The Performance of the Empirical Best Linear Unbiased Predictor in 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 Statistical Methods Division currently uses a combination of empirical best prediction, Calibration, and Horvitz-Thompson methods to estimate these statistics. These three estimators are evaluated through a Monte Carlo simulation experiment using the two census years data 2007 and 2012. The performance of the three estimators is compared through their mean squared errors and relative bias.

Keywords: Annual Survey of State and Local Government Finances; EBLUP; Calibration

1. Introduction

Every five years, the Economic Directorate of the U.S. Census Bureau conducts a census of approximately 91,500 local government units to collect data on their financial activities. In the years between two consecutive censuses (years ending with 2 and 7, e.g. 2007, 2012, and 2017) the Economic Directorate also conducts 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: This report is released to inform interested parties of research and to encourage discussion. Any views expressed on statistical, methodological, technological, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

We used three estimation methods for the 2013 ALFIN survey cycle. Originally, we planned to use calibration to estimate for both aggregate totals and low level values for most expenditures and revenues data, reverting to Horvitz-Thompson methods for all other data. Though calibration estimates performed well for stable data with relatively small yearly changes, the estimates degraded for volatile data with moderate or extreme changes from one year to another. To improve our methodology and address the shortcomings of calibration, an empirical best linear unbiased prediction (EBLUP) estimator was developed during the 2013 ALFIN survey cycle. EBLUP was used to make estimates for the most challenging, volatile revenues/expenditures data and tended to outperform calibration in these situations.

However, a thorough, rigorous evaluation is needed to compare estimators for the ALFIN and determine what conditions one estimator outperforms the others. For this research, we conducted an evaluation to compare the performance of three estimators: Horvitz-Thompson, calibration and EBLUP. We used data from the 2007 and 2012 Censuses of Governments: Finance (CoG-F) to carry out the evaluation.

2. Data

Data collected for the ALFIN are provided by local governments across the country. Each financial activity reported by local governments is assigned to an item code. These item codes can be grouped into one of four main categories: revenues, expenditures, assets and debts. Approximately four hundred item codes are included in these four categories. In the production environment, we currently select from the three estimators for only the expenditures and revenues item codes. 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. The annual statistics from the ALFIN data are published in two products: the downloadable file and viewable file. The downloadable file provides estimates of the total for each item code, both by state 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 some of the more notable detailed items. Statistics from the viewable file are given both by state and for the nation and published online in a nested table format.

The scale of the ALFIN statistics presents formidable challenges when making estimates. During non-census years, over 30,000 state-item code totals must be estimated for the annual downloadable file. The cell sizes, based on the number of local governments contributing to the state item code estimates, are often small ($n<10$), and design-based estimators such as Horvitz-Thompson can become unstable in these conditions. This research continues earlier efforts by Love *et al* (2013, 2014) to find alternate estimators that improve estimation stability and precision for the ALFIN.

Through small area estimation, we can calculate estimates and measures of variability for areas, or domains, with sample sizes that are too small for reliable direct estimation with traditional estimators such as Horvitz-Thompson. Using small area methods, the effective sample sizes can be increased by "borrowing strength" from similar domains with models and auxiliary data. Though the models can take a variety of forms, the overall goal is an appreciable increase in estimation accuracy over that of the direct estimator.

Small area methods offer promise as an alternative approach to handle the challenges posed by ALFIN estimation. Auxiliary data from the CoG-F can be leveraged through models and small area methods to improve ALFIN estimates. The use of small area estimation for ALFIN is appropriate because cell sample sizes by domain (state by item code pairs) cannot be controlled and are often too small for reliable direct estimation. The small cell sample sizes are the result of a sample design that is not a direct-element design. Instead, the sampled units are local governments, which have different combinations of item codes. 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 uses a two-phase sample design. In the first phase, a group of local governments is designated as certainties (weight=1) and included in the sample, while other local governments are selected using a stratified probability proportional-to-size (πPS) design (Särndal *et al*, 1992). In the second phase, a modified version of cutoff sampling (Dalenius & Hodges, 1959) is used to reduce the number of non-contributory municipalities, townships and special districts in the sample. This sample design was 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. Data from the 2012 CoG-F provides the auxiliary information used for the size variable and to identify certainty units on the frame.

The sample design 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, including: 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. Next, 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)

The traditional design-based Horvitz-Thompson (HT) estimator can be used to estimate the total for an item code c in state k:

$$\hat{t}_{kc} = \sum_{i \in S, i \in k} d_{ki} y_{kic} \tag{1}$$

Where the sampling weight $d_{ki} = \frac{1}{\pi_{ki}}$, and π_{ki} is the inclusion probability for unit i in state k, and units are summed in sample for state k. Note that in this evaluation, one government unit (i) can have multiple item codes (c).

The HT estimator is unbiased with respect to the sample design. But HT is also highly variable for small sample sizes, which occurs often for ALFIN estimates.

4.2 Calibration Estimator

Calibration methods are a form of reweighting, because they consist of adjusting the sample design weights so that survey estimates of totals agree with known population totals. These known population totals are auxiliary data obtained from external sources. Calibration estimators employ auxiliary data to adjust the original sampling weights to be consistent with a set of constraints known as the calibration equations. The calibration estimator is model-assisted because although models are used to construct the estimator, its statistical properties are calculated with respect to the probability sampling distribution (Valliant *et al*, 2000). Through their use of auxiliary data, calibration estimators can offer improvements in accuracy and precision versus direct estimators.

Starting with a finite population $U = \{ 1, \dots, i, \dots N \}$, a probability sample s ($s \subseteq U$) is drawn with a given sample design, and we assume that inclusion probabilities $\pi_i = \Pr(i \in s)$ and the joint inclusion probabilities $\pi_{ij} = \Pr(i \,\&\, j \in s)$ are always positive. Let y_i be the value of the variable of interest y for the i^{th} population element in U.Let x_i be the value of the auxiliary variable x for the i^{th} population element, where x_i can be a vector containing one or many variables.

Suppose $(y_i \;\; x_i)$ is observed for $i \in s$, and the population total $t_x = \sum_U x_i$ is known. Our goal is to find a set of weights (w_i) by adjusting the sample design weights $(d_i = \frac{1}{\pi_i})$ to meet the following constraint:

$$\hat{t}_x = \sum_s w_i x_i \qquad (2)$$

Deville and Särndal (1992) defined a distance function (g) to measure the proximity of the original weights (d_i) with the new weights (w_i). Their objective was to derive a new set of weights (w_i) that minimized deviations with the original sampling weights (d_i), which yield unbiased estimates. Deville and Särndal (1992, 1993) also derived a class of estimators, each corresponding to a different distance function. These estimators, which can satisfy the above constraint in equation (2) using a known population total, are known as calibration estimators.

Deville and Särndal used a series of LaGrangian multipliers to solve for (2). In practice, the new set of calibrated weights (w_i) is derived numerically through computer software. We currently use SUDAAN, developed by Research Triangle Institute, for calibration estimations. SUDAAN can solve for equation (2) and derive calibration weights through repeated linearization (Research Triangle Institute, 2012) by finding a distance function g so that:

$$\sum_{U} \boldsymbol{x_i} = \sum_{S} d_i \, \alpha(\boldsymbol{g}^T \boldsymbol{x_i})(\boldsymbol{x_i}) \tag{3}$$

where

$$\alpha(\boldsymbol{g}^T \boldsymbol{x}_i) = \frac{l(u-c) + u(c-l)\exp(A\boldsymbol{g}^T \boldsymbol{x}_i)}{(u-c) + (c-1)\exp(A\boldsymbol{g}^T \boldsymbol{x}_i)} \;\; and\; A = \frac{u-l}{(u-c)(c-l)}$$

The l, u, and c terms above are user defined, where l specifies the lower bound, u the upper bound and c is a centering parameter (Kott, 2011). Because this evaluation does not adjust for non-response or undercoverage, c is set to one. We use the default settings in SUDAAN for the upper and lower bounds, setting l to zero and u to infinity, respectively. In addition, we use a no-intercept model in SUDAAN for better model fit.

Two models are run to calculate calibration estimates for each state: one for expenditure item codes and one for revenue item codes. Total expenditures and revenues are calculated for each (non-certainty) government unit (i) in a state (k) by summing all the item codes (c) associated with a model (m) for each unit. Note that item codes with only one contributing non-certainty unit are excluded from the model and estimated using HT instead because SUDAAN requires at least two contributing units for each estimate. Known state totals from the 2007 CoG-F are used to calibrate so that:

$$t_k = \sum_{i \in U, i \in k, c \in m} x_{kic} \tag{4}$$

Our objective is to find a set of calibrated weights $\{w_i\}$ by adjusting the original design weights $\{d_i\}$ so that the distance between $\{w_i\}$ and $\{d_i\}$ is minimized and satisfies the following constraint:

$$\hat{t}_k = \sum_{i \in S, i \in k, c \in m} w_{ki} x_{kic} \tag{5}$$

For this evaluation, the 2007 CoG-F provides the auxiliary data used to calculate the item code terms (x_{kic}) in the above constraint. Then we can find calibration estimates for each item code in a model for a state using the set of calibrated weights $\{w_i\}$ from SUDAAN:

$$\hat{t}_{kc}^{CAL} = \sum_{i \in S, i \in k} w_{ki} y_{kic} \tag{6}$$

4.3 Proposed Method (EBLUP Estimator)

The goal of our evaluation is to develop and then compare estimators of ALFIN data. More specifically, we seek to estimate state totals of item codes under challenging conditions, where the item codes can have small cell sizes and/or experience volatility in their values over time.

In addition to the design based approach used by the Horvitz-Thompson estimator, and the model-assisted approach used by the calibration estimator, we are also interested in developing a model-based estimator for ALFIN data based on small area estimation methods. To start, consider a Linear Mixed Model (LMM) for a total **Y** on auxiliary variable (**X**) that includes both fixed (**β**) and random (**γ**) components, as shown in equation (7) below:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \quad (7)$$

For this evaluation, we propose the following nested unit-level model (see Chapter 7 in Rao, 2003):

$$\mathrm{Y}_{kic} = \beta_0 + \beta_1 X_{kic} + v_{ic} + \varepsilon_{kic} \quad (8)$$

Where Y_{kic} denotes the cth item code value for the ith government unit in state k for the current year; X_{kic} a corresponding item code value obtained from the most recent Census of Governments; β_0 and β_1 are the fixed effects, the unknown intercept and slope respectively. The v_{ic} are the random complement to the fixed X, or the small area specific random effects for our data; the ε_{kic} are errors in the individual observations $i = 1, \dots, N_c$.The distribution of the random effects corresponds to the deviations of item code log values from the value of $\beta_0 + \beta_1 log(X_{kic})$. In addition, we assume that:

$$v_{ic} \sim^{iid} N(0, \tau^2) \text{ and } \varepsilon_{kic} \sim^{iid} N(0, \sigma^2)$$

To account for the skewed nature of the data and reduce heteroscedasticity, we transform the variable (Y_{kic}) to a log scale:

$$log(\mathrm{Y}_{kic}) = \beta_0 + \beta_1 log(X_{kic}) + v_{ic} + \varepsilon_{kic} \quad (9)$$

Once the data are fit using equation (9) and diagnostics are used to assess goodness of fit, the following model-based predictor is used for $\hat{y}_{kic}$:

$$\hat{y}_{kic} = exp(\hat{\beta}_0 + \hat{\beta}_1 \log(X_{kic}) + \hat{v}_{ic}) \quad (10)$$

Where the estimated fixed and random parameters are computed using SAS® PROC MIXED procedure.

Thus, the state total estimate for an item code c can be obtained by using:

$$Y_{kc} = \sum_{i \in U, i \in k} y_{kic} \tag{11}$$

An estimate of Y_{kc} is given by:

$$\hat{t}_{kc}^{EBLUP} = \hat{Y}_{kc} = \sum_{i \in S, i \in k} y_{kic} + \sum_{i \in S^C, i \in k} \hat{y}_{kic} \tag{12}$$

Where $\hat{y}_{kic}$ is a model-dependent predictor of the non-sampled part (S^C) of the population U.

5. Evaluation Design

This evaluation uses data from the Finance components of the 2007 and 2012 Census of Governments. The universe is the intersection of 2007 data with 2012 data, including only the units surveyed during both census years. For simplicity, the universe is further restricted to include non-zero values on the variables of interest, or the four main groups of item codes. The universe for this evaluation is comprised 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 the three estimators: HT, Calibration and EBLUP. During the analysis, we computed the relative root mean squared error (RRMSE) and relative bias for each estimator from the 1000 samples. Our analysis includes item code totals for 49 states, but excludes Washington DC and Hawaii, because these do not have any non-certainty local governments in sample.

5.1 Mean Square Error (MSE)

Incorporating both the variance of an estimator and its bias, the mean square error (MSE) provides an important measure for comparing estimator quality. In this evaluation, we calculate MSE for all three estimators over all sample replicates. The MSE for a state-item code combination is calculated as follows:

$$\widehat{MSE}(\hat{t}_{kc}) = \frac{1}{1000} \sum_{r=1}^{1000} (\hat{t}_{kcr} - t_{kc})^2 \tag{14}$$

Where $\hat{t}_{kcr}$ is the estimated state total of an item code for one replicate (r), and t_{kc} is the true state total of an item code.

Similarly, the Relative Root Mean Squared Error (RRMSE) is calculated as:

$$\widehat{RRMSE} = \frac{\sqrt{\widehat{MSE}(\hat{t}_{kc})}}{t_{kc}} \qquad (15)$$

5.2 Relative Bias (RB)

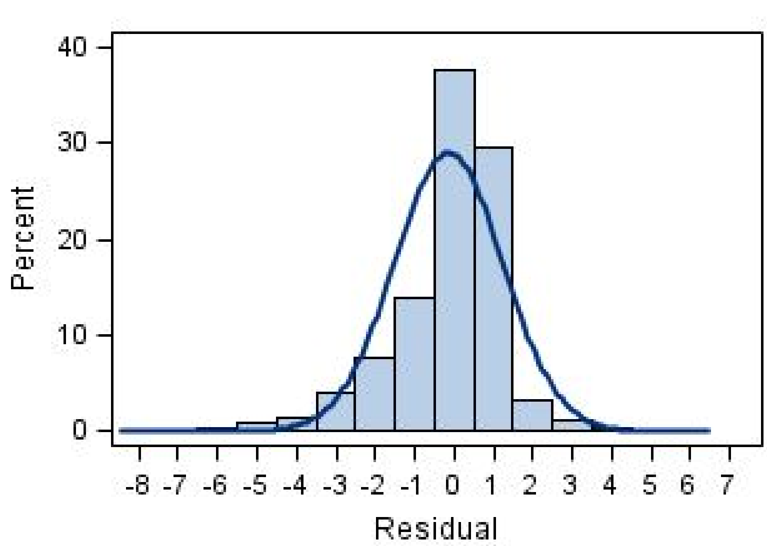
The bias of an estimator is measured as the difference between its expected value and the true value of the parameter being estimated. In our evaluation, relative bias is calculated for a state-item code combination as:

$$\widehat{RB}(\hat{t}_{kc}) = \frac{1}{1000} \sum_{r=1}^{1000} (\frac{\hat{t}_{kcr}}{t_{kc}} - 1) \qquad (16)$$

6. Results

Figure 1 provides an example of a distribution of residuals for a state (California) after undergoing a log transformation from equation (9) above. As shown in Figure 1, the normality assumption in the unit level model of the EBLUP estimator is satisfied.

Figure 1: Normality of the Residuals



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

A comparison of the estimated RRMSE of the three estimators is given in Table 1. The values indicate the number of times an estimator outperforms the others for $\widehat{RRMSE}$. These results are compiled from 1000 replicates with each replicate yielding state-item code estimates over 49 states for a total of 7015 possible estimates per replicate.

Table 1: Number of times Estimator outperforms the others for RRMSE (7015 cells = state by item code estimates)

HT	Calibration	EBLUP
369	180	5003

NOTE: Ties are not listed in Table 1; these can be attributed to 1461 state - item code estimates that are from cells having only certainty units, with 2 exceptions, where the Calibration estimator defaulted to HT.
Data Source: U.S. Census Bureau, 2007 and 2012 Census of Governments: Finance

After excluding the state-item code estimates from cells having only certainty units, EBLUP outperforms the other two estimators for over 5000 of the possible 5552 estimates. Table 1 demonstrates the strength of the EBLUP estimator for $\widehat{RRMSE}$, outperforming the other two estimators in over 90% of cells after ties are excluded.

One unexpected result from Table 1 is that the HT estimator outperforms the other two more times than the Calibration estimator. Table 2 provides a direct comparison between the HT and Calibration estimators for $\widehat{RRMSE}$. The results from Table 2 show that the HT estimator outperforms Calibration for $\widehat{RRMSE}$ in over 63% of the cells after excluding ties.

Table 2: Direct comparison between HT and Calibration -- number of times Estimator outperforms the other for RRMSE (7015 cells = state by item code estimates)

HT	Calibration
3502	2038

NOTE: Ties are not listed in Table 2; these can be attributed to 1461 state - item code estimates that are from cells having only certainty units, with 14 exceptions, where the Calibration estimator defaulted to HT.
Data Source: U.S. Census Bureau, 2007 and 2012 Census of Governments: Finance

Similarly, a comparison of the estimated relative bias ($\widehat{RB}$) for the three estimators is given in Table 3.

Table 3: Number of times Estimator outperforms the others for Relative Bias (7015 cells = state by item code estimates)

HT	Calibration	EBLUP
3011	1837	702

NOTE: Ties are not listed in Table 3; these can be attributed to 1,461 state - item code estimates that are from cells having only certainty units, with 4 exceptions, where the Calibration estimator defaulted to HT.
Data Source: U.S. Census Bureau, 2007 and 2012 Census of Governments: Finance

The results from Table 3 show that the HT estimator outperforms the other two estimators for estimated relative bias in over 3000 of the possible 5550 estimates, or in over 54% of cells after excluding ties. This result is consistent with expectations, because the HT estimator is unbiased with respect to sample design.

Tables 4 and 5 provide an overall comparison of the mean $\widehat{RRMSE}$ and average relative bias for the three estimators relative to cell size. Categories are formed for median cell sizes calculated over the 1000 sample replicates for the non-certainty (πPS) units only, or the units that could be included in a model for a state-item code estimate. The last two categories show that the median cell size can be zero, indicating some state-item code combinations are obscure and have only one or two contributing πPS units, but not for every sample replicate. Two separate categories are formed for the obscure state-item code estimates, reflecting that some of these estimates can also include contributing certainty units, while others are reliant on only the πPS units.

Table 4: Overall Estimator Comparison for Mean RRMSE by Cell Size (7015 cells = state by item code estimates)

Median Cell Size (πPS units only)	Number of Cells	Mean RRMSE		
		HT	Calibration	EBLUP
> 30	1063	0.0711	0.0695	0.0341
21-30	438	0.108	0.107	0.0411
11-20	674	0.169	0.177	0.0512
6-10	567	0.221	0.254	0.0767
1-5	1891	0.426	0.983	0.188
0*	636	0.787	0.786	0.141
0**	285	51.1	51.1	2.20

* Includes other contributing certainty units in the estimates.
** Estimates calculated only from πPS units (no certainty units).

NOTE: Table 4 excludes 1461 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 4 expand on the findings from Table 1, in that EBLUP clearly outperforms the other two estimators for mean $\widehat{RRMSE}$ over all size categories. In addition, the Calibration estimator outperforms HT for the largest size category (n>30), the Calibration and HT estimators are nearly equivalent for the second largest size category (20 < n ≤ 30), and HT outperforms Calibration over the three smaller size categories (1 ≤ n ≤ 20).

For the two obscure state-item code estimates (n≈0), the mean $\widehat{RRMSE}$ for HT and Calibration are approximately equivalent, which reflects how the Calibration estimator defaults to HT when there is only one contributing πPS unit. As expected, the mean $\widehat{RRMSE}$ for obscure state-item code estimates with no certainties using HT/Calibration reaches extreme values. With its use of non-sampled data and model-based approach, EBLUP can offer improved performance under these conditions.

Table 5: Overall Estimator Comparison for Average Relative Bias by Cell Sizes (7015 cells = state by item code estimates)

Median Cell Size (πPS units only)	Number of Cells	Average Relative Bias		
		HT	Calibration	EBLUP
> 30	1063	0.21%	0.49%	2.69%
21-30	438	0.36%	0.77%	3.07%
11-20	674	0.56%	0.92%	3.61%
6-10	567	0.60%	1.40%	5.05%
1-5	1891	3.61%	6.12%	11.8%
0*	636	2.41%	2.65%	12.5%
0**	285	4694%	4692%	204%

* Includes other contributing certainty units in the estimates.
** Estimates calculated only from πPS units (no certainty units).

NOTE: Table 5 excludes 1461 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

Similarly, the results from Table 5 expand on the findings from Table 3, with HT ranked highest in performance for average relative bias, followed by Calibration and then EBLUP over all size categories, with one exception. The exception is the second obscure state-item code category, where the estimates are calculated using only πPS units. Under these conditions, the average relative bias for HT/Calibration reach extreme values, while EBLUP outperforms due to the presence of non-sampled data using the model-based approach.

7. Conclusions

Our evaluation shows a dominant performance by EBLUP with respect to $\widehat{RRMSE}$, while HT outperforms the other two estimators for estimated relative bias due to its unbiasedness property. Based on this research, we plan to select a combination of HT and EBLUP estimators for use with production ALFIN estimations. In the future, we plan to evaluate different forms of the EBLUP estimator that could offer improved performance with relative bias.

8. References

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

Deville, J. and Särndal, C. (1992). "Calibration Estimators in Survey Sampling," Journal of American Statistical Association, 87, 376-382.

Deville, J., Särndal, C, and Sautory, O. (1993)."Generalized Raking Procedures in Survey Sampling," Journal of American Statistical Association, 88, 1013-1020.

Love, E., Barth, J. and Tran, B. (2014). "Evaluating Calibration Estimators for the Annual Survey of Local Government Finances," 2014 Joint Statistical Meetings.

Love, E. and Tran, B. (2013). "Evaluation Study of Calibration Estimation for the Annual Survey of Local Government Finance," 2013 Federal Committee on Statistical Methodology Research Conference.

Kott, P. S. (2011). "WTADJX is Coming: Calibration Weighting in SUDAAN when Unit Nonrespondents Are Not Missing at Random and Other Applications," 2011 Joint Statistical Meetings.

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

Research Triangle Institute (2012). SUDAAN Language Manual, Volumes 1 and 2, Release 11. Research Triangle Park, NC: Research Triangle Institute.

Särndal, C.-E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York, NY: Springer-Verlag.

Tran B. and Dumbacher, B. (2014). "An Evaluation of Different Small Area Estimators for the Annual Survey of Public Employment and Payroll," 2014 Joint Statistical Meetings.

Valliant, R., Dorfman, A.H. and Royall, R.M. (2000). *Finite Population Sampling and Inference*. New York, NY: John Wiley & Sons, Inc.