Benchmarking the Empirical Bayes to Decision-based Estimates in the Annual Survey of Public Employment and Payroll[1]

Justin Nguyen, Mark Bauder, Joseph Barth
Justin.D.Nguyen@census.gov; Donald.M.Bauder@census.gov;
Joseph.J.Barth@census.gov
U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-1912

Abstract
The Governments Division of the U.S. Census Bureau uses small area estimation techniques for several of its surveys. The Annual Survey of Public Employment and Payroll (ASPEP) yields estimates of the number of federal, state, and local government civilian employees and their gross payrolls. The ASPEP sample design is based on state and type of government (county, city, township, special district, and school district) as strata from which a proportional-to-size sampling design is applied. Estimation of government totals at the state and functional level, e.g., air transportation, public welfare, hospitals, etc. are produced. We used Empirical Bayes models to estimate the totals for the cells. At the state and national level aggregates, the totals obtained from the direct estimates are reliable due to big data. Furthermore, we obtain other reliable totals, Decision-based estimates, from which we benchmark. In this paper, we show how to use the Empirical Bayes estimation, and then benchmark the estimates to the direct estimates and Decision-based totals.

Keywords: Governments Unit, Small Area Estimation, Empirical Bayes, Decision-based, Benchmarking

1. Introduction

The Annual Survey of Public Employment and Payroll (ASPEP) is an annual survey conducted by the Governments Division (GOVS) of the U.S Census Bureau to measure the number of state and local government civilian employees and their gross payroll. The ASPEP provides state and local government data on full-time and part-time employment, part-time hours worked, full-time equivalent, and payroll statistics by government functions. Small area methods are used to calculate estimates of local government totals for combinations of state and local government functions. Government functions include, as example, fire fighters, police protection, education, libraries, etc. (See Appendix 1 for a list of government function codes and descriptions.)

---

[1]*Disclaimer: This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.*

The ASPEP is comprised of three components: a census of select federal agencies, a census of the 50 state governments, and a probability sample of local governments (cities, counties, townships, school districts, and special districts). Every five years, in the years ending in "2" and "7", the GOVS conducts the Census of Governments (CoG). The employment component, known as CoG-E, collects public employment and payroll data for over 90,000 governments in the United States.

In this paper, we show how to use the Empirical Bayes (EB) to estimate one variable of interest, the number of full-time employees for each state at function levels, and then benchmark the results to reliable totals at the state level and the Decision-based totals.

2. Sample Design

The sample design is a two-phase, stratified, systematic probability-proportional-to-size (πps) design. Strata are defined by state and types of governments (cities, counties, townships, school districts, and special districts). The size variable is Total Pay (total of full-time pay and part-time pay). It was required that all units in the District of Columbia and Hawaii be selected, so they are made initial certainties, i.e., selected with a probability of 1.0000. After the first phase of sampling, a modified version of cutoff sampling using the cumulative square root of the frequency method (Cheng & Corcoran, 2010) is used to reduce the number of non-contributory sub-counties and special districts in sample. The sample was designed to meet requirements for coefficients of variation at three percent at the state level.

The final sample was combined from the initial certainty, first-phase πps and the second phase πps. Together, approximately 10,149 units were selected for the survey. Table 1 and Table 2 below illustrate the sample breakdown by unit type and by government type.

Table 1: Sample Breakdown by Unit Type

Unit Type	Sample	Universe	Rate (%)
Initial certainty	23	23	100.00
Second certainty	3,674	3,674	100.00
Non-activity	244	18,100	1.35
πps	6,208	68,109	9.11
Total	10,149	89,906	11.29

Source: U.S Census Bureau, 2012 Census of Governments: Organization

Table 2: Sample Breakdown by Government Type

Government Type		Sample	Universe	Rate (%)
1	County	1,393	3,031	45.96
2	City	2,960	19,503	15.18
3	Township	577	16,354	3.53
4	Special District	2,769	36,955	7.49
5	School District	2,450	14,063	17.42
Total		10,149	89,906	11.29

Source: U.S Census Bureau, 2012 Census of Governments: Organization

3. Estimation Methodology

In this paper, we focus our estimation process on: (i) obtain the state totals using the Decion-based, (ii) estimate the total at the unit level using the Empirical Bayes, and (iii) benchmark the Empirical Bayes to the Decision-based at the government function code and at the state level.

3.1 Nested-error Regression Model

Under the nested-error regression methodology, we use a nested-error regression model similar to that of Battese, Harter and Fuller (1988) to predict totals for out-of-sample units. Let Y_{jfi} be the total of interest, such as the number of full-time employees, for function *f*, in unit *i* within state *j*. The nested-error regression model is:

$$\theta_{jfi} = \beta_{0,j} + \beta_{1,j} x_{jfi} + u_{jf} + e_{jfi} \tag{1}$$

where $\theta_{jfi} = \log(Y_{jfi})$, the predictor x_{jfi} is the log of the variable corresponding to Y_{jfi} from the most recent Census of Governments, $u_{jf} \sim N(0, \sigma_{u,j}^2)$ are independent random effects, and $e_{jfi} \sim N(0, \sigma_{e,j}^2)$ are error terms. We obtain estimates $\hat{\beta}_{0,j}$, $\widehat{\beta}_{1,j}$ and $\hat{u}_{jf}$ by fitting the model in (1) separately for each state, using only non-certainty, in-sample units. We refer to the resulting estimator of Y_{jfi} as Empirical Bayes because it is the Empirical Bayes estimator under the model. We define:

$$\hat{\theta}_{jfi}^{\mathrm{EB}} = \hat{\beta}_{0,j} + \hat{\beta}_{1,j} x_{jfi} + \hat{u}_{jf}$$

$$\hat{Y}_{jfi}^{\mathrm{EB}} = \exp(\hat{\theta}_{jfi}^{\mathrm{EB}})$$

Let S_{jf} be the set of in-sample units for state *j*, function *f*, and let S_{jf}^{C} be the corresponding set of out-of-sample units. We estimate Y_{jfi} by

$$\widehat{Y}_{jf}^{\mathrm{EB}} = \sum_{i \in S_{jf}} Y_{gfi} + \sum_{i \in S_{jf}^{\mathrm{C}}} \hat{Y}_{jfi}^{\mathrm{EB}}$$

3.2 Decision-based Model

The Decision-based helped to estimate the synthetic in each cell by providing a stable state total. Decision-based is a process of testing the possibility for combining large and small strata for a government type. This strengthened statistical models for the area of estimation. The state total was estimated by a calibration estimator (GREG) specified as follows:

$$\hat{t}_{y,GREG} = \hat{t}_{y,\pi} + \hat{b}(t_x - \hat{t}_{x,\pi}) \tag{2}$$

where $t_x = \sum_{i \in U} x_i$, $\hat{t}_{x,\pi} = \sum_{i \in S} \frac{x_i}{\pi_i}$, $\hat{t}_{y,\pi} = \sum_{i \in S} \frac{y_i}{\pi_i}$, $\hat{b} = \frac{\sum_{i \in S} (x_i - \bar{x})(y_i - \bar{y}) / \pi_i}{\sum_{i \in S} (x_i - \bar{x})^2 / \pi_i}$

π_i is the inclusion probability, and x_i is the auxiliary data for unit *i*.

The slope $\hat{b}$ was obtained by Decision-based (DB) process proposed by Cheng et al. (2009). The DB method improved the precision of estimates and reduced the mean square error of weighted survey total estimates. The idea was to test the equality of linear regression lines to determine whether we can combine data in different substrata. The null hypothesis H_0: $b_1 = b_2$, that is, the equality of the frame population regression slopes for two substrata. In large samples, $\hat{b}$ is approximately normally distributed, $\hat{b} \sim N(b, \sum)$. Under the null hypothesis, with two sub-strata U_1, U_2 from samples S_1, S_2 of sizes n_1, n_2, we have $\hat{b}_1 - \hat{b}_2 \sim N(0, \Sigma_{1,2})$ where $\hat{b}_1 \sim N(b, \sum_1)$, $\hat{b}_2 \sim N(b, \sum_2)$, and $\sum_{1,2} = \sum_1 + \sum_2$. Therefore, the test statistic is

$$(\hat{b}_1 - \hat{b}_2) \Sigma_{1,2}^{-1} (\hat{b}_1 - \hat{b}_2) \sim \chi_1^2 \qquad (3)$$

Our research showed that it was unnecessary to do the hypothesis for the intercept equality. Because our data analyses led us to observe that we never rejected the null hypothesis of equality of intercepts when we could not reject the null hypothesis of equality of slopes. This makes sense because the 2007 payrolls can be 0 essentially only if the 2002 payrolls are.

The critical value for a test based on (3) is obtained from a Chi-squared distribution with 1 degree of freedom. The test was performed with a significance level of $\alpha = 0.1$. If we cannot reject the null hypothesis, then the slopes estimated in S_1 and S_2 are accepted as the same, and the Decision-based estimator is equal to the GREG estimator for the union of two sample sets, that is, for $S = S_1 \cup S_2$. Otherwise, the Decision-based estimator is the sum of two separate GREG estimators of stratum totals, that is,

$$\hat{t}_{y,DB} = \begin{cases} \hat{t}_{y,greg} & \text{if } \mathrm{H}_0 \text{ is accepted} \\ \sum_{h=1}^{2} \hat{t}_{y,greg}^{h} & \text{if } \mathrm{H}_0 \text{ is rejected} \end{cases} \qquad (4)$$

where $\hat{t}_{y,greg}$ denotes the GREG estimator from the combined stratum S, while $\hat{t}_{y,greg}^{h}$ denotes the GREG estimator from substratum h from sample S_h. DB was produced for 50 states and Washington D.C. totals.

3.3 Benchmarking

In our previous research, the Decision-based estimator produced reliable totals at the state level. Therefore, we benchmarked the Empirical Bayes to the Decision-based estimates at the state level. The ratio for the benchmark within state j and unit i is:

$$r_i = \frac{\hat{y}_i^{\mathrm{DB}}}{\sum_j \hat{y}_{ij}^{\mathrm{EB}}} \tag{5}$$

$$\hat{y}_{ij}^{\mathrm{BM}} = (r_i)\left(\hat{y}_{ij}^{\mathrm{EB}}\right) \tag{6}$$

4. Results

In this research, we estimate the survey total of full-time employment using formulas (5) and (6) above. Table 3, Table 4, and Table 5 provide some examples of the results that compared the benchmarked Empirical Bayes to the Decision-based total at the government function level (financial administration, firefighters, other government administration, housing and community development, natural resources, police-other, etc.) These results ranged from large size of full-time employees (California) to medium size of full-time employees (Maryland) and small size of full-time employees (Montana). The results provide details at the government function code estimates as well as the state total level for the number of total full-time employees. In all cases, benchmarking provides consistent estimates with small coefficients of variation (CV).

Table 3: Comparison Between Benchmarking and Decision-based for State of California for the Number of Total Full-time Employees

Variable	Benchmark	Benchmark CV	Decision based	% difference
Financial Admin.	19,079	0.67%	20,047	4.83%
Firefighters	21,830	0.54%	21,819	0.05%
Other Gov. Admin.	19,254	1.09%	19,621	1.87%
Housing and Dev.	8,721	2.42%	8,222	6.07%
Natural Resources	5,292	3.13%	5,181	2.14%
Police-Other	19,046	0.97%	19,906	4.32%
…	…	…	…	…
Total	881,882		881,882	

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

Table 4: Comparison Between Benchmarking and Decision-based for State of Maryland for the Number of Total Full-time Employees

Variable	Benchmark	Benchmark CV	Decision based	% difference
Financial Admin.	2,113	1.18%	1,847	14.40%
Firefighters	5,876	0.30%	6,080	3.36%
Other Gov. Admin.	3,044	1.79%	2,691	13.12%
Housing and Dev.	2,254	0.75%	2,196	2.64%
Natural Resources	479	2.46%	412	16.26%
Police-Other	3,613	0.79%	3,541	2.03%
…	…	…	…	…
Total	179,074		179,074	

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

Table 5: Comparison Between Benchmarking and Decision-based for State of Montana for the Total Number of Full-time Employees

Variable	Benchmark	Benchmark CV	Decision based	% difference
Financial Admin.	515	2.36%	582	11.51%
Firefighters	448	1.08%	609	26.44%
Other Gov. Admin.	499	2.00%	564	11.52%
Housing and Dev.	135	1.04%	195	30.77%
Natural Resources	144	3.84%	116	24.38%
Police-Other	368	1.27%	461	20.17%
…	…	…	…	…
Total	21,162		21,162	

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

The results yielded if we ran our estimates at the government function code level (financial administration, firefighters, other government administration, housing and community development, natural resources, and police-other) from large areas to small areas (cities, counties, townships, school districts, and special districts), the percent difference between benchmark and Decision-based is smaller for states with a large number of full-time employees. States with few full-time employees yielded a large difference between the two methods. With the help of the Empirical Bayes estimator, we were able to produce estimates for small areas that is necessary to borrow strength from the larger related areas to form direct estimators that increase the effective sample size, and hence, increase the precision. To avoid the differences in our estimates between the two methods at the government function code level, benchmarking has helped to sum up to the state total level. This method is multiplying all the small government function code estimates by a constant factor so that the weighted total agrees with the direct estimate. With this raking method, we were able to achieve the same total level in comparison to the Decision-based results. We have conducted statistical testing procedures required at a 90 percent level of significance.

5. Limitation

Benchmarking helps to adjust total estimates so they agree with reliable totals (Decision-based totals). However, in contrast we will lose the design properties such as consistency in the cells. Another limitation that the Decision-based estimates are dependent on the survey sample size. If the sample size is large enough, then it appeared that the benchmark worked best. However, if we reduced the sample size, then this method provided a large relative bias.

6. Conclusion

Our other research has shown that the Empirical Bayes outperformed other estimators such as, Hortvitz-Thompson, SPREE (Structure Preserving Estimation), and the composite estimator. We were able to obtain reliable estimate totals using benchmarking. However, as mentioned above, benchmarking can lead to design inconsistency.

7. Future Research

There are a few remaining issues that need further research. We will explore in more detail the use of the Empirical Bayes with alternative assumptions other than normality. We will continue our research to identify methods that can keep design consistency while benchmarking.

Appendix 1: List of Government Function Code and Description

Government Function Code	Description
000	Totals for Government
001	Airports
002	Space Research and Technology
005	Correction
006	National Defense and International Relation
012	Elementary and Secondary-Instruction
014	Postal Service
016	Higher Education-Other
018	Higher Education-Instructional
021	Other Education
022	Social Insurance Administration
023	Financial Administration
024	Firefighters
025	Judicial and Legal
029	Other Government Administration
032	Health
040	Hospitals
044	Streets and Highways
050	Housing and Community Development
052	Local Libraries
059	Natural Resources
061	Parks and Recreation
062	Police Protection-Officers
079	Welfare
080	Sewerage
081	Solid Waste Management
087	Water Transport and Terminals
089	Others and Unallocable
090	Liquor Store
091	Water Supply
092	Electric Power
093	Gas Supply
094	Transit
112	Elementary and Secondary-Other Total
124	Fire-Other
162	Police-Other

References

Fuller, W. (2002), “Regression Estimating for Survey Samples”, *Survey Methodology*, 5 – 23.

Kott, P. (2012) “An Introduction to Calibration Weighting for Establishment Survey”, *Survey Methodology*, 8 – 10.

Battese, Harter and Fuller (1988), “Nested-error Regression Methodology”, *Survey Methodology*.

Brackstone and Rao (1976), “Raking Ratio Estimators”, *Survey Methodology*, 42 – 70.

Sarndal, C.-E. (2003), *Model Assisted Survey Sampling*, New York: Springer.

Tran, B. and Cheng, Y. (2010), “Application of Small Area Estimation for Annual Survey of Public Employment and Payroll.”

Cheng, C. and Corcoran, C. (2010), “Alternative Sample Approach for the Annual Survey of Public Employment and Payroll.”

Dumbacher, B. and Hill, C. (2014), “Description of the 2014 Sample Design for the Annual Survey of Public Employment and Payroll.”