Alternative Sample Allocations for the U.S. Current Employment Statistics Survey   August 2005

G. Erkens, J.Gershunskaya, and L. Huff, BLS

G. Erkens, Bureau of Labor Statistics, Office of Employment and Unemployment Statistics

Suite 4985, Postal Square Building, 2 Massachusetts Ave., N.E., Washington, DC 20212

Abstract

The Bureau of Labor Statistics began conducting the Current Employment Statistics Survey well before probability sampling was accepted as a standard for government surveys. The sample for this survey was recently (1997) converted from a quota sample to a probability sample. The resources for each State's probability sample were fixed to the size of its quota sample. Total resources for data collection are limited to the sum of all States' resources.

Since the inception of the probability redesign, some states experienced a rapid employment increase, causing State sample needs to change. We explored different methods for revising the resources available to each State while keeping the total number of resources fixed. One method involved equalizing relative standard errors (RSEs) between all States. We also reviewed several bounded optimum allocations to produce an estimate of national employment with the smallest sampling error under the given conditions. This paper reviews the methods we investigated and their effects on the national and State RSEs.

Keywords: Optimum Allocation, Relative Standard Error, Variance,

1. Introduction

The Current Employment Statistics (CES) survey began in 1915 as a quota sample—well before probability sampling was a standard for government surveys. Throughout the survey's history attempts were made to revise the CES for a probability design. Those attempts were refused for a variety of reasons. With the benchmark revision in 1991, there was greater pressure to revise the CES and implement a probability design. When the Bureau researched a probability design, the resources available to each state were set at their respective quota sample levels (Butani, Stamas, Brick 1997). Since the original research for the CES probability redesign, States' needs have changed. Many States' employment levels have increased dramatically, while the employment in other States has increased at a lower rate. Due to these changes in employment levels, the current allocation of CES resources affects the accuracy of both State and National CES estimates. With respect to their given employment levels, some States possess an insufficient number of resources. We attempted to revise the allocation of resources amongst the States in order to more accurately reflect their needs, while keeping the total number of resources for the CES survey fixed.

This paper contains the remaining sections. Section 2 contains an overview of the CES survey's design. Section 3 covers the allocation procedures CES uses to optimally allocate resources within a State. Section 4 details three different methods we used to reallocate resources among the States. Those methods include: equal State CVs, an unbounded National optimum allocation, and a bounded National optimum allocation. Section 5 displays the different procedures' effects on the State and National relative standard errors (RSEs) and sample sizes, and section 6 contains a summary of the methods and results.

2. Survey overview

The CES survey is a Federal-State cooperative survey. Statewide, local area, and nationwide estimates of employment are important aspects of the survey. The survey data is collected by the national office, and a State's data is shared with that State, allowing them to produce their

own statewide and local area estimates of employment.

The CES frame uses the unemployment insurance account (UI) information collected by the States. State offices collect information such as employment, wages, and industry coding. That information is then provided to the BLS National office which compiles the Quarterly Census of Employment and Wages (QCEW, formerly referred to as the ES-202), providing a universe count of all UIs and the establishments reported under each UI (that gives a multi-worksite report). The CES sample frame is constructed using 12 months of employment information for each UI account.

The CES survey uses a stratified random sample of UIs. Each UI account is assigned to a stratum based upon its size class and industry. Employment is summed up for all establishments or reporting units (RUs) in the UI for each month, and the size class is based upon the maximum employment over the past 12 months. Size class descriptions are given in table 1. For the industry, each UI is assigned a North American Industry Classification System (NAICS) supersector, which depends upon the industries of a UI's composite RUs. The NAICS supersectors are listed in table 2. The intersections of each supersector and size class make up the strata. With 8 size classes and 13 industries, there are 104 possible strata.

Size	Employment
1	0-9
2	10-19
3	20-49
4	50-99
5	100-249
6	250-499
7	500-999
8	1000+

Table 1: CES Size Classes

While the number of UIs selected from each stratum is typically determined by an optimum allocation, constraints exist.

Nearly all UIs are sampled from strata with large employment UIs (size class 8). CES also has the following requirements for each stratum:

1. Minimum sample of 2 UIs from each stratum, assuming that $N_h > 1$
2. Maximum sample weight of 100 for each stratum ($N_h/n_h <= 100$)
3. Minimum of 50 UIs sampled from each industry except mining, and a minimum of 30 UIs sampled from mining.

NAICs Supersector	NAICS Code
10—Mining	11,21
20—Construction	23
31—Manufacturing, Non-Durables	33, 321, 327
32—Manufacturing, Durables	322, 323, 324, 325, 326
41—Wholesale Trade	42
42—Retail Trade	44, 45
43—Transportation, Warehousing, Utilities	22, 48, 49
50—Information	51
55--Financial	52, 53
60--Professional	54, 55, 56
65--Educ. & Health Services	61, 62
70--Leisure	71, 72
80—Other Services	80

Table 2: CES Industries

3. Allocation overview

The CES survey uses an optimum allocation for each state. For a given cost, CES allocates sample to produce an estimate of statewide employment with the smallest sampling error. We first compute some basic statistics for each month on a State's file of UIs.

For each UI we possess 12 months of data. The stratum variance is computed for each month, and the stratum Pearson correlation is computed for each set of adjacent months. Variances are computed as described in Cochran (1977, chap. 2). The variance and correlation are then averaged for the year.

Average Variances and correlations are calculated for 5 years (the current year and four years previous), and the median is calculated for those 5 years (Butani, Stamas, Brick 1997). This method gives a robust result that will not change

drastically between years. After calculating the medians, we then compute an adjusted variance which is used in the allocation.

$$S_h^2* = 2(1-\rho_h)S_h^2$$

This formulation is a simplified version of the variance for a combined ratio estimator, for which we make the following assumptions:

1. The ratio of statewide employment change for each industry (R) between each pair of months is approximately 1.
2. The stratum variance for each month is approximately the same.
3. The stratum correlation for each pair of adjacent months is approximately the same.

The CES survey uses the following information for each stratum as input for the optimum allocation:

- N_h—Total number of UIs in stratum h
- $S_h^* - \sqrt{S_h^2 *}$
- c_h—Cost for stratum h

The cost is computed by calculating the average number of counties per UI in a stratum. After removing the cost for a State's certainty UIs, we use the remaining available cost to allocate sample UIs to each stratum. The optimum allocation runs through the following algorithm

1. For a given cost, optimally allocate sample UIs to the strata in a State. Sample size in each state is n.
2. Determine if overallocated to stratum h ($n_h > N$)$_h$ in State. If yes then the UIs in stratum h are sampled with certainty.
3. Remove cost of the certainty UIs from the cost (C_{St}) for the State.
4. After removing certainty strata from the calculations, repeat steps 1-3 until there is no more overallocation (Cochran 1977, chap. 5).

4. Updating Costs for each State

Since the beginning of the CES Redesign, State employments in different States have changed— sometimes substantially. During the past 15 years the change is even more dramatic. Employment has grown in all States, but some

States have shown exceptional growth. Figure 1 plots 14 year-old Statewide employment against their 14 year employment growth. The x-axis is plotted on a log scale, and tick marks on that axis occur at 250,000, 500,000, and multiples of 1,000,000. The last tick mark is at 8,000,000. The graphing symbol is the State's postal abbreviation (FIPS), and the line denotes the approximate average growth. Note that specific states have grown much more relative to their original employment levels. Large employment increases affect the National and States' estimates. In order to reflect the changing needs of some States, we attempted to devise some methods to find a more suitable distribution of CES resources amongst the States. To decide on a new distribution of CES resources, we used the following set of criteria.

1. The new distribution should provide a good estimate of National total employment, and the error for National total employment should not increase.
2. The new distribution should be "just".

The second criterion is the more nebulous of the two, so it is difficult to objectively assess. There is a need to balance between State sample needs and the inevitable cuts to some State sample sizes. This paper covers four of the methods we reviewed. Results for these methods are in section 5.

The first method we investigated equalized RSEs for all States. The following algorithm establishes a set cost for each State such that each State has the same RSE.

$RSE = .009$

$\delta = .002$

$C = $ the total number of non-certainty CES resources

1. Let

$$C_{st} = \frac{\sum_{h \in St} N_h S_h \sqrt{c_h}}{(CV * AE_{St})^2 + \sum_{h \in St} N_h S_h^2},$$

2. $RSE* = \begin{cases} RSE - \delta, \text{ if } \sum_{St=1}^{51} C_{St} < C \\ \\ RSE + \delta, \text{ if } \sum_{St=1}^{51} C_{St} > C \end{cases}$

3. $RSE = RSE*$

4. Stop when $abs(\sum\limits_{Sti=1}^{51} C_{St} - C) < 50$

For a given C_{St}, each State's resources are optimally allocated. If there is any overallocation in any State, then the entire process is repeated: removing the values for the certainty strata from C_{St} and the other calculations. Once there is no more overallocation in any of the States, the procedure ends.

The second method we reviewed was an unconstrained National optimum allocation. National optimum allocation uses the typical formula for optimum allocation (Cochran 1977). We used the following algorithm for the National optimum.

1. n_h = Sample size in stratum h. There are 5304 total possible strata.
2. C = Total CES Sample Resources – Resources used for Certainty Stratum
3. $n_h = C \dfrac{N_h S_h / \sqrt{c_h}}{\sum\limits_{h=1}^{H_{Nat}} N_h S_h \sqrt{c_h}}$, where H_{Nat}

    represents the total number of stratum in the Nation.
4. If overallocation in any stratum ($n_h > N_h$), then set $n_h = N_h$, and start again at step 1.

The last two allocations were constrained versions of the National Optimum allocation. Each one attempted to compromise between the current allocation of resources and the National optimum allocation. For the last two allocations we used the following constraints:

1. $\dfrac{N_{St}}{n_{st}} \leq 40$

2. $RSE = \max\left(RSE_{Current}, RSE_{National\,Optimum}\right)$
3. $n_{St\,New} \geq (1-p)n_{St\,Old}$
    a. $p = .15$
    b. $p = .25$

N_{st} is the total number of units in a State's frame, and n_{st} is the number of State sample units. Parts a and b of the last constraint were each part of a different allocation. These two constraints allow for a maximum sample decrease in each State of 15% and 25% respectively. The PROC NLP procedure in SAS was used for the two constrained allocations.

5. Results

Each of the allocations displayed unique advantages. The equal RSEs method kept the average RSE at its lowest point (.495), but the sample loss was extreme in many States. Sample gains are shown graphically in figure 2. The chart displays the sample size under the current allocation on the x-axis, and the new allocation is on the y-axis. The plotted line is a reference line with slope 1 and 0 intercept. The plotting symbol is the State FIPS code. All States above and below that line have a respective increase or decrease. Note that most States lie above the line, while a few States lie below the line. This characteristic shows that many States would gain sample under the Equal RSEs method, but that gain would come at the expense of very heavy losses for larger States such as California. We also reviewed a constrained version of the Equal RSEs method, in which the National RSE was bounded and State RSEs were allowed to vary slightly, but that revision still caused extreme sample losses in many large States.

The unconstrained National Optimum allocation offered the lowest National RSE, but the RSE for some States increased to over 2%. In this allocation the sample size losses in several States were also very extreme. A graphical presentation of the sample gain under this allocation is shown in figure 2. Note that California (CA)—the State with the largest sample under the current allocation—would gain sample. Other larger States would also gain sample: States such as Illinois (IL), New York (NY), and Florida (FL). Figure 3 gives the same information, but it focuses in on all States with an original sample size less than 7,000. Figure 3 exhibits sample gains for some States under the National Optimum allocation, but the majority of States would loose sample. Sample losses for some States would still be relatively extreme. Under this allocation the National RSE decreased (relative to the current allocation) to .072%.

These problems led to the constraints placed upon the National Optimum allocation. They reduced the steep State sample losses that occurred in the Equal RSE and unconstrained National Optimum allocations. The limit of a

25% sample loss helped to curb heavy losses in many States; nevertheless, this loss was still deemed too severe. Relatvie to the current allocation, the National RSE under this allocation decreased to .074%. Under the constraint of a 15% sample loss the National RSE decreased to .075%, and the sample gains and losses were—as expected—much less. Table 3 displays quantiles of sample gains under the different allocations.

6. Summary

We attempted to find some just method of reallocating resources amongst States in the CES survey. To achieve this goal we had two main objectives: States that need more sample should acquire more sample, and States should not experience a large sample loss. Each of the different methods provided certain benefits and drawbacks. The equal RSEs method did give more resources to States that needed it, but it accomplished this goal at the expense of very large and unacceptable sample losses in other States. This method also decreased the accuracy of the National CES estimate.

The unbounded National Optimum allocation also delivered more resources to States in need, but this transfer could only occur with heavy and unacceptable sample losses to some smaller States. Under this method the National RSE also decreased, but the RSEs for several States increased to an undesirable level.

The bounded National Optimum allocations put upper limits on the States' RSEs and lower limits on how much sample a State could loose. The limits helped us balance between our different objectives. States that needed more sample received it, but they did not receive as much sample as under the unbounded allocation.

The different variations of the National optimum allocation push us towards an allocation of resources that is proportional to the percentage of National employment contained in a State. This proportional allocation is relatively easy to justify as "fair" or "just". The final reallocation will be decided by the CES policy council.

The views expressed in this paper do not represent policy at the Bureau of Labor Statistics.

References

Butani, Shail. George Stamas, Michael Brick. (1997), "Sample Redesign for the Current Employment Statistics Survey". 20 June 97. http://www.bls.gov/ore/pdf/st970260.pdf

Cochran, William G. (1977), Sampling Techniques. New York: John Wiley & Sons, Inc.

Allocation	Minimum	5%	10%	25%	50%	75%	90%	95%	Maximum
Equal RSEs	-13124	-3769	-2302	-280.5	669	1541	1928	2200	2594
National Optimum	-2296	-2004	-1546	-999	-400	469	1714	3178	4944
National Optimum with 15% limit	-1535	-873.5	-511	-390	-89	123.5	474	1394	2899
National Optimum with 25% limit	-1860	-1403.5	-799	-586.5	-89	242	725	2061.5	3668

Table 3: Quantiles of States' sample size Changes for each New Allocation

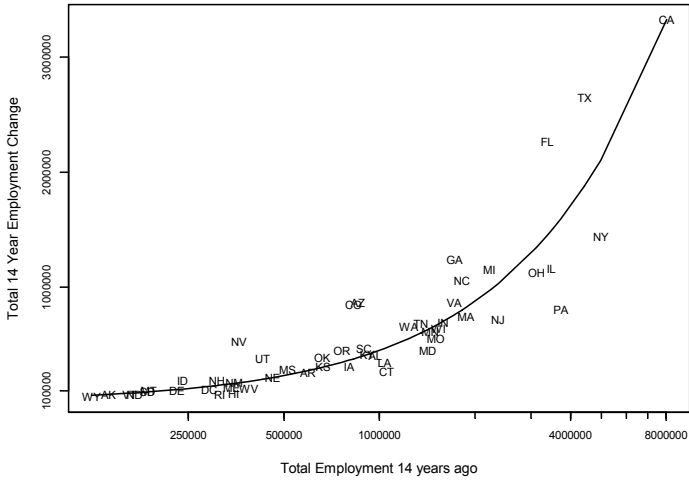## Plot of Total Employment and 14-year Employment Change

Figure 1: Plot of State Employment 14 years ago against 14 year growth. Tick marks on x-axis at 250,000, 500,000, and multiples of 1,000,000. The last tick mark is at 8,000,000.

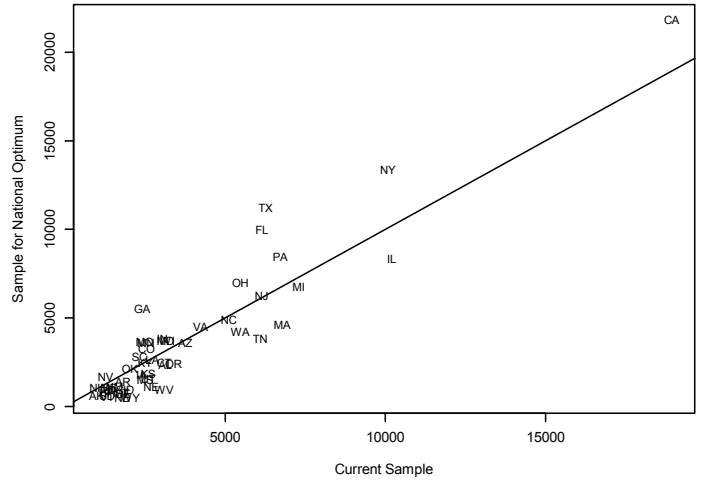## Sample Size Comparison Current Sample vs. National Optimum

Figure 3: Plot of Current CES sample sizes for all States against New sample size under Unconstrained National optimum allocation.

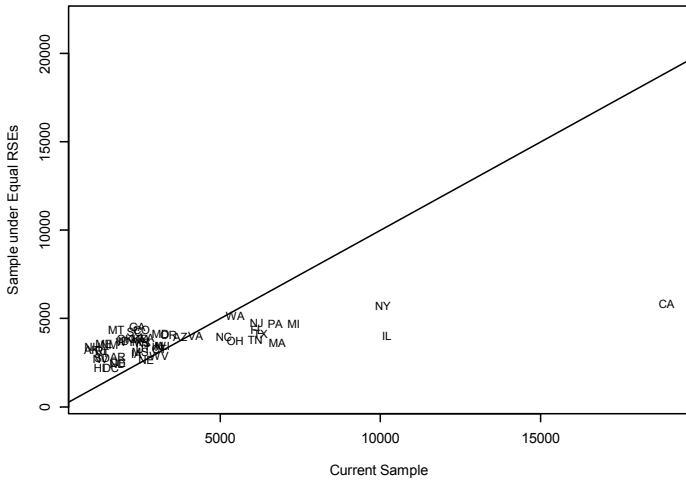## Sample Size Comparison Current Sample vs. Equal RSEs

Figure 2: Plot of Current CES sample sizes for all States against New sample size under the Equal RSEs allocation.

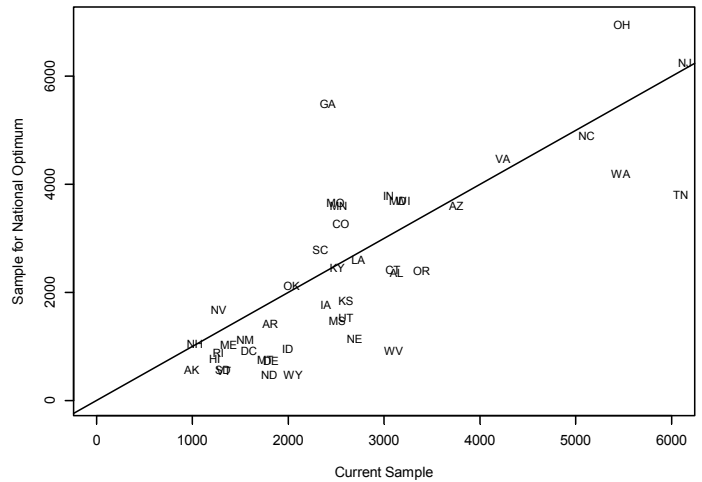## Sample Size Comparison Current Sample vs. National Optimum

Figure 4: Plot of Current CES sample sizes for all States against New sample size under Unconstrained National optimum allocation. Graph is cropped to view States with current CES sample size less than 7,000.