RESEARCH REPORT SERIES
(Statistics #2020-04)

Comparison of Small Area Models for Estimation of U.S. County Poverty Rates of School Aged Children Using an Artificial Population and a Design-Based Simulation

Carolina Franco[1]

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

Center for Statistical Research & Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: August 24, 2020

Comparison of small area models for estimation of U.S. county poverty rates of school aged children using an artificial population and a design-based simulation*

Carolina Franco, U.S. Census Bureau

August 25, 2020

Abstract

We use the design-based simulation of Maples et al. (2014), which repeatedly samples from an artificial population comprised of 2008-2012 American Community Survey (ACS) unit-level data, to compare different small-area estimation models for county-level rates of school-aged children in poverty in the United States. We compare a Binomial Logit Normal (BLN) model, a Fay-Herriot model on rates, and a Fay-Herriot model on log-transformed counts inspired by the model used in the production of the official estimates by the Census Bureau's Small Area Income and Poverty Estimates (SAIPE) Program. We also explore the effect of estimating the sampling variance on the relative performance of the models, using design-based variance estimates as well as estimates from a Generalized Variance Function (GVF). The GVF of Franco and Bell (2013, 2015) yields a considerable reduction in the Mean Squared Errors (MSEs) of the estimates of the sampling variances of the direct

*The views expressed in this report are those of the authors and not the U.S. Census Bureau. All U.S. Census Bureau disclosure avoidance guidelines have been followed and estimates have been approved for release by the U.S. Census Bureau Disclosure Review Board. DRB approval number: CBDRB-FY19-592 estimators relative to the design-based estimates, but a smaller reduction in the MSEs of the corresponding estimates of effective sample sizes. This seems to lead to an overall reduction of MSEs of the model predictors when using the GVF estimates rather than their design-based counterparts to fit the Fay Herriot model on the rates, but not so for the BLN model. Overall, the BLN model has a very modest advantage over the other two alternatives in the sense of having lower MSEs for the majority of pseudo-counties in the artificial population both when the sampling variances are known or when they are estimated.

KEYWORDS: SAIPE, BLN Model, Fay-Herriot Model, Artificial Population

1 Introduction

The U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program produces estimates of poverty for different age groups and levels of geography in the U.S. Here, we focus on comparing models for the estimation of county poverty rates of school-aged (5-17 years old) children.

The SAIPE program's county-level production model is a Fay-Herriot (1979) model, where the response variables are logged poverty count estimates from the American Community Survey (ACS), based on one year of data collection (called "1-year estimates"). Covariates are available from tabulations of tax records from the Internal Revenue Service (IRS) and from the Supplemental Nutritional Assistance Program (SNAP). One covariate is derived from the 2000 Census. For more information about SAIPE, see https://www.census.gov/programs-surveys/saipe.html.

The Fay-Herriot model is a classic area-level small area estimation model. Small area estimation models typically assume common relationships between the quantities of interest and auxiliary variables across the domains using mixed models to obtain improved predictions relative to the direct survey-weighted estimates. For a review of small area estimation, see Rao and Molina (2015). We test three different small area estimation models for estimating poverty rates for school-aged children in poverty by fitting them to repeated samples from an artificial population comprised of unit-level survey data drawn from the

2008-2012 American Community Survey (ACS).

The ACS 5-year unit-level data is treated as the true population or universe for the purposes of this simulation. The data are grouped into pseudo census tracts and pseudo counties. There are 487 pseudo-counties in this artificial population, which is less than the total number of counties in the US (3143 in 2012). This is a limitation of this simulation, but it still serves as a starting point for model evaluation and comparison. For ease of exposition we usually refer to the pseudo-counties as counties hereafter, except where the distinction is necessary.

The artificial population described above is sampled 1000 times with a simplified design resembling that of the ACS, featuring stratification, systematic sampling, and a simple non-response mechanism. More details on the simulation design can be found in Maples et al. (2014).

The sample sizes vary slightly among simulations because the ACS non-response is mimicked through a simplified random mechanism. The average sample sizes over the simulations for the pseudo-counties range from $239 - 21,890$ households, with a median of 881. In contrast, the 2012 1-year ACS sample for counties ranged from less than 15 to about $56,000$ households with a median of 239. Note that the minimum and median sample sizes are much smaller on average for the simulated samples. This is partly because the simulation was designed with Census tract estimation in mind. This is a drawback of this simulation for the purpose of this evaluation

The models that will be compared include a univariate version of the Binomial Logit Normal (BLN) Model described in Franco and Bell (2013, 2015), a Fay-Herriot model on poverty rates (FHR), and a Fay-Herriot model on log-transformed counts (FHL). As mentioned above, SAIPE uses a FHL model in the production of official estimates of school-aged children in poverty. The motivation for using a BLN model is that it naturally handles skewness and estimates of zero poverty for a county. Note that although the FHL model can also handle skewness, estimates of zero poverty are problematic when attempting to take a log transformation. These estimates are dropped in the model fitting in the SAIPE production model, though (synthetic) model predictions are still formed for those counties.

Slud and Maiti (2011) study the effect of this left-censoring, though using data from the Current Population Survey (CPS), which were used in the production of official SAIPE estimates through 2004.

We also implement the FHR model for comparison due to its simplicity. Though there is no need to drop zeros for fitting the FHR, assuming a symmetric distribution for the estimates of poverty rates may be deemed unrealistic. Moreover, estimates of zero poverty can also be problematic when implementing the FHR model since design-based estimates of the variances of observations of zero are defined as zero, an unrealistic estimate. This problem also arises when computing the design-based effective sample size estimate that is required for the BLN model, which is inversely proportional to sampling variance estimate, and hence undefined for observations of zero poverty. In the simulation, there are very few estimates of zero poverty, namely 29 total in the entire $487,000$ county samples. For the ACS 1-year estimates of county school-aged children in poverty, the incidence of zeros is typically higher, for instance for 2012 about 3.5% of counties had estimates of zero. How we deal with estimates of zero poverty will not have a major impact on the results here, because of the small number of such estimates in the simulation, but could have more of an impact in modeling the real county poverty data, and a larger impact on the related problem of modeling school-aged children in poverty at the Census tract or school-district level, where the incidence of zeros is higher. Nonetheless, we give the details on how we dealt with zeroes when implementing the models in the simulations in Appendix A.

Models similar to the BLN model studied here were explored in the context of SAIPE in Slud (2004), and at the unit level in Slud (2000), and Maiti and Slud (2002) but the analysis in those papers was tailored to CPS data.

The auxiliary variables available for the artificial population include the number of child tax exemptions in poverty and the total number of child tax exemptions for each county. These are derived from tabulations of tax records obtained for statistical purposes under an agreement with the Internal Revenue Service (IRS). The covariate used for the BLN model is the county logit-transformed pseudo IRS poverty rate, which is defined as the number of child exemptions in poverty divided by the total number of tax child exemptions. For the FHR model the covariate used is the untransformed county IRS pseudo poverty rate, and for the FHL model both the log of the county total number of child exemptions and the log of the county number of child exemptions in poverty are used as covariates. Though the FHL model has one more parameter that needs to be estimated, we conjecture that parameter estimation plays a minor role here because there is a large number of small areas which should provide fairly accurate parameter estimates. Note than in the SAIPE production model, additional covariates are available, including the logged number of county SNAP benefits recipients, the log of the estimated county population age 0-17, and the log of the Census 2000 county estimate of the number of related children in poverty. For more information, see https://www.census.gov/programs-surveys/saipe.html.

The analysis is done using the software R (R core team, 2017). For the BLN model we use the function *glmer* in the package *lme4* (Bates et al., 2015), and for the Fay-Herriot model we use the function *mseFH* in the *sae* package (Molina and Marhuenda, 2015). These functions are used to implement an empirical Bayes modeling approach using maximum likelihood for parameter estimation.

In our analysis, we first use the true sampling variances in the model fitting to separate the effect of estimating the sampling variances and effective sample sizes from the problem of selecting the best model form, though clearly the two problems are intertwined. The true sampling variances would not be known in practice, but they can be determined with accuracy from the simulation by averaging over the simulation replications, and using the true population proportions computed from the artificial population. We also compare the models fitted using estimates of the sampling variances and effective sample sizes, including a design-based estimate, and an estimate from the Generalized Variance Function (GVF, see Wolters, 1985) explored in Franco and Bell (2013, 2015).

Section 2 describes the three models. Section 3 discusses simulation results comparing these models using the true sampling variances and effective sample sizes in the fitting. Section 4 discusses the results using design-based estimates of sampling variances and effective sample sizes. Section 5 discusses implementing the GVF in Franco and Bell (2013) in the artificial population, and explores the reductions in MSEs in the estimates of sampling variance and effective sample size that result from using the GVF vs. the design-based estimates. Section 6 gives results comparing the alternative models using these GVF variance estimates. Section 7 provides some concluding remarks.

2 Models

2.1 Fay-Herriot Models

Suppose Y_i is the population characteristic of interest for area i, and y_i is the direct survey estimate of Y_i. Let e_i be the sampling error in y_i, assumed to be $N(0, v_i)$, independent with v_i known. Let u_i be the area i random effect, assumed to be *i.i.d.* $N(0, \sigma_u^2)$ and independent of the e_i. For m small areas, the Fay-Herriot model is

$$y_i = Y_i + e_i \qquad i = 1, \ldots, m \tag{1}$$

$$Y_i = \mathbf{x}_i'\beta + u_i \tag{2}$$

For the FHR model in our simulation, y_i is the estimate of the rate of school-aged children in poverty for county i, and Y_i is the true poverty rate in the artificial population. Then v_i is the sampling variance in the direct estimates of poverty rate. In this study we compare the effect on the model predictor's MSEs of using the true sampling variances, the direct estimates of sampling variances, or the GVF estimates of sampling variances in alternative implementations.

For the FHL model, y_i is the estimate of log-transformed counts of school-aged children in poverty, v_i is its sampling variance, and Y_i is the true log-transformed counts of poverty rates in the artificial population. The results are transformed back to the original scale using the properties of the log-normal distribution. See Bell et al. (2016) for more details.

The covariates x_i used in each of the models were already discussed in Section 1. The models are fit to each of the 1000 samples drawn from the artificial population, predictions are formed, and their MSEs are computed taking averages over the simulation replications to compute expectations over the design, and using the true population values of the parameters of interest from the artificial population.

2.2 BLN model

Before defining the BLN model, we discuss briefly how to modify the sample count and sample size in order to account for the complex sampling design. The idea is that these quantities are adjusted using the design-effect (Kish, 1965), which approximately captures the effect of the complex sampling on the direct estimates. See also Franco and Bell (2013) for discussion of this implementation of the BLN model.

Suppose $\hat{p}_i$ are the direct ACS estimates of poverty rates; $\tilde{p}_i$ are preliminary estimates of p_i based on $\hat{p}_i$ defined such that they cannot be zero. These preliminary estimates are obtained from the direct estimates and the available covariates through a non-linear regression as in Franco and Bell (2013).

Suppose that $\widehat{\mathrm{Var}}(\hat{p}_i)$ is the estimate of sampling variance of $\hat{p}_i$, in this simulation either design-based or GVF-based. We define the (estimated) *effective sample sizes* $\tilde{n}_i$ and (estimated) *effective sample counts* $\tilde{y}_i$ as:

$$\tilde{n}_i = \tilde{p}_i(1-\tilde{p}_i) \Big/ \widehat{\mathrm{Var}}(\hat{p}_i), \tag{3}$$

$$\tilde{y}_i = \tilde{n}_i \times \hat{p}_i \tag{4}$$

Note that the non-zero preliminary estimates of p_i are used in order to avoid effective sample sizes of zero, which do not make sense here. The effective sample size is an estimate of the sample size one would need under Simple Random Sampling to obtain the same variance as in the complex sampling scheme.

The true effective sample size n_i^*, which can be computed from the simulations but would not be known in practice, replaces $\tilde{p}_i$ and $\widehat{\mathrm{Var}}(\hat{p}_i)$ with their true quantities in the right side of (3).

The BLN model using the estimated effective sample sizes is defined as follows:

$$\tilde{y}_i | p_i, \tilde{n}_i \sim \text{Bin}(\tilde{n}_i, p_i) \qquad i = 1, \ldots, m \tag{5}$$

$$\text{logit}(p_i) = \mathbf{x}_i' \beta + u_i; \tag{6}$$

where $u_i \overset{i.i.d.}{\sim} N(0, \sigma_u^2)$, and $\text{logit}(p_i) = \log[p_i/(1-p_i)]$. The covariate $\mathbf{x}_i$ is the logit transformed pseudo poverty rate for county i.

Of course, one could fit the model with the true effective sample size n_i^* if it were known, replacing also $\tilde{y}_i$ with $y_i^* = n_i^* \times \hat{p}_i$. We discuss this case in the next section.

3 Model comparison using true sampling variances and effective sample sizes

In this section, when fitting the models we use the true sampling variances and effective sample sizes for each of the counties. Many small-area models, including the ones studied here, have an underlying assumption that either the sampling variance or the effective sample size is known. In practice these are not known and need to be estimated. We can compute the "true" quantities here since we have access to the universe and true population quantities and we can find accurate expected values by averaging over the 1000 simulation replications.

The empirical Prediction Mean Squared Errors of the predictors of the small area means Y_i are calculated for each county over the 1000 simulation replications. We refer to those simply as the MSE's.

The BLN model predictors have lower MSE's than those of the FHL model for about 68% of the counties, and lower MSE's than those of the FHR model for about 54% of the counties. The FHR model predictor has lower MSE's than those of the FHL model for about 53% of the counties. Table 1 summarizes the distributions of the county ratios of MSEs among the different models.

There is some redundancy in including the summary statistics for both a ratio and its inverse, as we do in many of the tables in this report, since many of these can be inferred from

Table 1: Summary of distribution of county ratios of MSEs of FHL, FHR, and BLN models computed over the simulation replications, where the models are fitted using the true sampling variances and effective sample sizes

MSE Ratio	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
FHL/BLN	0.19	0.94	1.12	1.21	1.34	4.94
BLN/FHL	0.20	0.75	0.89	1.02	1.07	5.39
FHR/BLN	0.24	0.82	1.04	1.36	1.36	8.92
BLN/FHR	0.11	0.73	0.96	1.02	1.22	4.13
FHR/FHL	0.07	0.65	0.95	1.44	1.41	19.85
FHL/FHR	0.05	0.70	1.05	1.29	1.53	13.82

each other (for instance, one can obtain the first quartile of the inverse by inverting the third quartile of the original, the maximum as the inverse of the minimum, etc.). However, we allow this redundancy because it permits easy and quick comparison of the models by looking at both ratios simultaneously.

Note that the median of the FHL/BLN ratio is 1.12 and that the first and third quartiles are larger for this ratio than its reciprocal, while the maximum and minimum are somewhat comparable between the two ratios. These results suggests that the BLN model tends to perform better than the FHL in the sense of having lower MSEs for most counties. The median percentage decrease in MSEs of using the BLN model vs. the FHL is 11% . The median precentage decrease fom using the BLN vs the FHR model is lesser, only 4%. It is unclear which model to select among the FHR and FHL based on Table 1.

We also compare the model MSEs for all three models with the variances of the direct estimators, where these are again calculated over the simulation replications. Although for all three models the majority of counties have MSEs that are lower than the variances of the direct estimators, there are some counties for which this does not hold. There are 48 such counties for the FHL model, 33 for the FHR, and 28 for the BLN model. Table 2 summarizes the distributions of ratios of model MSE's and variances of the direct estimator for the counties. The table suggests a moderate advantage for the BLN model in terms of tending to have lower MSEs. For the majority of counties there is a significant decrease in MSEs from modeling. Note, for instance, that the median decrease in MSE for the BLN model relative to the variance of the direct estimator is about 76%.

Table 2: Summary of county ratios of MSEs of model predictions (FHL, FHR, and BLN) and the variance of the direct estimator computed over the simulation replications, where the models are fitted using the true sampling variances and effective sample sizes

MSE Ratio	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
FHR/DIR	0.01	0.12	0.29	0.42	0.53	3.94
FHL/DIR	0.01	0.11	0.29	0.44	0.65	3.58
BLN/DIR	0.01	0.11	0.24	0.38	0.54	2.95

4 Model comparison using design-based estimates of sampling variances

In this section we perform a similar analysis as in Section 3, but instead of using the true sampling variances and effective sample sizes, which would not be known in practice, we use the corresponding design-based estimates when fitting the models. These variance estimates were computed for each sample from the artificial population using the Successive Differences Replication method (SDR, see Fay and Train, 1995), which is the method used by the ACS (see U.S. Census Bureau, 2014).

The BLN model still has lower MSEs than the FHL model for most counties using the design-based estimates of sampling variances and effective sample sizes, but only for 54% of the counties. The BLN model has lower MSEs than the FHR model for 66% of the counties. And the FHL model has lower MSE's than the FHR model for 58% of the counties.

Using the design-based estimates of sampling variance, the FHL model's performance relative to the other two models seems to improve, compared to the case of known sampling variances. This suggests that in some sense the direct estimates of sampling variances of log count estimates are "better" than the direct estimates of sampling variances of rate estimates. Table 3 provides a comparison of the CVs of these estimates of sampling variances, by providing summaries of the distribution of county ratios of CVs. The table shows a tendency for the CVs to be lower when computing the sampling variance estimates of the log count estimates than when computing the sampling variance estimates of the rate estimates.

Table 3: Summary of distribution of county ratios of CV's of design-based (SDR) estimates of sampling variances for rates and log-transformed counts

CV Ratio	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
SDR rates/SDR log count	0.47	1.08	1.20	1.20	1.34	1.68
SDR log count/SDR rate	0.60	0.75	0.83	0.85	0.93	2.15

We return to comparing the MSEs of the FHR, FHL, and BLN model predictors. Table 4 shows summary statistics of the distribution of county ratios of MSEs among the different models. This table suggests that the BLN model still has a slight edge over the FHL model in terms of tending to yield lower MSEs, since the median of the FHL/BLN ratio is greater than one and the other quartiles are slightly bigger than those of the reciprocal, although the max is considerably larger for the BLN/FHL than its reciprocal. The advantages of using the BLN model over the FHL model appear more tame here than in the case where sampling variances and effective sample sizes are known.

Table 4: Ratios of MSEs of FHL, FHR, and BLN models computed over the simulation replications, where the models are fitted using design-based (SDR) estimates of the sampling variances and effective sample size

MSE Ratio	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
FHL/BLN	0.10	0.77	1.04	1.20	1.42	6.20
BLN/FHL	0.16	0.70	0.96	1.07	1.29	10.43
FHR/BLN	0.33	0.93	1.13	1.66	1.66	18.43
BLN/FHR	0.05	0.60	0.88	0.87	1.07	3.03
FHR/FHL	0.12	0.75	1.13	2.03	1.91	37.39
FHL/FHR	0.03	0.52	0.88	1.10	1.34	8.46

We also compare the model estimates to the direct estimates, by again looking at the distribution of ratios of model MSEs to direct sampling variance estimates, shown in Table 5. The table above suggests that the performance of the BLN and FHL are somewhat comparable when using the design-based estimates of sampling variances, with a slight edge for the BLN model, as suggested by comparing the mean, 3rd quartile, and maximum of these ratios. Note that the minimum, 1st quartile, and median are similar for the FHL and BLN models. We again see significant benefits from using models as opposed to the direct estimators. The number of cases in which the direct estimator has a variance that is lower than MSEs of the model predictors are 51, 34, and 35 for the FHL, FHR, and BLN models respectively.

Table 5: Ratios of MSEs of FHL, FHR, and BLN models and the variance of the direct estimator computed over the simulation replications, where the models are fitted using design-based estimates of sampling variances and effective sample sizes

MSE Ratio	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
FHL/DIR	0.01	0.11	0.27	0.46	0.64	4.28
FHR/DIR	0.04	0.18	0.34	0.45	0.56	3.93
BLN/DIR	0.02	0.12	0.26	0.38	0.53	2.92

5 Estimating the sampling variances via a GVF

In this section, we consider using a Generalized Variance Function (GVF) to improve the estimates of the sampling variances of the direct estimates of poverty rates. Specifically, we use the GVF in Franco and Bell (2013). Franco and Bell give steps for implementation of the GVF. This involves first finding preliminary estimates of the poverty rates that are strictly positive. We use the same least squares regression as Franco and Bell to compute these preliminary estimates, and also try a weighted version, with the weights being proportional to the sample size. As the estimates are very similar under both approaches, we use the unweighted version in subsequent work. Franco and Bell (2013) recommend dropping all counties that have less than 25 households in the sample before fitting the GVF, based on results from Maples (2012). Because the artificial population has larger sample sizes than the actual ACS 1-year estimates, there are no such counties so nothing is dropped.

Table 6 compares the MSE's of the design-based and GVF estimates of the sampling variance by looking at the distribution of their ratios over the counties. For most counties we see a large reduction in the MSEs of the sampling variances estimates from using the GVF rather than the direct estimation approach. In fact, the median reduction is 85%. For all but 42 counties, the MSE of the GVF estimates is less than the MSE of the direct estimates of sampling variances. This suggest that using the GVF estimates when fitting the small area models, where applicable, could lead to improved predictions.

Table 6: Summary measures of counties' ratios of MSE's of design-based and GVF estimates of the sampling variances, where the MSE's are computed over the simulations

MSE ratio	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Direct/GVF	0.24	2.66	6.77	13.92	17.65	112.50
GVF/Direct	0.01	0.06	0.15	0.33	0.38	4.24

There is also a reduction in MSEs of the estimates of the effective sample size when using the GVF estimates of sampling variances to compute them, though the gains do not seem quite as striking. Table 7 summarizes the distribution of the county ratios of MSEs. Note, for instance, that the median decrease in MSE of the effective sample size estimates is 32%, compared to the 85% decrease in the MSEs of the sampling variance estimates in Table 6. There are 135 (about 28%) counties for which the GVF estimates have higher MSEs than the direct estimates of effective sample size.

Table 7: Summary measures of counties' ratios of MSEs of design-based and GVF estimates of the effective sample size, where the MSEs are computed over the simulation replications

MSE ratio	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Direct/GVF	0.19	0.93	1.48	1.84	2.16	20.78
GVF/Direct	0.05	0.46	0.68	0.89	1.08	5.15

6 Model comparisons using GVF estimates of variances

While using the GVF estimates instead of the direct estimates of sampling variances tend to considerably decrease the MSEs of the FHR model predictors, the results are mixed for the

BLN model predictors. While use of the GVF decreases the MSEs of the FHR model predictors for about 67% of the counties, with the median decrease being 18%, for the BLN model only about half the counties show improved MSEs from using the GVF estimates. This could potentially be because the decrease in MSEs of the estimated sampling variances from using the GVF vs the direct estimates tends to be much larger than the decrease in the MSEs of the estimated effective sample sizes, as shown in Section 5. The former is a direct input to the FHR model while the latter is a direct input to the BLN model. Table 8 further compares the MSEs of the model predictors based on GVF estimates to those based on the direct estimates of sampling variances by showing their ratios for the BLN and FHR models separately. From Table 8 one might conclude that the MSE results are actually worse for the BLN model when using the GVF estimates.

Table 8: Summary measures of counties' ratios of MSEs of models based on design-based and based on GVF estimates of the effective sample size, where the MSEs are computed over the simulation replications

MSE ratio	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
BLN Direct/GVF	0.14	0.57	1.00	1.09	1.37	6.48
BLN GVF/Direct	0.15	0.73	1.00	1.35	1.75	7.41
FHR Direct/GVF	0.50	0.88	1.22	1.32	1.62	4.76
FHR GVF/Direct	0.21	0.62	0.82	0.91	1.14	2.00

When directly comparing the MSEs of the model predictors of the BLN and FHR models using the GVF estimates for the fitting, we find that the MSEs of the BLN model are less than those of the FHL model for 53% of the counties, and that the MSEs of the FHR model are less than the MSEs of the BLN model for 60% of the counties. The MSEs of the FHR model predictors are less than the MSEs of the FHL model predictors for 54% of the counties.

We see that using the GVFs for fitting the BLN and FHR models has given an edge to the FHR model, but not to the BLN model, in terms of reducing the MSEs of the model predictors overall. One may then want to compare how the BLN model predictions obtained using the direct sampling variance estimates compare to the FHR model predictions that used the GVF sampling variance estimates. In this case the MSEs of the FHR model predictors are larger than those of the BLN model predictors for 54% of the counties, again suggesting a mild advantage for the BLN model. Table 9 further compares the two models. There, we also compare the FHR MSEs fitted with the GVF estimates to the FHL MSEs fitted using the direct estimates of sampling variances. Again, we see a moderate advantage from using the BLN model, but the winner in terms of reducing MSEs overall between the FHR fitted with GVF estimate of sampling variance and the FHL model using the direct estimates of sampling variance is less clear.

Table 9: Ratios of MSEs of FHL, FHR, and BLN models computed over the simulation replications, where the models are fitted using design-based (SDR) estimates of the sampling variances and effective sample size

MSE Ratio	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
FHR-GVF/BLN-Direct	0.21	0.69	1.07	1.33	1.62	10.08
BLN-Direct/FHR-GVF	0.10	0.62	0.94	1.08	1.46	4.66
FHR-GVF/FHL-direct	0.05	0.71	0.95	1.44	1.48	18.37
FHL-direct/FHR-GVF	0.05	0.68	1.05	1.25	1.40	22.06

In Table 9 we examine the ratios of model MSEs and the variances of the direct estimates, using the GVF estimates of sampling variances in fitting of the FHR model and the direct estimates of sampling variances in fitting the other two models (note some of the entries of this table were already presented in Table 5). This table suggests an advantage of the BLN model over the FHR model, even when the former is fitted with GVF estimates of sampling variances, based on the summary statistics being either lower for the BLN/Direct ratio or very similar. We had already discussed a slight advantage of the BLN model over the FHL model when using the direct sampling variance estimates in Section 4.

7 Conclusions

The simulation shows an advantage from using the BLN model with the design-based estimates of sampling variances compared to the other models that use estimated sampling variances, but the advantage is moderate.

Table 10: Ratios of MSEs of FHL, FHR, and BLN models and the variance of the direct estimator computed over the simulation replications, where the FHL and BLN models are fitted using design-based estimates of sampling variances and effective sample sizes, while the FHR is fitted using the GVF estimates of sampling variances

MSE Ratio	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
FHL-Direct/DIR	0.01	0.11	0.27	0.46	0.64	4.28
FHR-GVF/DIR	0.01	0.13	0.29	0.43	0.53	4.04
BLN-Direct/DIR	0.02	0.12	0.26	0.38	0.53	2.92

The GVF of Franco and Bell (2013, 2015) yields a considerable reduction in the MSEs of the estimates of the sampling variances of the direct estimators, but a smaller reduction in the MSEs of the corresponding estimates of effective sample sizes. This in turn seems to lead to an overall reduction of MSEs of the model predictors when using the GVF estimate to fit the FHR model, but not so for the BLN model. Hence the design-based estimates appear be a better choice to use with the BLN model, whereas the GVF estimates appear to work best for the FHR model. For the FHL model, the GVF estimates are not appropriate since they are designed for proportions, not log-counts.

Overall, the BLN model fit using the direct estimates of effective sample sizes tended to perform slightly better than the FHR model fit using the GVF estimates of sampling variances in the sense of having smaller MSEs for the majority of counties. And it also performed moderately better than the FHL model using the design-based estimates of sampling variance, in the same respect. The results also suggest that sampling variance estimation can have an important impact on model selection, so that the two problems should be considered jointly.

Future research should include testing these models in a simulation design featuring several counties with smaller sample sizes, to more closely mimic the problem of producing official estimates using the true ACS one year estimates of children in poverty. The latter estimates also feature a higher incidence of estimates of zero poverty than our simulation design. The BLN model is better suited to handle estimates of zero than the FHL model, which requires dropping them from the model fitting. However, when using the BLN model with the design-based estimates of sampling variances, alternative estimates of sampling variances should be used for counties with estimates of zero poverty rate, perhaps from a GVF.

Note that a comprehensive model comparison would also look at several model diagnostics tools and external validity checks, but this is not pursued here. Moreover, a bivariate BLN may be more suitable for SAIPE production than its univariate counterpart, so that the previous, non-overlapping ACS 5-year estimates can be jointly modeled with the one year estimates of interest, in order to improve precision. The simulation currently does not support this type of bivariate modeling. A limited study of model diagnostics of the bivariate BLN model for SAIPE data, as well as a comparison between the bivariate BLN and the SAIPE production model is included in Franco and Bell (2015), but more evaluations are needed. This paper is one part of these further evaluations.

More research also needs to be devoted to finding GVFs for estimating sampling variances and design-effects, and in particular, the BLN model may benefit from a GVF that yields lower MSEs for estimation of the effective sample size. Another line of future research is to study how well the reported MSEs track the true MSEs for each of the models, as estimation of MSEs is an important component of small area estimation programs. That is, to compare the MSEs computed under the assumption each model is true to the actual MSEs, which can be ascertained with accuracy from the simulations but usually not in practice.

References

Bates, D. Maechler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1, 1-48

Bell, W. R., Basel, W. W., and Maples, J. J. (2016). An overview of the U.S. Census Bureau's Small Area Income and Poverty Estimates program. Chapter 19 in *Analysis of Poverty Data by Small Area Estimation*, M. Pratesi (ed.), West Sussex, UK: Wiley, 349-378.

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.

Fay, R. E. and Train, G. (1995). Aspects of survey and model-based postcensal estimation of income and poverty characteristics and states and counties. *Proceedings of the Government Statistics Section* , Alexandria, VA: American Statistical Association, 154–159.

Franco, C. and Bell, W. R. (2013). Applying bivariate binomial/logit normal models to small area estimation. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 690–702, URL http://ww2.amstat.org/sections/srms/Proceedings/.

Franco, C. and Bell, W. R. (2015). Borrowing information over time in binomial/logit normal models for small area estimation. *Statistics in Transition (new series)* and *Survey Methodology*, joint issue on Small Area Estimation, **16**, 563–584, URL http://stat.gov.pl/en/sit-en/issues-and-articles-sit/previous-issues/volume-16-number-4-december-2015/.

Maiti, T. and Slud, E. V. (2002). Comparison of Small Area Models in SAIPE. Census Bureau preprint.

Maples, J., (2012) An Examination of the Relative Variance of Replicate Weight Variance Estimators for Ratios Through First-Order Expansions. *Proceedings of the American Statistical Association, Survey Research Methods Section*. 4067-4077, URL http://www.asasrms.org/Proceedings/index.html

Maples, J., Franco, C., Basel, W., Posey, J. and Wieczorek, J. (2014). An Empirical Artificial Population an Sampling Design for Small-Area Model Evaluation. Technical Report, U.S. Census Bureau.

Molina, I. and Marhuenda, Y. (2015) sae: An R Package for Small Area Estimation. *R Journal*, 7, 1, 81-98.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. URL https://www.R-project.org/

Slud, E. V. (2000). Models for Simulation and Comparison of SAIPE Analyses. SAIPE Working Papers. URL https://www.census.gov/programs-surveys/saipe/library/working-papers.2000.html

Slud, E. V. (2004). Small Area Estimation Error in SAIPE using GLM versus FH Models. *Proceedings of the American Statistical Association, Section on Survey Research Methods*. 4403-4409, URL http://www.asasrms.org/Proceedings/index.html

Slud, E. V., and Maiti, T. (2011). Small-area estimation based on survey data from a left-censored Fay-Herriot model. *Journal of Statistical Planning and Inference*, 141, 11, 3520-3535.

U.S. Census Bureau (2014). *American Community Survey Design and Methodology (version 2.0, January 2014)*, URL https://www.census.gov/programs-surveys/acs methodology.html.

Wolter, K. M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.

A Handling of zero estimates

To fit the BLN model, we need estimates of the effective sample counts of children in poverty and of the effective sample sizes. The design-based estimates of effective sample sizes are not well-defined for zero counts. When fitting the BLN model using the design-based estimates of sampling variance, poverty rate estimates of zero are arbitrarily assigned an effective sample size estimate of 1, so that they have minimal impact on the model fitting. This is sensible as observations of zero are associated with small sample sizes. The FHR and FHL models require estimates of the sampling variances of the direct estimates of the poverty rates and the log-transformed poverty counts, respectively. For fitting the FHR model using the design-based estimates of sampling variance, we use an arbitrary high number as the estimate of the sampling variance of an observation of a zero poverty rate, namely 0.25, so such observations have a small impact on the model fitting. Using the design-based estimate of zero would give undue influence and weight to such observations. For the FHL model, we deal with observations of zero poverty rates by assigning them a very high sampling variance estimate (10^8), effectively dropping them from the fitting and giving zero weight to the direct estimates for the counties with estimates of zero poverty.

The "fix-ups" described above for dealing with observations of zero poverty rates are not needed when using the GVF estimates of sampling variances and effective sample sizes for the BLN and FHR models.