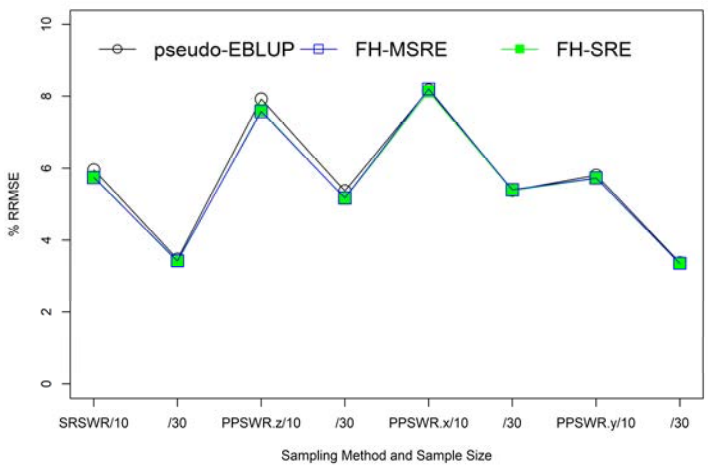
Rao, J.N.K. (2003), *Small Area Estimation*, New York: John Wiley & Sons, Inc.

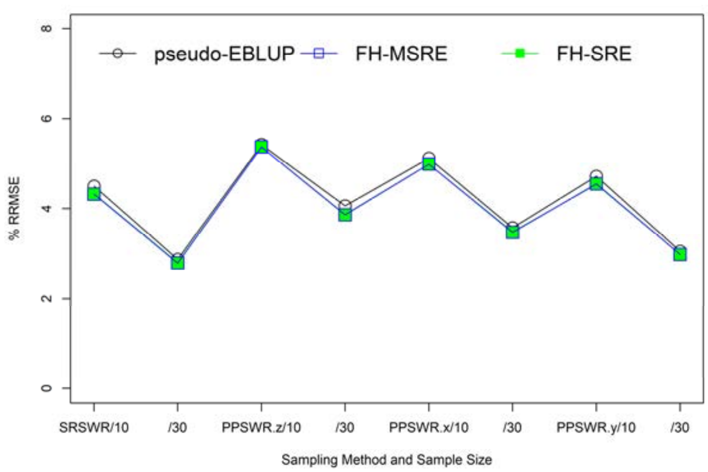
Rao, J.N.K. (2003) and Molina, I. (2015), *Small Area Estimation, 2nd Edition*, Hoboken, NJ: John Wiley & Sons, Inc.

You, Y. and Rao, J.N.K. (2002), "A Pseudo Empirical Best Linear Unbiased Prediction Approach to Small Area Estimation Using Survey Weights," *The Canadian Journal of Statistics, 30,* 431-439.

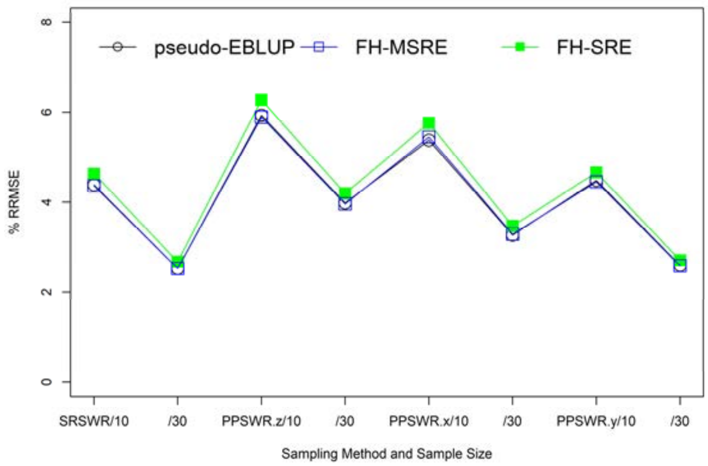
Appendix

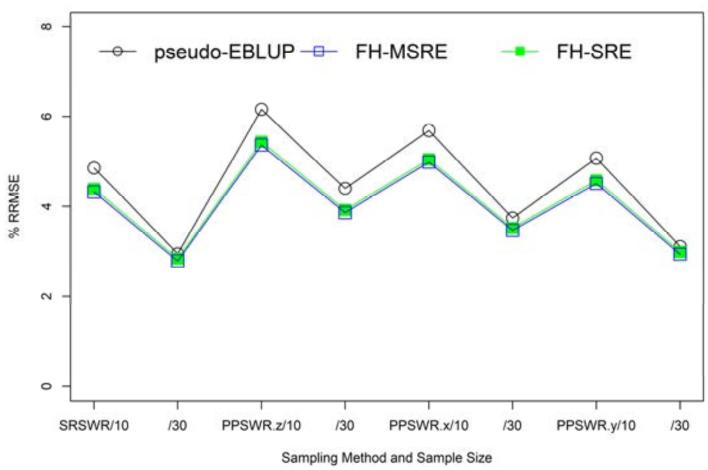


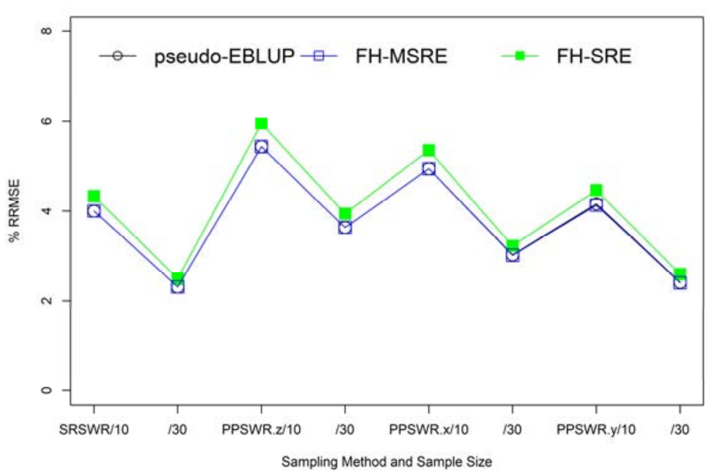




Figures A1-A3: Percent relative root mean square errors for Populations 1-3, respectively







Figures A4-A6: Percent relative root mean square errors for Populations 4-6, respectively