Comparative Study of MSE Estimates for Small Area Models Under Different Sampling Variances*

Kyle Irimata[†] Gauri S. Datta[‡†] Jerry Maples[†]

Abstract
The Fay-Herriot model is a popular linear mixed effects model for estimating small area means. Many approximations for the mean squared error (MSE) of the empirical best linear unbiased predictor of the small area means have been produced for this model. Amongst other things, these MSE approximations depend on the estimated sampling variance and estimates of the model parameters. However, estimation of the random effects variance can be difficult when the sampling variances are comparatively large and dispersed, which in turn can impact the estimation of the MSE. We compare the estimation of the random effects variance and the corresponding MSE estimates for various approaches such as those proposed by Prasad and Rao and by Fay and Herriot, under different sampling variance patterns and random effects distributions. For illustration, we use data from the American Community Survey and tax records to estimate childhood poverty in the U.S.

Key Words: Small area estimation, Fay-Herriot, Mean squared error

1. Introduction

Small area estimation (SAE) deals with estimation of one or more characteristics of subpopulations which may lack samples that are large enough to produce reliable estimates. In this case, small areas often refer to geographic areas, in which the available sample from an area may be small. For instance, the American Community Survey (ACS) is an annual survey administered by the U.S. Census Bureau, which collects more detailed information on the U.S. population than does the decennial census. These data can be used to investigate outcomes such as poverty status for children in small areas, such as counties. In these cases, the sample sizes for some small areas may be very small, or even zero. SAE is often primarily focused on point estimates, such as prediction of the mean for each small area. However, estimation of the mean squared error (MSE) is also an important and challenging problem (Pfeffermann 2013).

The Fay-Herriot model, which is an area level model, is very popular in SAE for producing small area estimates. For m small area population means, denoted by $\theta_i, i = 1, ..., m$, with direct estimates given by Y_i, the model takes the form

$$\theta_i = x_i'\beta + v_i \qquad Y_i = \theta_i + e_i \qquad ,$$

where x_i denotes the p vector of covariates, with coresponding vector of regression coefficients β. We assume that the sampling errors $e_i \sim N(0, D_i)$ are independent, and also independent of the random model errors v_i. We further assume that $E(v_i) = 0$, and $V(v_i) = \sigma^2$. Let us denote $\mathbf{Y} = (Y_1, ..., Y_m)^T$, $X = (x_1, ..., x_m)^T$, $\Sigma = D + \sigma^2 I$, where $D = diag(D_1, ..., D_m)$ and I is the identity matrix. We assume that $rank(X) = p$.

The best linear unbiased predictor (BLUP) of θ_i for the Fay-Herriot model when σ^2 is known is given by:

$$\tilde{\theta}_i(\sigma^2, Y) = Y_i - B_i(Y_i - x_i^T\tilde{\beta}),$$

*This report is released to inform interested parties of ongoing research and to encourage discussion. The views expressed are those of the authors and not those of the U.S. Census Bureau

[†]Center for Statistical Research and Methodology, U.S. Census Bureau

[‡]Department of Statistics, University of Georgia

which is a weighted average of the direct survey estimate Y_i and the synthetic regression predictor $x_i^T\tilde{\beta}$, where $\tilde{\beta} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y$. Here, $B_i = D_i/(\sigma^2 + D_i)$ is the shrinkage coefficient, which shrinks more to the synthetic regression for large sampling variances D_i. Note that the estimation of the small area mean θ_i is dependent on the estimate of the variance term σ^2 through B_i and $\tilde{\beta}$. Rao and Molina (2015) provide a thorough overview of small area models.

Many estimators of σ^2 have been proposed. One of the least computationally burdensome estimators was introduced by Prasad and Rao (1990), and others (Lahiri and Rao, 1995). This closed form estimator is given by:

$$\hat{\sigma}^2_{PR} = \frac{Y^T(I-H)Y - tr[D(I-H)]}{m-p}, \tag{1}$$

where $H = X(X^TX)^{-1}X^T$.

Fay and Herriot (1979) originally proposed an estimator based on the solution to the estimating equation below

$$Q(\sigma^2) = Y^T[\Sigma^{-1} - \Sigma^{-1}X(X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}]Y = m-p, \tag{2}$$

where the corresponding estimator of the variance is denoted by $\hat{\sigma}^2_{FH} = Q^{-1}(m-p)$, provided $Q(0) > m-p$. Here, $Q(\sigma^2)$ is a monotonically decreasing function of σ^2.

Other approaches include the use of the maximum likelihood (ML) or residual maximum likelihood (REML) to obtain estimates of the variance (Datta and Lahiri 2000).

One challenge with estimating the variance is the issue of non-positive estimates. Estimates of σ^2 obtained using the PR approach may be negative, while estimates obtained using ML or REML can be zero, and the FH estimating equation may not have a positive solution. In these cases, the variance estimates must be adjusted to be positive. Thus, we define $\hat{\sigma}^2 = \max(0.1 * \min_i(D_i), \tilde{\sigma}^2)$, where $\tilde{\sigma}^2$ is obtained using one of the estimation methods discussed above.

We denote the corresponding empirical best linear unbiased predictor (EBLUP) of θ_i based on the estimate of σ^2 as $\hat{\theta}_{i,EBL} = \tilde{\theta}_i(\hat{\sigma}^2, y)$. For the purpose of computing the MSE, we decompose the prediction error $\hat{\theta}_{i,EBL} - \theta_i$ as

$$\hat{\theta}_{i,EBL} - \theta_i = \{\theta_{i,BP}(\beta, \sigma^2, Y) - \theta_i\} + \{\tilde{\theta}_i(\sigma^2) - \theta_{i,BP}\} + \{\hat{\theta}_{i,EBL} - \tilde{\theta}_i(\sigma^2)\},$$

where $\theta_{i,BP} = Y_i - B_i(Y_i - x_i^T\beta)$.

Using this decomposition, and assuming the normal distribution for the random effects, the MSE of the EBLUP is

$$\begin{aligned} MSE(\hat{\theta}_{i,EBL}) &= E\{\theta_i(\beta, \sigma^2, Y) - \theta_i\}^2 + E\{\tilde{\theta}_i(\sigma^2) - \theta_i\}^2 + E\{\hat{\theta}_{i,EBL} - \tilde{\theta}_i(\sigma^2)\}^2 \\ &= g_{1i}(\sigma^2) + g_{2i}(\sigma^2) + E\{\hat{\theta}_{i,EBL} - \tilde{\theta}_i(\sigma^2)\}^2, \end{aligned}$$

where $g_{1i} = D_i(1 - B_i)$ and $g_{2i} = B_i^2 x_i^T\{X^T\Sigma^{-1}X\}^{-1}x_i$.

In general, there is no simple expression of the third term of this MSE decomposition. Much work has been done under the assumption of normality of random effects to produce approximations of the third term, accurate to $o(m^{-1})$. See for instance Prasad and Rao (1990), Datta and Lahiri (2000) or Datta, Rao and Smith (2005) and also Battese, Harter and Fuller (1988). However, there is comparatively less work in cases where the normality of the random effects does not hold. It can be shown that this MSE approximation to the order $o(m^{-1})$ depends on σ^2 and the fourth moment of the random effect v_i. We introduce a bootstrap approach to estimating MSE. Section 2 discusses the botstrap distribution, as well as the bootstrap estimation of the MSE. In Section 3 we provide the results of a set of simulation studies to compare the bias of various estimators of MSE, including the bootstrapping approach. Section 4 discusses a small data example, and Section 5 includes some concluding remarks.

2. Bootstrap MSE

In the absence of normality of the random effects, bootstrap offers one option for estimating the MSE. The use of bootstrapping in general for producing estimates of MSE is not new. For example, Pfeffermann and Correa (2012) proposed a parametric bootstrapping approach for estimating MSE, while others such as Hall and Maiti (2005) suggested a non-parametric boostrap. In this work, we propose the use of a distribution matching the second and fourth moments of the random effects for use in estimating MSE under non-normality. Hall and Maiti (2005) demonstrated that a three-point distribution is adequate to match a distribution with three moments - the mean, variance and fourth moment. We followed their recommendation in developing a bootstrap distribution. Initially, we pursued the use of a three-point distribution to match these moments; however, preliminary simulations revealed two issues. One was that the distribution was not always suitable for obtaining MSE estimates with low bias due to the coarseness of the distribution. Secondly, in some cases the estimates of the kurtosis were extremely large, due to the bounding from the skewness as discussed in Section 2.2. Thus, we incorporated a continuous modification to the bootstrap distribution to smooth the random effect distribution, as well as piecewise function based on the estimate of the kurtosis.

2.1 Bootstrap Distribution

For the model error random effect v_i, we want to draw our bootstrap samples from a distribution matching the second and fourth moments calculated from the sample. We define

$$d = \begin{cases} \frac{1}{3} & 1 < \hat{\gamma}_2 < 3 \\ \frac{1}{12} & \hat{\gamma}_2 \geq 3 \end{cases}, \tag{3}$$

where $\hat{\gamma}_2$ is the estimate of the kurtosis of v_i. Estimation of γ_2 is discussed in Section 2.2. We use d to calculate

$$c = \max\Big(\frac{2d(3-\hat{\gamma}_2) \pm \sqrt{2(\hat{\gamma}_2 - 3)d(1-6d)}}{4(3-\hat{\gamma}_2)d^2 + 2d(1-6d)}\Big), \tag{4}$$

which will be positive. Next set $\alpha = \frac{1}{\sqrt{1+2dc}}$.

We will draw F bootstrap samples. For the f^{th} bootstrap sample, draw $U_{i,f} \sim Uniform(0,1)$ and let

$$W_{i,f} = \sqrt{c}[I(U_{i,f} < d) - I(U_{i,f} > 1-d)]. \tag{5}$$

Also, draw $z_{i,f} \sim N(0,1)$. Then, the bootstrap random effect, denoted by $v^*_{i,f}$ is given by

$$v^*_{i,f} = \alpha(z_{i,f} + W_{i,f})\hat{\sigma}, \tag{6}$$

for $i = 1, ..., m$. Based on these bootstrapped random effects, we have

$$\theta^*_{i,f} = x_i^T\hat{\beta} + v^*_{i,f}. \tag{7}$$

Next, draw the bootstrap sampling errors $e^*_{i,f} \sim N(0, D_i)$ and define our bootstrap direct estimate $y^*_{i,f}$ as

$$y^*_{i,f} = \theta^*_{i,f} + e^*_{i,f}. \tag{8}$$

Based on the bootstrap sample $\mathbf{y}^*_f = (y^*_{1,f}, ..., y^*_{m,f})^T$, we refit the model using the bootstrapped dataset in the same way as the original dataset and obtain $\hat{\sigma}^{2*}_f$ using any chosen estimator and $g_{1i}(\hat{\sigma}^{2*}_f)$. The EBLUP of $\theta^*_{i,f}$ is computed as $\hat{\theta}^*_{i,f}$ for $i = 1, ..., m$.

Note that:

$$E[g_{1i}(\hat{\sigma}^2)] = g_{1i}(\sigma^2) + b_i(\phi) + o(m^{-1}) \tag{9}$$

$$E[g_{1i}(\hat{\sigma}_f^{2*})] = g_{1i}(\sigma^2) + 2b_i(\phi) + o(m^{-1}), \tag{10}$$

where , $\phi = (\sigma^2, \gamma_2)^T$ and $b_i(\phi)$ is of order $O(m^{-1})$ which is the result of the bias and variance in estimating ϕ. Let

$$mse^*_{F,i} = \frac{1}{F}\sum_{f=1}^{F}(\hat{\theta}^*_{i,EBLUP,f} - \theta^*_{i,f})^2.$$

Then,

$$\begin{aligned} E[mse^*_{F,i}] &= E\Big[\frac{1}{F}\sum_{f=1}^{F}(\hat{\theta}^*_{i,EBLUP,f} - \theta^*_{i,f})^2\Big] \\ &= g_{1i}(\sigma^2) + b_i(\phi) + g_i(\phi) + o(m^{-1}), \end{aligned} \tag{11}$$

and $g_i(\phi)$ is of order $O(m^{-1})$.

We note that

$$E\Big[g_{1i}(\hat{\sigma}^2) - \frac{1}{F}\sum_{f=1}^{F} g_{1i}(\hat{\sigma}_f^{2*})\Big] = -b_i(\phi) + o(m^{-1}) \tag{12}$$

Using equation (11) and (12), we define the bootstrap estimator of MSE is

$$\begin{aligned} mse_{i,bootstrap} &= \frac{1}{F}\sum_{f=1}^{F}(\hat{\theta}^*_{i,EBLUP,f} - \theta^*_{i,f})^2 + \Big[g_{1i}(\hat{\sigma}^2) - \frac{1}{F}\sum_{f=1}^{F} g_{1i}(\hat{\sigma}_f^{2*})\Big] \\ &= g_{1i}(\hat{\sigma}^2) + \frac{1}{F}\sum_{f=1}^{F}[(\hat{\theta}^*_{i,EBLUP,f} - \theta^*_{i,f})^2 - g_{1i}(\hat{\sigma}_f^{2*})] \end{aligned}$$

It can be shown that $E[mse_{i,bootstrap}] = MSE(\hat{\theta}_{i,EBL}) + o(m^{-1})$.

2.2 Estimation of the Skewness and Kurtosis

We now discuss estimation of the skewness and kurtosis γ_1 and γ_2. These estimators are used in the bootstrap distribution which matches the second and fourth moment of the random effect, discussed in Section 2.1. Note that

$$E\Big[\sum_{i=1}^{m}\hat{e}_i^3\Big] = \sum_{i=1}^{m}E[v_i^3] + O(1) = m\sigma^3\gamma_1 + O(1),$$

where $\hat{e}_i$ is the residual based on ordinary least squares. Then, our estimator of the skewness coefficient is given by:

$$\hat{\gamma}_1 = (\hat{\sigma}^2)^{-3/2}\frac{1}{m}\sum_{i=1}^{m}\hat{e}_i^3. \tag{13}$$

and based on

$$\begin{aligned} E\Big[\sum_{i=1}^{m}\hat{e}_i^4] &= \sum_{i=1}^{m}[\gamma_2\sigma^4 + 6\sigma^2 D_i + 3D_i^2\Big] + O(1) \\ &= m\gamma_2\hat{\sigma}^4 + 6\sigma^2\sum_{i=1}^{m}D_i + 3\sum_{i=1}^{m}D_i^2 + O(1), \end{aligned}$$

an estimator of the kurtosis coefficient is

$$\tilde{\gamma}_2 = \frac{\sum_{i=1}^{m} \hat{e}_i^4 - 6\hat{\sigma}^2 \sum_{i=1}^{m} D_i - 3\sum_{i=1}^{m} D_i^2}{m\hat{\sigma}^4} \tag{14}$$

Note that the estimator of the skewness coefficientt $\hat{\gamma}_1$ may be positive or negative. However, since $\gamma_2 \geq \gamma_1^2 + 1$, we must have the corresponding inequality satisfied by $\hat{\gamma}_1$ and $\tilde{\gamma}_2$ (Pearson 1916). If that inequality is not satisfied by $\tilde{\gamma}_2$ and $\hat{\gamma}_1$, we modify the definition of $\tilde{\gamma}_2$ as:

$$\hat{\gamma}_2 = \max\Big(\tilde{\gamma}_2, (\hat{\gamma}_1^2 + 1)(1 + \frac{1}{\sqrt{m}})\Big). \tag{15}$$

In the above, we used the multiplier $(1 + \frac{1}{\sqrt{m}})$ so that the equality $\hat{\gamma}_2 = \hat{\gamma}_1^2 + 1$ does not hold. We note that for the purposes of our bootstrap distribution, the estimation of the skewness coefficient is used only in the bounding of the kurtosis.

3. Simulation Study

We conducted a simulation study in R to compare the performance of the bootstrapping method of estimating MSE with traditional estimators. This simulation considered the bootstrapping approach using three estimators of σ^2, including the Fay-Herriot, maximum likelihood and residual maximum likelihood, respectively represented as FH(B), ML(B) and REML(B) in the tables and figures. Each setting was investigated with 1,000 bootstrap samples. We provided comparisons to the traditional estimators introduced by PR/LR, FH, as well as using ML and REML. Simulations were set up following a subset of the simulations conducted by Datta, Rao and Smith (2005), as well as under a sampling variance pattern based on the public use 2017 ACS 5-year data (U.S. Census Bureau 2018).

3.1 Datta-Rao-Smith Simulation Setup

Following a subset of the simulation setup utilized in Section 5 of Datta, Rao and Smith (2005), we conducted simulations using two sampling variance patterns for 15 small areas. Pattern (1), corresponding to Pattern (a) in Datta, et al. had five unique, relatively small sampling variance values: 0.7, 0.6, 0.5, 0.4, 0.3. Pattern (2), which corresponds to Pattern (c) in Datta, et al., had relatively dispersed sampling variance values: 4.0, 0.6, 0.5, 0.4, 0.1. We did not consider Pattern (b) as we were investigating more extreme sampling variance patterns. For each of the sampling variance patterns, we set the random effects variance, $\sigma^2 = 1$. We investigated the normal and double exponential distributions for the random effects. We used a covariate with the coefficient $\beta = 1$, with the values of the covariate generated from an exponential distribution with mean 1.

Ten thousand datasets were generated for each of the simulation setups. With 15 small areas and five distinct sampling variance values per pattern, each of the unique sampling variance values was represented in three small areas per dataset. The results of the simulations were grouped by the value of the sampling variances.

The sampling variances under Pattern (a) were relatively small, and not too dispersed. For this sampling variance pattern, we saw that the percentage bias when using the PR, FH, ML and REML are similar to the resluts obtained by Datta, Rao and Smith (2005). Generally, there was little discrepancy among these four approaches, regardless of whether the normal or double exponential distribution were used for the random effects distribution. We also saw that the relative biases for the bootstrapping approaches were in line with the biases observed for the traditional estimators. The results of these simulations are given in Table 1.

Table 1: Percent Bias Results for Pattern (a)
$m = 15$, 1,000 bootstrap samples, 10,000 datasets, $\sigma^2 = 1$

	D_i	PR	FH	ML	REML	FH(B)	ML(B)	REML(B)
Normal	0.7	-1.70	-2.10	-3.50	-2.50	-2.50	-3.30	-2.50
	0.6	-1.60	-2.10	-3.60	-2.60	-2.30	-3.40	-2.80
	0.5	3.00	2.50	1.90	2.20	1.90	1.60	1.90
	0.4	4.40	4.10	7.00	4.10	1.80	4.30	1.80
	0.3	3.20	2.80	4.70	3.20	2.40	3.50	2.40
Double Exp	0.7	-0.20	-1.50	-3.80	-2.60	-2.90	-5.80	-3.90
	0.6	-1.90	-3.10	-4.90	-4.10	-4.60	-7.20	-5.50
	0.5	1.90	0.80	0.50	0.60	-1.10	-3.00	-1.40
	0.4	3.80	2.90	6.40	2.90	-1.50	0.30	-1.50
	0.3	4.50	4.10	7.7	4.60	0.40	0.40	0.40

Table 2: Percent Bias Results for Pattern (c)
$m = 15$, 1,000 bootstrap samples, 10,000 datasets, $\sigma^2 = 1$

	D_i	PR	FH	ML	REML	FH(B)	ML(B)	REML(B)
Normal	4.0	-0.30	-3.70	-7.10	-3.00	-0.80	-4.50	1.40
	0.6	52.20	-1.00	-5.70	-2.80	0.00	1.80	3.80
	0.5	62.90	-3.90	-7.80	-5.40	-2.90	-0.30	1.00
	0.4	97.50	0.90	0.60	-0.30	-0.30	5.10	3.70
	0.1	669.90	3.10	4.00	1.00	2.10	8.10	5.20
Double Exp	4.0	0.30	-5.80	-11.90	-7.50	-2.90	-9.60	-3.50
	0.6	63.70	-3.10	-9.90	-6.20	-2.40	-4.10	-0.50
	0.5	84.10	-1.10	-7.40	-4.10	-0.60	-1.30	1.90
	0.4	120.20	2.10	0.60	0.30	0.30	3.50	3.60
	0.1	849.00	8.50	10.50	6.50	4.30	9.50	6.50

Sampling variance Pattern (c) had more spread as compared to Pattern (a), and also had a larger maximum sampling variance. The results for the four traditional approaches were comparable to those obtained by Datta, Rao and Smith (2005). For these four approaches, we observed large percentage of bias when using the PR estimator, as well as some negative bias when utilizing the FH, ML or REML. The bootstrapping approaches yielded relative biases that were similar in magnitude to those of the FH, ML and REML variances. The results for Pattern (c) are given in Table 2.

3.2 ACS Simulation Setup

We conducted simulations using sampling variances based on public use ACS 5-year data in order to investigate more extreme situations, with larger sampling variances relative to the sampling variance patterns discussed previously. The sampling variances were generated using the percentiles of ACS 5-year county level data. We considered random effects following the normal distribution, as well as the double exponential distribution with 50 small areas, with $\sigma^2 = 1$. We used a covariate with the coefficient $\beta = 1$, with the values of the covariate generated from an exponential distribution with mean 1. The small areas

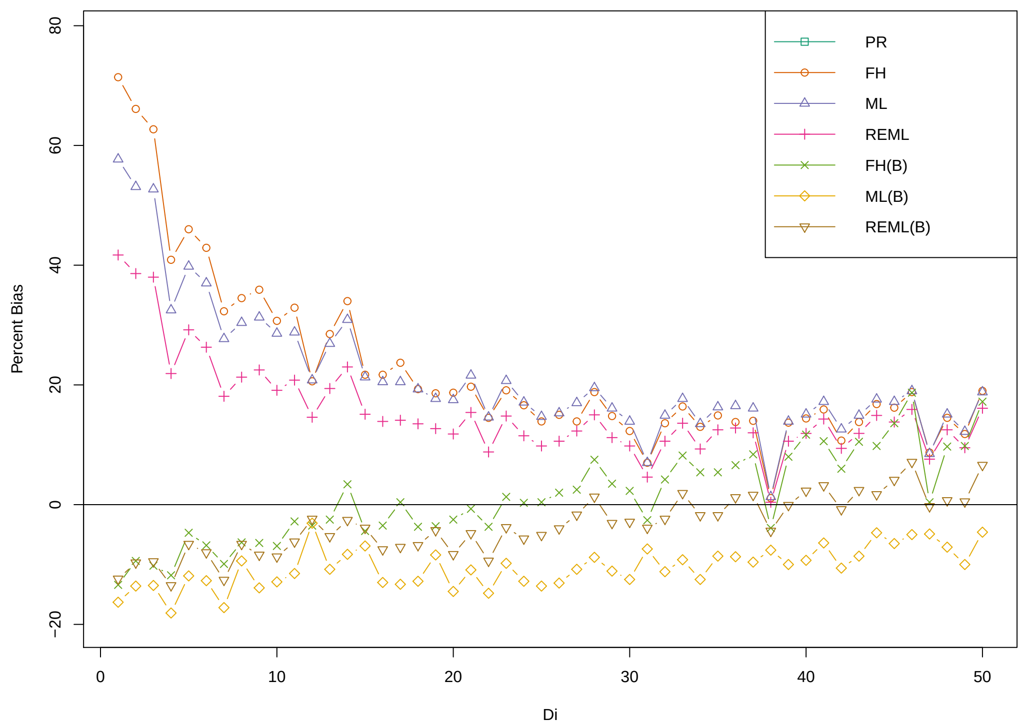

Figure 1: Percent Bias: ACS setup, Normal Distribution
$m = 50$, 1,000 bootstrap samples, 10,000 datasets, $\sigma^2 = 1$

were ordered by increasing sampling variance. Ten thousand simulations were conducted for each random effects distribution, and 1,000 bootstrap samples were used within each simulation.

Figure 1 provides a plot of the percent bias for the four traditional estimators of MSE, as well as the bootstrap using the FH, ML and REML estimators for variance when the random effects are generated following the normal distribution. We observed that the traditional MSE estimators had large relative biases, though this bias leveled off for the larger values of the sampling variance. The PR had extremely large relative bias, and for the most part is not visible in this plot. The bootstrapping approaches performed similarly to one another, with some negative bias for the smaller sampling variances, with this bias generally trending toward zero as the sampling variance increases. In general, the bootstrap approaches had less bias in absolute value as compared to the traditional estimators. The bootstrap with REML appeared to perform best out of these estimators, with respect to the relative bias. The bootstrap with ML estimator of variance appeared to underestimate the true MSE across all areas. A similar conclusion was reached for the double exponential random effects distribution (see Figure 2).

Figure 2 provides a plot of the percent bias for the four traditional estimators of MSE, as well as the bootstrap using the FH, ML and REML estimators for variance when the random effects are generated following the double exponential distribution. These results were largely similar to those under normality. We saw large, positive bias with the PR, FH, ML and REML, while the bootstrapping approaches had less bias in magnitude, though this bias was negative for smaller values of the sampling variances. Again, the bootstrap with REML presented the most appealing performance with respect to the relative bias.

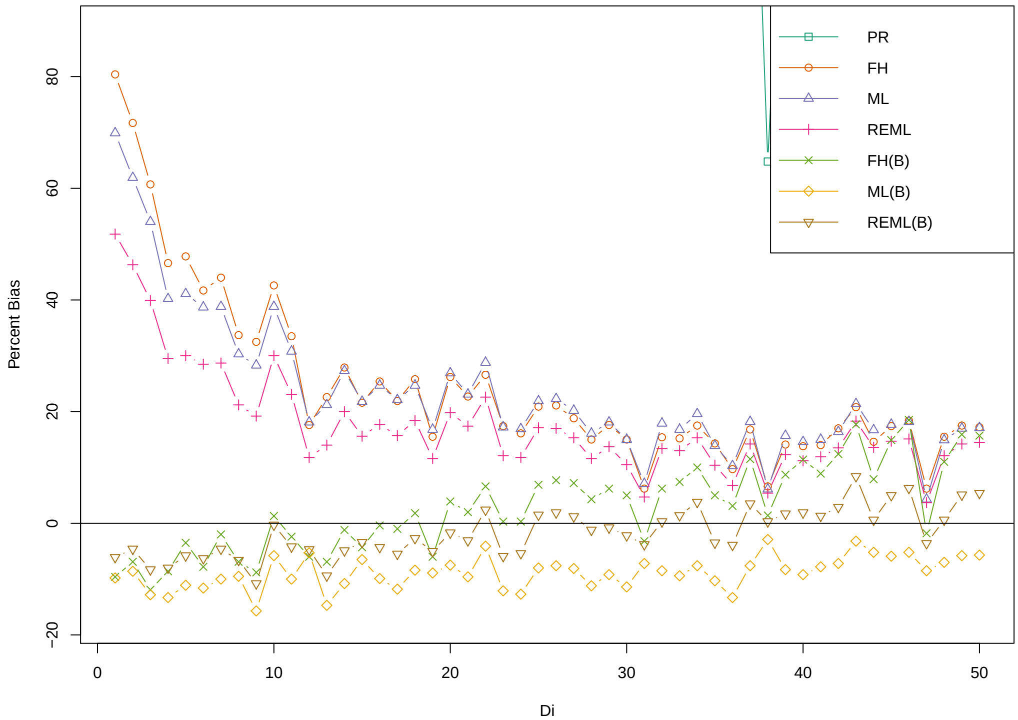

Figure 2: Percent Bias: ACS setup, Double Exponential Distribution
$m = 50$, 1,000 bootstrap samples, 10,000 datasets, $\sigma^2 = 1$

Overall, the simulations following the Datta, Rao and Smith (2005) setup showed that when the sampling variances are small or moderate, as compared to the random effects variance, the bootstrapping approaches perform similarly to the traditional approaches. However, in cases where the sampling variances were large, the traditional approaches yield estimates of MSE that have large relative biases. On the other hand, the bootstrapping approaches were comparatively less biased in the more extreme scenarios. The bootstrapping approaches are thus preferrable for estimating MSE, especially given that the random effects distribution is unknown in practice.

4. Data Analysis

As an illustration of the differences among these approaches, we analyzed modified public use data from the 2017 5-year ACS (U.S. Census Bureau 2018). We considered the poverty rate for children under the age of 18 in the 67 counties in Pennsylvania as our outcome of interest. The rate of Supplemental Nutrition Assistance Program (SNAP) benefit usage for each of the counties was included as a predictor (U.S. Census Bureau 2019).The sampling variances were scaled to be larger for illustrative purposes and the small areas were sorted by increasing sampling variance.

Figure 3 provides a plot of the estimated MSE for each of the four traditional estimators of MSE, as well as the bootstrap with three different variance estimators. The MSE estimates obtained using the Prasad-Rao estimator were very large for counties with small sampling variances, which stood in stark contrast to estimates obtained by any of the other approaches. The ML, REML, bootstrap ML and bootstrap REML had similar performance

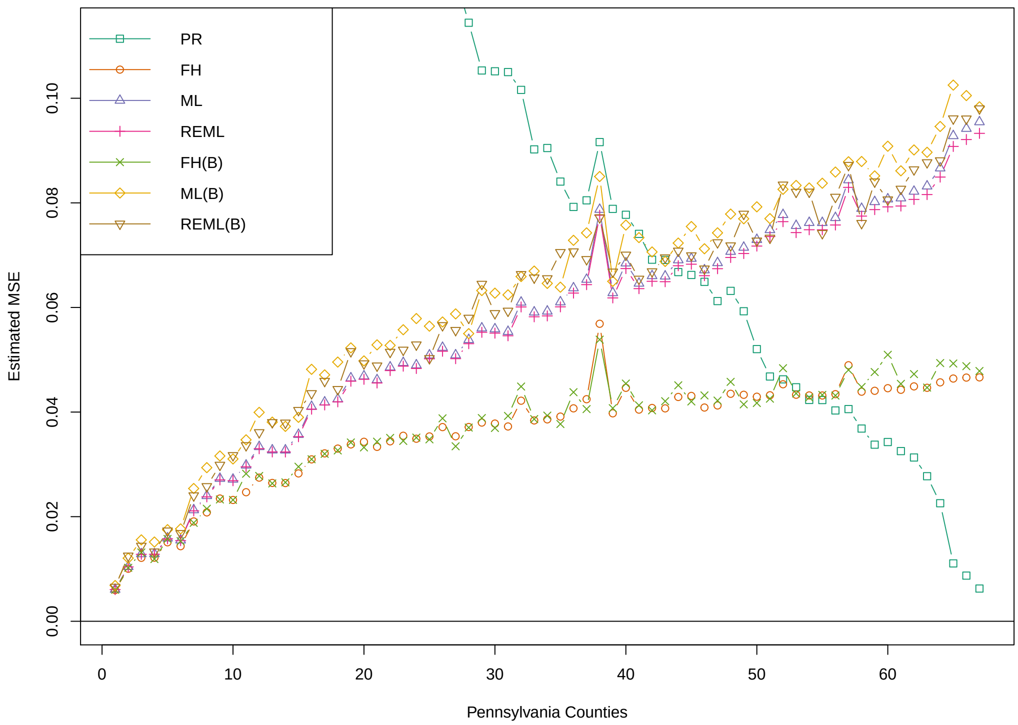

Figure 3: Estimated MSE for ACS Example

to one another. The estimates of MSE for the FH and bootstrap FH were clustered together, with smaller estimates of MSE as compared to other approaches. It is worth noting that though the estimates of MSE differed, the true MSE was unknown.

The Prasad-Rao estimated values of the MSE of the EBLUPs seemed very unreasonable. The estimator $\hat{\sigma}^2_{PR}$ in equation (1) produced a negative estimate in our application. This negative value was truncated upward to $0.1 * \min(D_i) = 0.000636$. This value appeared to be a severe underestimate of the variance (which is supported from a comparison with the other estimates of σ^2; the FH, ML and REML estimates, respectively, are 0.0466, 0.0894, 0.0930). A very low PR estimate of σ^2 reduces the contributions of g_{1i} and g_{2i} in the estimated MSE; the dominant term for all small areas in this application is the g_{3i} term. A simple algebraic calculation will show, for any of the estimation methods of σ^2, that the g_{3i} term will be maximized for a small area i for which $D_i = 2\hat{\sigma}^2$ (see Rao, Molina 2015). Since the D_i are arranged in an ascending order and $D_1 > 2 * \hat{\sigma}^2_{PR}$, the corresponding g_{3i} will be maximized at $i = 1$. The domination of the g_{3i} term and the monotonic decrease (with i) of this term badly affect the PR MSE estimates, resulting in unreliable estimated MSE for all the small areas.

5. Conclusions

Though point estimates are usually the primary focus in small area modeling, estimation of the MSE is also an important consideration. In this work, we investigated a bootstrapping approach that is free of distributional assumptions. We utilized a bootstrapping distribution matching the variance and kurtosis of the random effect. In the case of small, less dispersed sampling variances relative to the model error variance, the relative biases of the MSE estimators using bootstrapping were similar to those produced using more traditional MSE estimators. However, for larger and more dispersed sampling variances, the bootstrap estimators incurred less bias as compared to traditional estimators. Based on the simulations discussed in this work, the bootstrap approach using REML to estimate the variance had the best performance in the more extreme situations, while maintaining lower bias in less extreme settings. Looking forward, additional simulations will be conducted to investigate the performance of the bootstrap MSE under other conditions.

REFERENCES

Battese, G.E., Harter, R.M., and Fuller, W.A. (1988). An Error-Components Model for Prediction of Country Crop Areas Using Survey and Satellite Data. Journal of the American Statistical Association, 83, 28-36.

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

Datta, G., Rao, J.N.K., and Smith, D. (2005). On Measuring the Variability of Small Area Estimators under a Basic Area Level Model. Biometrika, 92(1), 183-196.

Fay, R.E. and Herriot, R.A. (1979). Estimatates of income for small places: an application of James-Stein procedures to census data. J. Amer. Statist. Assoc. 74, 267-277.

Hall, P. and Maiti, T. (2005). Nonparametic estimation of mean squared prediction erorr in nested-error regression models. Annals of Statistics. 34.

Lahiri, P., and Rao, J.N.K. (1995). Robust estimation of mean squared error of small area estimators. J. Amer. Statist. Assoc. 90, 758-766.

Pearson, K. (1916). Mathematical Contributions to the Theory of Evolution. - XIX. Second Supplement to a Memoir on Skew Variation. Philosophical Transactions of the Royal Society of London A, 216(546): 429-457.

Pfeffermann, D., and Correa, S. (2012). Empirical bootstrap bias correction and estimation of prediction mean square error in small area estimation. Biometrika, 99(2), 457-472.

Pfeffermann, D. (2013). New Important Developments in Small Area Estimation. Statistical Science, 28(1), 40-68.

Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. J. Amer. Statist. Assoc. 85 163-171.

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

U.S. Census Bureau (2018). *2014-2017 American Community Survey 5-year Public Use Microdata Samples*. Retrieved from https://www.census.gov/data/developers/data-sets/acs-5year.2017.html

U.S. Census Bureau (2019). *2017 SAIPE Model Input Data*. Retrieved from https://www.census.gov/data/datasets/time-series/demo/saipe/model-tables.html