Outlier Research in the Annual Survey of Local Government Finances

Peter Schilling
Peter.Schilling@census.gov
Redouane Betrouni
Redouane.Betrouni@census.gov
Bac Tran
Bac.Tran@census.gov
U.S. Census Bureau, 4600 Silver Hill Road, Washington DC 20233

Abstract

The Annual Survey of Local Government Finances (ALFIN) is conducted by the U.S. Census Bureau and provides statistics about the financial activities of state and local governments across the nation. The Economic Directorate makes thousands of estimates at the state and local levels based on ALFIN, and uses small area methods due to low estimation cell sizes. The presence of outliers in ALFIN data is a concern due to violation of model-based assumptions. In this paper, we evaluate the use of transformations in the small area mixed models to handle outliers. Our research uses a Monte Carlo simulation experiment with data from census years 2007 and 2012 to conduct the evaluation.

Key words: robust estimation; small area; mixture models; outliers

1. Introduction

The Economic Directorate of the U.S. Census Bureau is responsible for conducting a census every five years of approximately 91,500 local government units to collect data on their financial activities. During the interim between two consecutive censuses (years ending with 2 and 7, e.g. 2007, 2012, and 2017) the Economic Directorate also directs the Annual Survey of Local Government Finances (ALFIN), a nationwide sample survey covering all local governments in the United States. Estimates published from the ALFIN are aggregated from the five local government types: counties, municipalities, townships, special districts, and school districts, in conjunction with data collected from the Annual Survey of School Finances. The Economic Directorate publishes local level aggregates from the ALFIN along with corresponding state level aggregates from the Annual Survey of State Government Finances. Statistics from these two surveys are used to estimate the government component of the Gross Domestic Product, allocate some federal grant funds, and provide information to assist in public policy research. More information about the ALFIN can be found at: http://www.census.gov/govs/local.

Disclaimer: Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

In non-census years, the Economic Directorate publishes hundreds of statistics for each state. The scale of ALFIN estimates, along with small estimation cell sizes, pose great challenges for traditional design-based estimators, such as Horvitz Thompson, which can become unstable in such conditions. To meet these challenges, we introduced small area estimation methods for ALFIN (Schilling *et al*, 2016), which can improve estimate reliability through the use of models and auxiliary data.

Though small area methods can improve the reliability of estimates, the presence of outliers can challenge basic assumptions, because they suggest the finite population is not represented by the base model used to make estimates. This issue is particularly relevant to ALFIN, where some portions of its complex data structure can experience substantial volatility from one year to another. Because outliers can be highly influential and skew estimates, they can prompt difficult decisions. Though removal of outliers could lead to improvements in some estimates, it is also a drastic step, because outliers can contain important information that should not be discarded. Alternatively, bias correction terms can be applied to the base model of small area estimators, as done by Chambers and Tzavidis (2006), and Chambers *et al* (2009).

In their ground breaking work, McLachlan and Basford (1988) described the effectiveness of mixture models, or the use of multiple distributions, as a means to expose any grouping that may underlie experimental data. This approach is promising in the context of outlier research, where outlying observations are assumed to be of a different nature than the other observations. Later, McLachlan *et al* (1997) explained how mixture models with multivariate normal distributions can be fitted to data through maximum likelihood via the expectation-maximization (EM) algorithm.

Our approach for this research involves a modified form of the linear mixed model often used in small area estimation. As shown by Gershunkaya (2010) and then Trinh and Tran (2016), this approach assumes the underlying distribution is a mixture of two normal distributions, where outliers are assigned to the distribution with higher variance and “regular” observations are assigned to the other. Thus, the modified model can explain the activity of *all* unit observations, and produce estimates that account for outliers.

2. Data

ALFIN data are collected from local governments across the country. Financial activities reported by local governments are assigned to item codes, which can be grouped into four main categories: revenues, expenditures, assets and debts. Approximately four hundred ALFIN item codes are included in these four categories. Small area methods are currently used in the production environment to estimate statistics for expenditures and revenues only. For all other item codes, the Horvitz Thompson estimator is used.

The ALFIN consists of a sample of local governments along with school district data provided by the Annual Survey of State Government Finances. Annual statistics from the ALFIN data are published in two products: the downloadable file and viewable file. The downloadable file includes estimates of the total for each item code, in all the states and for the nation at three different levels: local governments, state governments and combined state and local governments. In contrast, the viewable file provides aggregates of item code totals for the four main categories, as well as totals for notable detailed items. Statistics from the viewable file are given for all the states and for the nation and published online in a nested table format.

With the large number of item code estimates for each state in the downloadable file, the scale of ALFIN estimates can be challenging for any estimation strategy. During non-census years, over 30,000 state-item code totals must be estimated for the annual downloadable file. The cell sizes are based on the number of local governments that contribute to the state item code estimates are often small ($n<10$), which can lead to instability in design-based estimators such as Horvitz-Thompson. But small area methods offer a more robust approach, as effective sample sizes can be increased by “borrowing strength” from similar domains using models and auxiliary data.

Small area estimators are well suited to meet the challenges posed by ALFIN data. Auxiliary data from the CoG-F can be leveraged through models and small area methods to improve ALFIN estimates. Small area methods are especially applicable to ALFIN estimation because the cell sample sizes by domain (state by item code pairs) cannot be controlled and are often too small for reliable direct estimation. The sampled units in the ALFIN sample design are local governments, which can have different combinations of item codes. Thus, the ALFIN sample design is not a direct-element design, and can produce small cell sample sizes for estimation. In addition, the item codes associated with a local government can vary over time, and obscure item codes can be associated with small local governments, which can have low selection probabilities.

3. Sample Design

The ALFIN employs a two-phase sample design. The first phase designates a group of local governments as certainties (weight=1) and includes them in the sample, while selecting the remaining local governments using a stratified probability proportional-to-size (πPS) design (Särndal *et al*, 1992). The second phase reduces the number of non-contributory municipalities, townships, and special districts in the sample using a modified form of cutoff sampling (Dalenius & Hodges, 1959). Data from the 2012 CoG-F provides the auxiliary information for the size variable and helps identify certainty units on the frame. This sample design was originally implemented in 2014 and allows the Economic Directorate to reduce sample size and respondent burden for small cities, townships and special district governments, while maintaining estimate precision and data quality.

The sample design for this research was implemented using a multi-step process. First, large governments were designated as initial certainty units. Next, strata were defined for the remaining units through a combination of state and government types. Four of the five local government types were sampled by this design: counties, municipalities, townships and special districts. Next, in the first stage of the design, a stratified πPS sample was selected, where the size variable was defined as the maximum of total expenditures and a second variable that could be total taxes, total revenues, or long-term debt, depending on the government type. Then, a cut-off point was calculated for the second stage of the design using the cumulative square root of the frequency method (Dalenius & Hodges, 1959), to distinguish between small and large government units in the municipal and special district strata. Finally, the strata with small-size government units were subsampled. For municipal strata, subsampling was carried out using a simple random sampling design; for special district strata, subsampling was accomplished through systematic sampling.

4. Estimation Methods

4.1 Direct Estimator (Horvitz Thompson)

A traditional design-based Horvitz-Thompson (HT) estimator for ALFIN data that estimates the total for area m is:

$$\hat{Y}_m = \sum_{i \in s_m} w_i y_i \tag{1}$$

Where the sampling weight $w_i = \frac{1}{\pi_i}$, and π_i is the selection probability for unit i in area m. Though the HT estimator is unbiased with respect to the sample design, it becomes unstable for small sample sizes, which are common in ALFIN estimates.

4.2 EBLUP Estimator

A model-based estimator for ALFIN data based on small area methods was developed earlier by Schilling *et al* (2016). This model is a form of the Empirical Best Linear Unbiased Predictor, or EBLUP. Consider an estimator of Y_m, the total for small area m:

$$\hat{Y}_m = N_m \left[f_m \overline{y}_m + (1 - f_m) \hat{\overline{Y}}_{mr} \right] \tag{2}$$

Where y_{mj} is the population of the j_{th} unit within the m_{th} area.

$\overline{y}_m = \frac{y_m}{n_m} = \frac{\sum_{j=1}^{N_m} y_{mj}}{n_m}$: the sampled population mean;

$f_m = \frac{n_m}{N_m}$: the sampling rate;

$\hat{\overline{Y}}_{mr}$ is the predictor of the mean of the non-sampled part of the m_{th} area.

Simplifying (1) results in:

$$\hat{Y}_m = y_m + \hat{Y}_{mr} \tag{3}$$

Where $\hat{Y}_{mr}$ is the predictor of the total of the non-sampled part of the m_{th} area.

The predictor $\hat{Y}_{mr}$ is obtained for ALFIN estimator using a linear mixed model:

$$y_{mj} = \boldsymbol{x}_{mj}^{T} \boldsymbol{\beta} + u_m + \varepsilon_{mj} \tag{4}$$

for j = 1 …. n_m, m = 1 … M, and

$$u_m \sim^{iid} N(0, \tau^2) \text{ and } \varepsilon_{mj} \sim^{iid} N(0, \sigma^2)$$

Where $\boldsymbol{x_{mj}}$ is a vector of auxiliary variables for an observation j in area m, $\boldsymbol{\beta}$ is the corresponding vector of regression parameters; u_m are random effects, and ε_{mj} are the errors in the independent observations. To account for the skewed nature of the data and reduce heteroscedasticity, the model in (4) is transformed to a log scale.

$$log(Y_m) = \beta_0 + \beta_1 \, log(X_{mj}) + u_m + \varepsilon_{mj} \tag{5}$$

After fitting the data using (5), the following model-based predictor is used for $\hat{y}_{mj}$:

$$\hat{y}_{mj}^{EBLUP} = \hat{\beta}_0 + \hat{\beta}_1 \, log(X_{mj}) + \hat{u}_m \tag{6}$$

Then, a total estimate for state-item code m can be derived as:

$$\hat{y}_m^{EBLUP} = \hat{Y}_m = \sum_{j \in S_m} y_{mj} + \sum_{j \in S_m^C} \hat{y}_{mj}^{EBLUP} \tag{7}$$

Where $\hat{y}_{mj}^{EBLUP}$ is a model-dependent predictor of the non-sampled part (S_m^C) of the population U_m.

Earlier research (Schilling *et al*, 2016) has shown that small area estimators with linear mixed models can outperform direct estimators for ALFIN data, particularly for state – item code combinations with small sample sizes. However, volatility, or large changes from one year to another, can degrade the performance of linear models in certain estimation cells. This research could lead to the development of modified forms of the model in (4) that can compensate for the effects of outliers. Ultimately, modified small area estimators that can account for outlier effects could lead to more reliable ALFIN estimates.

4.3 N2 Estimator

To account for the adverse effects of outliers, we developed an alternative form of the linear mixed model in (4). As implemented by Gershunkaya (2010), and later by Trinh and Tran (2016), the estimator (denoted N2) is based on a scale mixture of two normal distributions, having a common mean and different variances.

$$y_{mj} = \boldsymbol{x_{mj}^T} \boldsymbol{\beta} + u_m + \varepsilon_{mj} \tag{8}$$

for j = 1 …. n_m, m = 1 … M, and $u_m \sim^{iid} N(0, \tau^2)$,

$\varepsilon_{mj} | z_{mj} \sim^{iid} (1 - z_{mj}) N(0, \sigma_1^2) + z_{mj} N(0, \sigma_2^2)$,

and $z_{mj} | \pi \sim Bin(1; \pi), \sigma_2 > \sigma_1$

where $z_{mj} | \pi$ = “Mixture part” indicator is a random binomial variable, with π = Probability of belonging to part 2 of the mixture

Though the N2 estimator appears to take the same form as the original linear mixed model in (4), its underlying distribution and assumptions are very different. With this approach, every observation is assigned a conditional probability of belonging to part 2 of the mixture, or the "outlier" (higher variance) distribution. Thus, estimates of **β** are robust to outliers, because outlying observations would have a greater probability of being assigned to the higher variance distribution in the mixture (Gershunkaya, 2010).

The following procedure was used for the N2 estimator.

Step 1: Estimating Model Parameters using the EM Algorithm

We used the expectation-maximization (EM) algorithm to estimate the parameters in equation (8) above, with $\theta = (\sigma_1, \sigma_2, \tau, \pi, \beta)$ denoting the set of model parameters. Consider $\theta^{(p)} = (\sigma_1^{(p)}, \sigma_2^{(p)},\ \tau^{(p)},\ \pi^{(p)},\ \beta^{(p)})$ as the set of model parameter values after the p^{th} iteration. Then at the $(p+1)^{th}$ iteration, compute:

E (expectation) - step:

$$z_{mj}^{(p+1)} = \frac{\frac{1-\pi^{(p)}}{\sqrt{\sigma_2^{(p)2} + \tau^{(p)2}}} \exp\left[-\frac{\left(y_{mj} - \boldsymbol{x}_{\boldsymbol{mj}}^{\boldsymbol{T}}\,\boldsymbol{\beta}^{(\boldsymbol{p})}\right)^2}{2\left(\sigma_1^{(p)2} + \tau^{(p)2}\right)}\right]}{\frac{\pi^{(p)}}{\sqrt{\sigma_1^{(p)2} + \tau^{(p)2}}} \exp\left[-\frac{\left(y_{mj} - \boldsymbol{x}_{\boldsymbol{mj}}^{\boldsymbol{T}}\,\boldsymbol{\beta}^{(\boldsymbol{p})}\right)^2}{2\left(\sigma_1^{(p)2} + \tau^{(p)2}\right)}\right] + \frac{1-\pi^{(p)}}{\sqrt{\sigma_2^{(p)2} + \tau^{(p)2}}} \exp\left[-\frac{\left(y_{mj} - \boldsymbol{x}_{\boldsymbol{mj}}^{\boldsymbol{T}}\boldsymbol{\beta}^{(\boldsymbol{p})}\right)^2}{2\left(\sigma_2^{(p)2} + \tau^{(p)2}\right)}\right]}$$

$$w_{mj}^{(p+1)} = \frac{1 - z_{mj}^{(p+1)}}{\sigma_1^{2(p)}} + \frac{z_{mj}^{(p+1)}}{\sigma_2^{2(p)}} \tag{8.1}$$

$$\bar{y}_m^{(p+1)} = \left(\sum_{j=1}^{n_m} w_{mj}^{(p+1)}\, y_{mj}\right) \Big/ \sum_{j=1}^{n_m} w_{mj}^{(p+1)} \tag{8.2}$$

$$\bar{\boldsymbol{x}}_{\boldsymbol{m}}^{(\boldsymbol{p+1})} = \left(\sum_{j=1}^{n_m} w_{mj}^{(p+1)} x_{mj}\right) \Big/ \sum_{j=1}^{n_m} w_{mj}^{(p+1)}$$

$$V_m^{(p+1)} = 1 \Big/ \left(\sum_{j=1}^{n_m} w_{mj}^{(p+1)} + \frac{1}{\tau^{(p)2}}\right)$$

$$u_m^{(p+1)} = V_m^{(p+1)}\ \left(\bar{y}_m^{(p+1)} - \bar{\boldsymbol{x}}_{\boldsymbol{m}}^{\boldsymbol{T}(\boldsymbol{p+1})} \boldsymbol{\beta}^{(\boldsymbol{p})}\right) \sum_{j=1}^{n_m} w_{mj}^{(p+1)} \tag{8.3}$$

M (maximization) - step:

$$\pi^{(p+1)} = \left(\sum_{m=1}^{M} \sum_{j=1}^{n_m} z_{mj}^{(p+1)}\right) \Big/ n \tag{8.4}$$

$$\sigma_1^{(p+1)2} = \frac{\sum_{m=1}^{M} \sum_{j=1}^{n_m} \left(1 - z_{mj}^{(p+1)}\right) \left[\left(y_{mj} - \boldsymbol{x}_{\boldsymbol{mj}}^{\boldsymbol{T}} \boldsymbol{\beta}^{(\boldsymbol{p})} - u_m^{(p+1)}\right)^2 + V_m^{(p+1)}\right]}{\sum_{m=1}^{M} \sum_{j=1}^{n_m} \left(1 - z_{mj}^{(p+1)}\right)}$$

$$\sigma_2^{(p+1)2} = \frac{\sum_{m=1}^{M}\sum_{j=1}^{n_m} z_{mj}^{(p+1)}\left[\left(y_{mj} - \boldsymbol{x_{mj}^T}\boldsymbol{\beta^{(p)}} - u_m^{(p+1)}\right)^2 + V_m^{(p+1)}\right]}{\sum_{m=1}^{M}\sum_{j=1}^{n_m} z_{mj}^{(p+1)}}$$

$$\tau^{(p+1)2} = \sum_{m=1}^{M}\left(u_{mj}^{(p+1)} + V_m^{(p+1)}\right)/M$$

$$\beta^{(p+1)} = \frac{\sum_{m=1}^{M}\sum_{j=1}^{n_m} w_{mj}^{(p+1)}\boldsymbol{x_{mj}}\left(y_{mj} - u_m^{(p+1)}\right)}{\sum_{m=1}^{M}\sum_{j=1}^{n_m} w_{mj}^{(p+1)}\boldsymbol{x_{mj}^T x_{mj}}} \qquad (8.5)$$

As shown in (8.1), when z_{mj} increases, the effects of the unit on (8.2) are down-weighted. In contrast to the original EBLUP estimator, the N2 estimator can offer robustness to outliers, because it can account for and adjust to their effects.

Step 2: After converging to the solution, we can find a predictor ($\boldsymbol{Y_{mr}}$) for the non-sampled data:

$$\boldsymbol{\hat{Y}_{mr}^{N2}} = \boldsymbol{x_{mr}^T}\,\hat{\beta}^{N2} + (N_m - n_m)\hat{u}_m^{N2}$$

Where $\hat{\beta}^{N2}, \hat{u}^{N2}$ are computed from the EM algorithm and $x_{mr}^T = \sum_{j=n_m+1}^{N_m} \hat{x}_{mj}^T$

Step 3: The estimator of Y_m is then:

$$\boldsymbol{\hat{Y}_m^{N2}} = y_m + \boldsymbol{\hat{Y}_{mr}^{N2}}$$

Similar to the original EBLUP estimator, the N2 estimator employs small area methods, because it incorporates information from multiple sources, including both sampled and non-sampled portions. The N2 estimator borrows strength by combining data from the area m, but also the non-sampled portions, which are taken from all areas. Unlike the original EBLUP estimator, the N2 estimator provides a means to adjust for the effects of outliers.

Model Transformations:

Both the original EBLUP estimator and N2 estimator have underlying unit-level, nested error models that assume errors are distributed normally. In practice, these assumptions are rarely satisfied and adjustments are often required before the estimators are used with production data. Issues with the normality assumption can often be addressed with transformations. As discussed by Schilling *et al* (2016), the log transformation helped address normality issues with the original EBLUP estimator. For the N2 estimator, we explored both the log and square root transformations. Though both transformations yielded similar results in terms of RRMSE, the square root transformation converged to solution much faster when estimating EM algorithm parameters, which was consistent with findings from Trinh and Tran (2016). Thus, the square root transformation was selected to address normality issues with the N2 estimator.

The modified unit level model from (8) is:

$$\sqrt{y_{mj}} = \beta_0 + \beta_1\sqrt{x_{mj}} + u_m + \varepsilon_{mj} \tag{9}$$

Then, we can predict y_{mj} with the inverse transformation:

$$\hat{y}_{mj}^{N2} = \left(\hat{\beta}_0^{N2} + \hat{\beta}_1^{N2}\sqrt{x_{mj}} + \hat{u}_m^{N2}\right)^2 \tag{10}$$

Finally, a total estimate for state-item code m can be derived as:

$$\hat{y}_m^{N2} = \hat{Y}_m = \sum_{j\in S_m} y_{mj} + \sum_{j\in S_m^C} \hat{y}_{mj}^{N2} \tag{11}$$

Where $\hat{y}_{mj}^{N2}$ is a model-dependent predictor of the non-sampled part (S_m^C) of the population U_m.

5. Evaluation Design

Our research includes data from the Finance components of the 2007 and 2012 Census of Governments. The universe is created from the intersection of 2007 data with 2012 data, and is restricted to only the units surveyed during both census years. For simplicity, the universe is further limited to include only non-zero values on the variables of interest, or the four main groups of item codes. The universe for this research consists of approximately 85,850 units.

The 2007 CoG-F provides the auxiliary data, and also serves as the sampling frame. The production sampling design is applied to select 1000 replicated samples on the 2007 and 2012 CoG-F data. For each sample replicate we estimate the 2012 state totals for both expenditure and revenue item codes using three estimators: HT, EBLUP, and N2. During the analysis, we computed the relative root mean squared error (RRMSE) for each estimator from the 1000 samples. Our analysis includes item code totals for two of the states with the greatest variety of ALFIN item code estimates: New York and California.

Relative Root Mean Squared Error (RRMSE):

Based on the Mean Squared Error (MSE), which incorporates both the variance of an estimator and its bias, the Relative Root Mean Squared Error (RRMSE) provides an important measure to compare estimator quality. The RRMSE is calculated as:

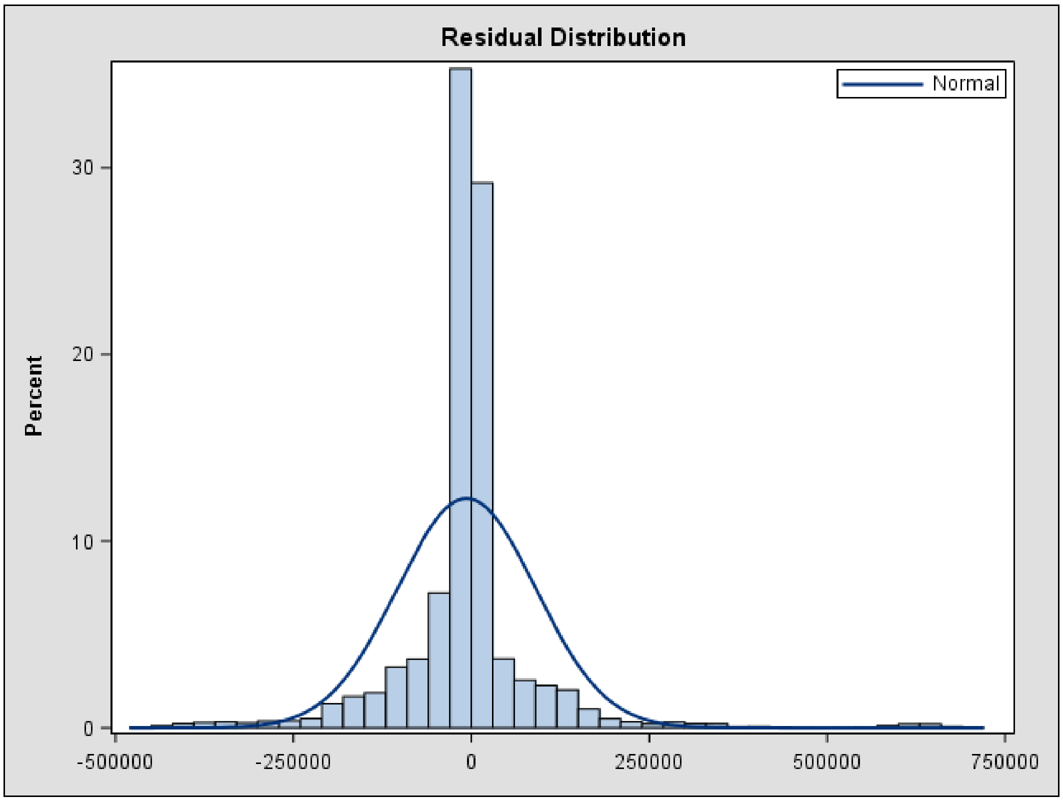
$$\widehat{RRMSE} = \sqrt{\frac{1}{n_{rep}}\sum_{k=1}^{n_{rep}}\left(\frac{\hat{Y}_{m,k} - Y_m}{Y_m}\right)^2} \tag{12}$$

Where n_{rep} is the number of sample replicates, $\hat{Y}_{m,k}$ is the estimated value of small area m for the k_{th} replicate, and Y_m is the true value for small area m, or the state-item code combination.

6. Results

Figure 1 shows a distribution of residuals after applying a square root transformation from equation (7) above for California revenues over the entire set of 1000 replicates. As shown in Figure 1, the plot is both unimodal and symmetric, and the normality assumption in the unit level model for the N2 estimator is approximately satisfied.

Figure 1: Normality of residuals for CA revenues



Data Source: U.S. Census Bureau, 2007 and 2012 Census of Governments: Finance

A comparison of the estimated RRMSE for the three estimators is provided in Table 1. These values denote the number of times one estimator outperforms the others for $\widehat{RRMSE}$. The results in Table 1 are compiled from 1000 replicates, which produce state-item code estimates for two states: California and New York. A total of 330 state-item code estimates are calculated for each replicate.

Table 1: Number of times estimator outperforms others for RRMSE (cells = state by item code estimates)

State	HT	EBLUP	N2
CA	3	51	88
NY	10	42	82

NOTE: Ties are not listed in Table 1; these can be attributed to 54 state - item code estimates that are from cells having only certainty units.
Data Source: U.S. Census Bureau, 2007 and 2012 Census of Governments: Finance

The state-item code estimates from cells having only certainty units are excluded from Table 1. In both states, the N2 estimator outperforms the other two in approximately 60% of the remaining cells.

Table 2 provides a breakdown of the results from Table 1 relative to estimation cell size. The results in Table 2 are aggregated from both states (CA, NY) from Table 1. The Table 2 categories are based on the median cell sizes calculated over the 1000 sample replicates for non-certainty (πPS) units only, or the units that are included in the model for a state-item code estimate. In the last two categories, the median cell size can be zero, because some state-item code combinations are obscure, with only one or two contributing πPS units, which do not appear in every sample replicate. The last two categories of the obscure state-item code estimates indicate that some of the estimates can include contributing certainty units, while others are calculated from πPS units only.

Table 2: Overall estimator performance for RRMSE by cell size (cells = state by item code estimates)

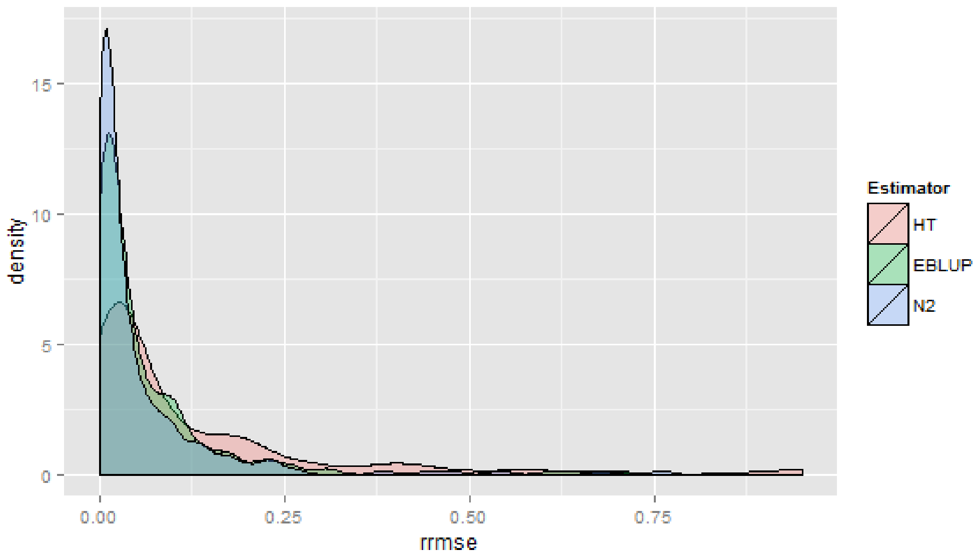
Median Cell Size (πPS units only)	Number of Times Estimator Outperformed		
	HT	EBLUP	N2
> 30	4	24	57
21-30	3	5	20
11-20	2	9	28
6-10	2	13	20
1-5	1	27	33
0*	1	13	5
0**	0	2	7

* Includes other contributing certainty units in the estimates.
** Estimates calculated only from πPS units (no certainty units).
NOTE: Ties are not listed in Table 2; these can be attributed to 54 state - item code estimates that are from cells having only certainty units.
Data Source: U.S. Census Bureau, 2007 and 2012 Census of Governments: Finance

The results from Table 2 show that N2 outperforms the other two estimators in most of the size categories. One of the last two categories, which include obscure state-item code estimation cells, is an exception. In the first obscure state-item code category, which includes certainty units, EBLUP outperforms the other two estimators. In the second obscure state-item code category, which include only πPS units, the N2 estimator outperforms the other two.

A density plot comparing RRMSE values for the three estimators is given in Figure 2, which excludes values greater than one (mostly for the HT estimator) for clearer visualization.

Figure 2: Estimator RRMSE for values<=1
(cells = state by item code estimates)

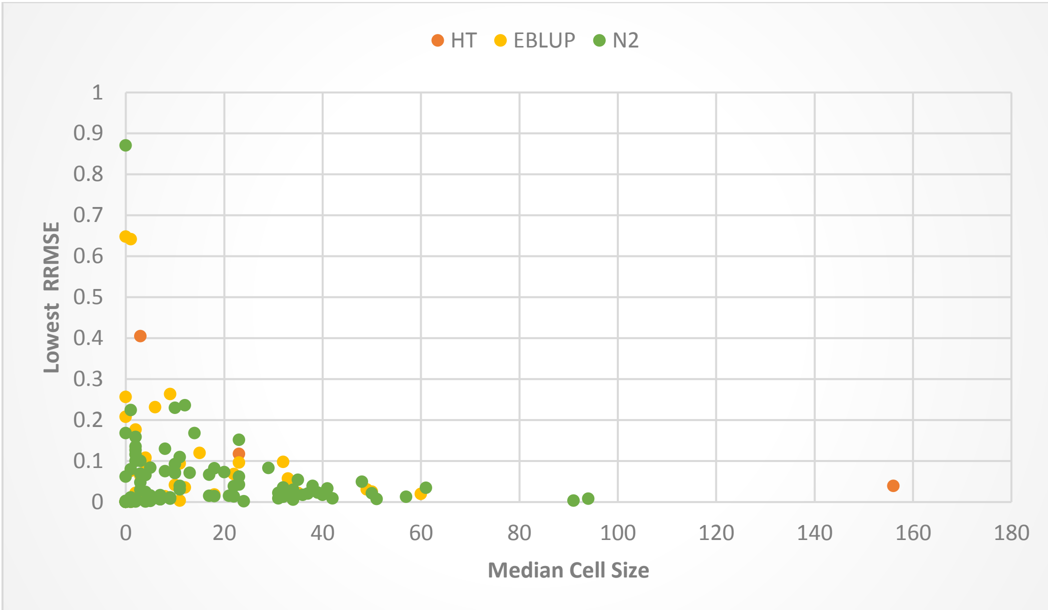


Data Source: U.S. Census Bureau, 2007 and 2012 Census of Governments: Finance

Figure 2 has a sharp peak near zero for the N2 estimator, with another sharp peak by EBLUP and then a shallower peak for the HT estimator. Both the N2 and EBLUP estimators have relatively short tails following their sharp peaks. In contrast, Figure 2 shows a broader, longer tail for the HT estimator. These results demonstrate a similar degree of high performance for $\widehat{RRMSE}$ in the N2 and EBLUP estimators versus lower performance for HT.

To examine the findings from Tables 1 and 2 and Figure 2 more closely, we created two scatter plots of the lowest $\widehat{RRMSE}$ value of the three estimators versus median estimation cell size by state, which are given in Figures 3 and 4 below. Note that the median cell sizes in the scatter plots are for πPS units only. The low $\widehat{RRMSE}$ values for the three estimators are shown using different colors on the two scatter plots.

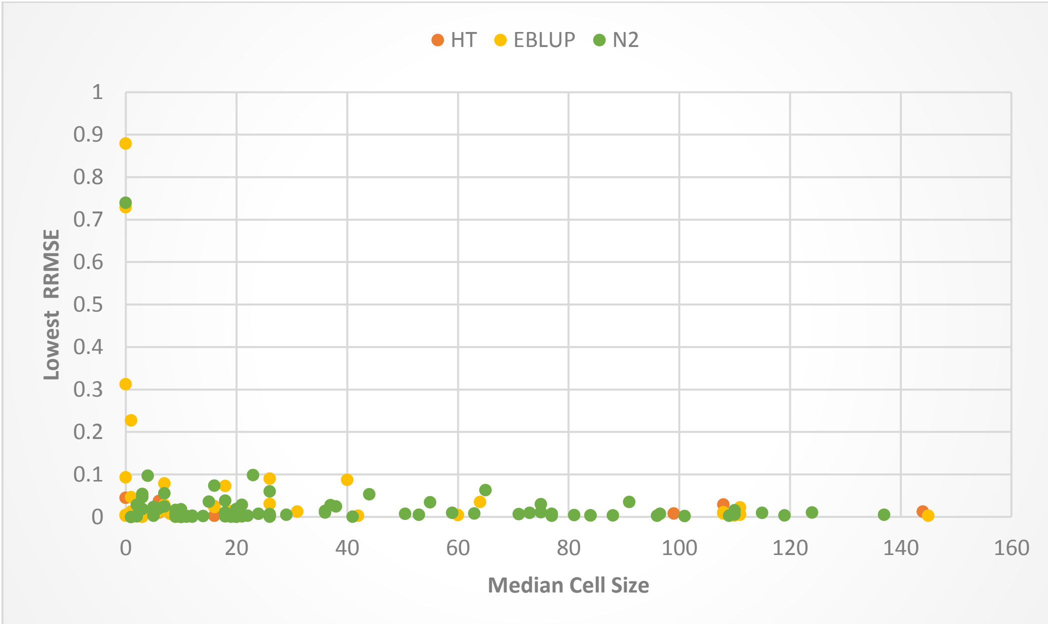
Figure 3: Lowest RRMSE vs. median cell size (CA πPS units only)
(cells = state by item code estimates)



Data Source: U.S. Census Bureau, 2007 and 2012 Census of Governments: Finance

As expected, the data from Figure 3 show higher $\widehat{RRMSE}$ values for smaller median estimation cell sizes and falling $\widehat{RRMSE}$ values as the median cell size increases. The scatter plot also confirms values from Table 1 for California in that most of the data points represent either the EBLUP or N2 estimators, with few data points for HT. Though similar profiles can be observed on Figure 3 for the N2 and EBLUP estimators, the N2 data points tend to have lower values and be clustered closer to zero. In addition, more data points for the N2 estimator on Figure 3 confirms the results from Table 1, showing that it outperforms EBLUP in the number of instances it has the lowest $\widehat{RRMSE}$ value.

Figure 4: Lowest RRMSE vs. median cell size (NY πPS units only) (cells = state by item code estimates)



Data Source: U.S. Census Bureau, 2007 and 2012 Census of Governments: Finance

The data from Figure 4 also show that $\widehat{RRMSE}$ values fall as median estimation cell sizes increase. Based on the number of data points on the scatter plot, the relative performance of the three estimators can be ranked, with the N2 estimator outperforming the other two, followed by EBLUP and then HT. This ranking is consistent with the results in Table 1 for New York. In contrast with Figure 3, there is less overall scatter among the data points, with relatively few $\widehat{RRMSE}$ values above 0.1 and most data points clustered near the zero value.

7. Conclusions

Our evaluation shows that the N2 estimator clearly outperformed the EBLUP and HT estimators with respect to $\widehat{RRMSE}$. Thus, ALFIN estimation gains are possible by adding outlier robustness to small area estimators. In the future, we plan to expand this research to other, less populated states, where the estimation challenges are greater. In addition, we plan to explore whether further improvements to the N2 estimator are possible using overall bias correction (Gershunkaya, 2010), which could potentially correct for biases in certain outlying estimation cells.

8. References

Chambers, R. and Tzavidis N. (2006). "M-quantile Models for Small Area Estimation", *Biometrika* 93, 255-268.

Chambers, R., Chandra, H., Salvatti, N. and Tzavidis N. (2009). "Outlier Robust Small Area Estimation", Invited Presentation, ISI 2009, South Africa.

Dalenius, T. and Hodges, J.L. (1959). “Minimum Variance Stratification”, *Journal of the American Statistical Association,* 54, 88-101.

Gershunskaya, J. (2010). “Robust Small Area Estimation Using a Mixture Model”, 2010 Joint Statistical Meetings.

McLachlan, G.J. and Basford K.E. (1988). *Mixture Models: Inference and Applications to Clustering.* New York; Marcel Dekker.

McLachlan, G.J, Peel, D. and Prado, P. (1997). “Clustering via normal mixture models”, *Proceedings of the American Statistical Association (Bayesian Statistical Science Section)*, 98-103.

Schilling P., Bertrouni R. and Tran B. (2016). “The Performance of the Best Linear Unbiased Predictor in the Annual Survey of Local Government Finances”, 2016 Joint Statistical Meetings.

Trinh, G.T. and Tran B. (2016). “Outliers in the Annual Survey of Public Employment & Payroll: Small Area Estimation Approach”, 2016 Joint Statistical Meetings.