Application of Fay's Method for Variance Estimation in the National Compensation Survey Benefits Products

Jonathan J. Lisic[1], Omolola E. Ojo[2]
[1] Bureau of Labor Statistics, 2 Massachusetts Ave., N.E., Room 3160, Washington, DC 20212, Lisic_j@bls.gov
[2] Bureau of Labor Statistics, 2 Massachusetts Ave., N.E., Room 3160, Washington, DC 20212

Abstract
Data on employee benefits is one of the key products produced by the National Compensation Survey (NCS). The NCS uses a stratified three-stage sample: areas (PSUs), establishments, and occupations. The weighting procedure includes adjusting sample weights for establishment non-response, occupation non-response, any unusual data collection situations, and post-stratifying to known employment control totals. This paper illustrates how Fay's method of balanced repeated replication (BRR) was applied to estimates generated from this complex sample design to obtain variance estimates for the NCS benefits products. The following topics will be presented: an introduction to the NCS stratified three-stage sample; a description of our utilization of Fay's method, including formation of variance strata, variance PSUs, and replicates; a detailed analysis of the magnitude of estimated variances (BLS 2008).

Key Words: Replication, Multi-stage sampling, Variances

1. Introduction

Employee benefits measures are one of the key products derived from the integrated National Compensation Survey (NCS) sample. These measures cover the incidence and detailed provisions of selected employee benefit plans. Incidence data are presented as the percentage of employees who have access to, or participate in, a broad selection of benefits. Provisions data are available for benefits, such as paid vacations and holidays, disability insurance (short-term and long-term), and life and health insurance.

Variances for these measures have not been published since the integration of several existing wage and benefit surveys into the current NCS survey close to a decade ago. This year the NCS survey will resume publication of these variances. This paper will give a brief outline of the National Compensation Survey, describe estimates we will obtain variances for, provide a description of the NCS implementation of Fay's method of balanced repeated replication, and present our analysis of the variances.

2. Description of the National Compensation Survey

The National Compensation Survey (NCS) is an establishment survey of pay and benefits conducted by the Bureau of Labor Statistics (BLS). It is used to produce three general types of survey outputs: employment cost data, employee benefits data, and national and locality wage data. The employment cost data includes the Employment Cost Index (ECI), a series of indexes that track quarterly and annual changes in wages and benefit costs, and quarterly cost level information on the cost per hour worked of each component of pay and benefits. The employee benefits data includes the incidence and provisions of selected employee benefit plans and are published once a year. The national and locality wage data include annual publication of occupational wages for a sample of localities, census divisions, and for the nation as a whole. All state and local governments and private sector industries, except for farms and private households, are covered in the survey. Only employees that receive a market wage are covered.

The BLS Quarterly Census of Employment and Wages (QCEW) serves as the sampling frame for the NCS survey and was used as the administrative data for this study. The QCEW is created from State Unemployment Insurance (UI) files of establishments, which are obtained through the cooperation of the individual state agencies.

The integrated NCS sample consists of five rotating replacement sample panels. Each of the five sample panels will be in sample for five years before being replaced by a new panel selected annually from the most current frame. The NCS sample is selected using a three-stage stratified design with probability proportionate to employment sampling at each stage. The first stage of sample selection is a probability sample of areas; the second stage is a probability sample of establishments within sampled areas; and the third stage is a probability sample of occupations within sampled areas and establishments.

The first stage of the NCS sample occurs at the national level across geographic areas. These Primary Sampling Units (PSUs) are based on the 2003 Office of Management and Budget (OMB) area definitions. Under the OMB definition there are three types of statistical areas. These area types are defined as Metropolitan, Micropolitan, and Combined Statistical Areas. Combined Statistical Areas (CSAs) are defined as a combination of adjacent Metropolitan and Micropolitan areas that meet certain conditions. Outside of these areas exists a number of counties. These counties are referred to as Outside Core Based Statistical Areas (CBSA). For selection purposes these outside CBSA's are organized into clusters to create more heterogeneous primary sampling units.

In 2004, a new area sample was selected for the NCS. This sample contained 152 areas. In this sample 57 areas were selected with certainty, where certainty areas are defined as having employment greater than 80 percent of the final sampling interval, which is obtained through an iterative process. The remaining areas consisted of 60 non-certainty metropolitan areas, 22 non-certainty Micropolitan areas, and 13 non-certainty outside CBSA county clusters.

The second stage of this design occurs at the establishment level within each selected area. This stage consists of a sample of nonagricultural private business establishments and State and local government operations, stratified by ownership and industry. Within each of the strata we employ PPS systematic sampling with frame employment as the measure of size (MOS) To ensure that no unit has a probability of selection greater than one, we remove all units that would be selected with certainty before the sampling process, and set their sampling weights to one. After the sample of establishments is selected, it is used for the third stage of the sampling process.

The third stage of this design occurs at the occupational level within each selected establishment. A sample of jobs is drawn from each of these establishments using PPS systematic sampling. To ensure that jobs are defined consistently across all establishments the Standard Occupational Classification (SOC) manual is used to classify these jobs, based upon duties, into occupations. After this selection and classification we create our smallest aggregate unit known as a quote, which is a distinct combination of time or incentive pay, level, union membership, full-time or part-time status, and occupation.

Our study focuses on the incidence and provisions estimates within the National Compensation Survey. These estimates are provided through proportions and averages. The proportional estimates are used to measure incidences such as access or participation in a particular benefit plan, or the percentage covered by a specific provision. Under access, we measure the percentage of people who have access to a particular benefit plan or specific benefit feature, or will have access after a service requirement is met. Participation is measured as the percentage of people who have access to a particular benefit plan or specific benefit feature and participate. Averages include provisions such as annual deductibles or lifetime maximum. Only those that are covered by these provisions are included in these estimates. All of these estimates are defined over a particular domain such as blue or white collar, industry, establishment size, type of benefit plan, and plan details.

The formulas used to calculate access incidence, participation incidence, and provisions are defined below: Access is defined by

$$A_D = \frac{\sum_{q \in D} W_q X_q}{\sum_{q \in D} W_q} \times 100$$

where D is the domain of interest. W_q is the final quote weight for quote q, and X_q is defined in the set {0, 1} where one implies employees within a particular quote have access to this benefit while zero implies that these employees do not have access. Participation is defined in a similar manner as

$$I_D = \frac{\sum_{q \in D} \sum_{j \in q} W_q P_{qj}}{\sum_{q \in D} W_q} \times 100$$

where D is the domain of interest, W_q is the final quote weight for quote q, and P_{qi} is the percentage of workers in quote q participating in plan j. Averages, such as of deductibles or contribution cost, are calculated by the ratio of the weighted deductible or contribution costs of participants divided by the weighted total participation as seen below:

$$\hat{\hat{Y}}_D = \frac{\sum_{q \in D} \sum_{j \in q} W_q Y_{qj} P_{qj}}{\sum_{q \in D} \sum_{j \in q} W_q P_{qj}}$$

Where D is the domain of interest, W_q is the final quote weight for quote q, P_{qj} is the percentage of workers in quote q participating in plan j, and Y_{qi} is the employee contribution or deductible. The weight W_q is calculated as the product of the number of employees within an establishment at initial collection divided by both the probability of selecting this establishment and the assigned number of quotes. However, revisions are made to the weight to account for establishment and occupational non-response and differences between the selected and the responding unit (BLS 2008).

3. Utilization of Fay's Method

National level estimates for benefit data consist of an expansive collection of domain estimates. Domain estimates are simply ratio estimates based upon a subpopulation. Unfortunately just like ratio estimates, calculation of exact variances measures for domain estimates from a complex survey are both difficult and timely to compute. Therefore a computationally efficient approximation of the true variance is desired.

Methods of approximation for variance of domain estimates include Taylor Series, Balanced Repeated Replication, Jackknife, and Fay's method of Balanced Repeated Replication. Paben (1999) has done prior analysis comparing these methods of variance estimation for domain estimates in the NCS. Through his survey he analyzed the bias associated with each variance estimation technique over NCS data. He concluded that all of these methods produced similar variance estimates, and are asymptotically valid. Therefore it would make the most sense for us to choose the method that is easiest to compute and implement in a production environment.

We determined that Fay's method of Balanced Repeated Replication was our best choice. This decision was made due to the ease of implementation and to provide consistency in variance estimation methods across all NCS estimates. The only potential issue with this choice is that Fay's Method has a tendency to underestimate the variance for small sample sizes (Paben, 1999). Fortunately due to privacy and stability concerns we are unlikely to produce variance estimates where this condition would be an issue.

Fay's method of Balanced Repeated Replication can be considered a generalization of traditional Balanced Repeated Replication, therefore the following description of Fay's Method will be identical to traditional BRR except where noted. Fay's method of BRR variance estimation deals with the stratified sampling problem of calculating a variance estimate when there are only two PSUs in each stratum.

The approach Fay's method of BRR takes to deal with this issue is to use a pertubation of the weights within each PSU and stratum combination to produce an estimate similar to the original. We will refer to this new estimate as a pseudo-replicate since each pseudo-replicate is not a distinct subsample as in the case of other methods such as random groups. Once several pseudo-replicates are calculated, the sum of the squared differences between the original estimate $\hat{Y}_r$, and the R pseudo-replicates $\hat{Y}_r$, can then be divided by $R(1-k)^2$ to produce an estimate of the variance as seen below:

$$Var(\hat{Y}) = \frac{1}{R \cdot (1-k)^2} \sum_{r=1}^{R} (\hat{Y}_r - \hat{Y})^2$$

There are two important notes about this method described. First, the only difference between BRR and Fay's method of BRR is the function that perturbs the weights. Fay's method of BRR defines the function such that if the weights in either the first or second PSU in a stratum are increases by a proportion k, then the weights in the other PSU within that stratum is decreased by *k*. Under traditional BRR the value of *k* is one, this implies that we double the weight of one PSU and discard the other.

The second note is that the selection of which PSUs to increase for the formation of pseudo-replicates is not done in an arbitrary way. Ideally we would like to choose all possible pseudo-replicate possiblities; unfortunately the calculation of all possible selections may be impossible for a large number of strata. Instead a balanced selection allows us to choose only a subset of these pseudo-replicates without any loss of information.

To define a balanced replication we will first need to formalize the construction of each pseudo-replicate. Let's consider the representation of a pseudo-replicate as a column vector of size *L*, where *L* is the total number of strata. This representation allows us to associate each stratum with a distinct element in the vector. Using this relationship we can then let each element in the vector correspond to one of the two possible perturbations of weights. We typically associate the value of one with the weight perturbation mentioned earlier involving an increase of the weights in the first PSU, and the value negative one with the weigth perturbation involving increasing the weights in the second PSU. Therefore each pseudo-replicate can be represented simply by a vector whose elements have two different values. We will then formalize the value assigned to each element of a particular pseudo-replicate as $\varphi_{lr} \in \{-1,1\}$, where φ_{lr} would be the pertubation assignment associated with the stratum *l* of pseudo-replicate *r*. A balanced replication would then be defined as a set of *R* pseudo-replicates such that they have an orthogonal row space as defined by:

$$\sum_{r=1}^{R} \varphi_{lr} \cdot \varphi_{qr} = 0, \forall l \neq q$$

However, this condition is not guaranteed for all values of *R*, but can be satisfied if number of pseudo-replicates is equal to the smallest multiple of four greater than or equal to the number of strata.

A simple way to obtain this balanced design is to use a Hadamard matrix. A Hadamard matrix is a square matrix with mutually orthogonal columns and rows where all elements are either one or negative one. If we let each replicate corespond to a row, and each stratum correspond to one column we can then use the associated matrix values to determine the pertubation for a particular replicate and stratum combination.

With this balanced selection we can then show that for linear estimators we would have an unbiased variance estimator (Wolter, 1985). This isn't necessarily the case for non-linear estimators, but empirical studies have shown that the bias is similar to other variance estimation techniques and is similar to linearization (Paben, 1999).

Fay's method of BRR, and traditional BRR may also be generalized to deal with the case where there are more than two PSUs in a strata. To deal with this issue we can then regroup the strata into a set of variance strata then randomly assign each PSU, or units within the PSU for multi-stage designs, to one of two variance PSUs of roughly equal weight.

The National Compensation Survey variance estimates are produced by Fay's method of BRR with k set to 0.5. The PSUs within this survey are combined to form variance strata by seperating the certainty from the noncertainty areas.

Each certainty area is given its own variance stratum. Within these variance strata the variance PSUs are formed by dividing establishment quotes in one of two ways. If an establishment was selected with certainty we will divide it's quotes between both variance PSUs. If the establishment was not selected with certainty we will place all of it's quotes in the same variance PSU. In both cases the initial variance PSU assignment is determined by a random start from a bernoulli distribution. The remaining quotes and establishments then follow an alternating assignment until they are exhausted.

The remaining variance strata are formed by combining two or more areas, where each area is assigned to one of the two PSUs. Since this is a five panel survey, once a unit has been assigned to a variance PSU it will retain its assignment until it exits the survey.

4. Analysis of the Magnitude of Estimated Variances

To analyze the variance for Incidence and Provisions estimates in the NCS survey we utilized data from the *Employee Benefits in Private Industry in the United States, 2005* publication. The initial objective was to look at the variances through an assessment of the distribution of standard errors of the estimates. Primarily this focus was to see how many of the estimates have reasonably useful values to draw meaningful economic conclusions.

To perform this analysis we utilized prior research at the Bureau of Labor Statistics on incidence and provisions reliability (Guciardo, 2008). The goal of this prior research was to define a set of criteria to evaluate the reliability of the domain estimators. The result was the creation of a hybrid test that used a combination of percentage relative standard error and standard error depending on the size of the estimate. This followed the rational that relative standard errors are a useful measure of stability for domain estimates except when they are near zero. To remedy this situation standard errors are employed below a pre-determined ratio size to measure stability.

To apply these criteria we first removed estimates from the *Employee Benefits in Private Industry in the United States, 2005* publication that either had no reported variance, had less than 30 item level measurements, or for percentage estimates outside the range of 0.5% to 99.5% since results are reported in integer percentages. This provided us with a total of 12,458 estimates with corresponding standard errors and relative standard errors. Unfortunately this data was from the prior area design, and therefore was limited to 63 variance strata for all areas. However, the results should be sufficiently similar to the redesign to not make this an issue.

Table 1: Hybrid Test Results for *Employee Benefits in Private Industry in the United States, 2005*

%RSE	**Standard Error**					
	Total	**< 1**	**1 - 1.5**	**1.5 - 2**	**2 - 3**	**> 3**
Total	100%	25%	15%	13%	18%	29%
> 50	3%	1%	0%	0%	0%	1%
40 - 50	5%	2%	1%	0%	1%	1%
30 - 40	10%	3%	1%	1%	2%	3%
20 - 30	17%	6%	2%	2%	2%	5%
10 - 20	24%	6%	3%	3%	3%	8%
< 10	42%	7%	7%	7%	10%	11%

Applying the hybrid test to this data we obtained the results within Table 1. Columns in this table represent ranges of standard error, while the rows represent ranges of relative standard error. Each cell then represents the percentage of the 12,458 estimates that fall within these bounds. Using this information we can see that five percent of all estimates had standard errors greater than three or relative standard errors exceeding 30%. Pushing the relative standard error back to 20% only increased the total percentage of failures to 10%. These more exteme estimates with larger standard errors and relative standard errors were primarily regional estimates. In fact 75% of the estimates that failed at the 30% level were regional domain estimates. The remaining domain estimates had a small number of participating quotes, or were borderline failures.

Attempts were made to address the issue of stability of these variance estimates. These attempts included a traditional empirical simulation based upon the construction of a frame representative of our population. Unfortunately for most of our estimates there was not sufficient similar data to build a sufficiently large representative frame. Initial work was done on variance estimates from the defined benefits products since these estimates are produced from a single panel. These single panel estimates seemed likely candidates for our simulation; however two issues prevented further work. The first issue was that only three panels existed that could be assembled to create a representative frame; this is due to different defined benefit data, collection methodology differences, and considerable durations of time between comparable panels. The second issue concerns the high degree of overlap between panels. The combinations of these two issues presented us with an insufficient frame to measure stability.

5. Summary

In this study, we have produced and analyzed estimated variances within National Compensation Survey Benefit products from the *Employee Benefits in Private Industry in the United States, 2005* publication. Applying a combination of these standard errors and relative standard errors in the form of a hybrid test, we were able to measure the reliability of domain estimates within these products. Attempts were also made to measure the stability of these variance estimates; unfortunately due to the lack of an adequate frame this measurement was not possible. With future research we hope to address the measurement of stability for variance estimates within the National Compensation Survey Benefits products more completely.

References

Dippo, C. S., Fay, R. E., Morganstein, D. H. (1984) Computing Variances from Complex Samples with Replicate Weights. 1984 Proceedings of the Section on Survey Research Methods, American Statistical Association, Washington, D.C., 489-494 “Fay’s Method for Variance Estimation”

Fay, R. E. (1984), Some Properties of Estimates of Variance Based on Replication Methods. Proceedings of the American Statistical Association Survey Research Methods Section. 495-500

Fay, R. E. (1989). Theory and Application of Replicate Weighting for Variance Calculations. Proceedings of the Section on Survey Research Methods, American Statistical Association, 212-217.

Guciado, C. J., Izsak, Y., (2008) Incidence and Provisions Reliability Research, Technical Report, U.S. Bureau of Labor Statistics

Izsak, Y., Ernst, L. R., Paben, S. P., Ponikowski, C .H. and Tehonica, J. (2003). Redesign of the National Compensation Survey. 2003 Proceedings of the American Statistical Association, Section on Survey Research Methods, [CD-ROM], Alexandria, VA: American Statistical Association.

Izsak Y, Ernst, L. R., McNulty E,. Paben, S. P., Ponikowski, C. H., Springer G., and Tehonica, J. (2005). Update on the Redesign of the National Compensation Survey. 2005 Proceedings of the American Statistical Association, Section on Survey Research Methods, [CD-ROM], Alexandria, VA: American Statistical Association

Judkins, D. (1990). Fay’s Method for Variance Estimation. Journal of Official Statistics, 6, 223-240. Comparison of Variance Estimation Methods for the National Compensation Survey

Kovar, J. G. (1985). Variance estimation of nonlinear statistics in stratified samples. Methodology Branch Working Paper N. 85-052E, Statistics Canada

Paben S. P. (1999) Comparison of Variance Estimation Methods for the National Compensation Survey, Proceedings of the Section on Survey Research Methods, American Statistical Association.

Rao, J.N.K., and Shao, J. (1996). On Balanced Half Sample Variance Estimation in Stratified Sampling. Journal of the American Statistical Association, 91, 343-348.

Rao, J.N.K., and Shao, J. (1999). Modified Balanced Repeated Replication for Complex Survey Data. Biometrika, 86, 403-415

U.S. Bureau of Labor Statistics (2008) BLS Handbook of Methods, National Compensation Measures, Ch. 8

U.S. Bureau of Labor Statistics (2007) National Compensation Survey: Employee Benefits in Private Industry in the United States, 2005. Washington D.C.

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

Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics