COMPARISON OF VARIANCE ESTIMATION METHODS FOR THE NATIONAL COMPENSATION SURVEY

Steven Paben, Bureau of Labor Statistics
BLS, 2 Massachusetts Ave., NE, Room 3160, Washington, DC 20212

Key Words: Ratio Estimation, Taylor Series, BRR, Jackknife, Fay's Method

INTRODUCTION

The National Compensation Survey (NCS) is a business establishment survey for occupational wages and benefits. This survey includes data on broad occupational classifications such as white-collar workers, major occupational groups (MOGs) such as sales, and individual occupations such as cashiers. Another feature of this survey is that it includes data by level of the job. The job level of an occupational series is derived from generic standards that apply to all occupations and occupational groups.

An essential part of NCS is the estimation of mean wages for different localities. In this study, an artificial Metropolitan Statistical Area (MSA) was created by combining NCS data from 16 different localities to serve as a sampling frame for 100 simulated samples. Then, the mean square error (MSE) for the 100 sample estimates was compared to variance estimates obtained with the linearized Taylor Series method of variance estimation and three different methods of replication: balance repeated replication (BRR), Fay's method, and a sample jackknife method.

SAMPLE DESIGN

NCS uses a rotating panel design with three stages of selection used in selecting each panel. The first stage of selection is of geographic area PSUs, which consist of both Metropolitan Statistical Areas (MSAs) and non-metropolitan counties. Much of the focus in NCS, including all the analysis in this paper, is on the production of locality estimates, that is estimates for individual MSAs for which the first stage of sampling is not an issue. Consequently, this stage of sampling is not addressed here. In the second stage of sampling, establishments are selected pps from industry strata, with total employment the measure of size. The sampling frame from which the establishments are selected is constructed from the unemployment insurance universe.

In the third stage of sampling, occupations are selected separately from each establishment. Typically, the occupational selections are done from a complete list of in scope employees for the establishment obtained from the respondent. (Certain cases of employees, such as those who set their own pay are out of scope). A systematic equal probability sample of employees is selected. Then, for each selected employee, wage data is obtained for all employees with the same detailed job as the selected employee within the particular establishment. For example, if one of the employees selected is a full time, grade 9, non-union accountant, whose earnings are time based (as opposed to incentive based), then data is collected for all employees satisfying these criteria for that establishment. Consequently, the equal probability selection of employees is equivalent to a pps selection of detailed jobs. The number of occupational selections in each establishment depends upon the size of the establishment.

The weight for each employee in a selected job is obtained by taking the product of the reciprocal of the probability of selecting the establishment, the reciprocal of the probability of selecting the job given that the establishment is selected, and nonresponse adjustment factors for establishment and occupational nonresponse.

METHODS

In order to compare variance estimation methods, we artificially created a "medium-sized" locality or MSA population from which we could draw simulated samples. The artificial MSA was created using 1997 NCS wage data from 16 different localities. Based on the sampling weights of a typical "medium-sized" MSA, the appropriate number of establishments were determined and created in each industry and size class. Since the establishments were originally selected using a pps design, we did not have enough data for workers from small establishments (generally establishments with less than 100 employees) to create all of the small establishments for the artificial population. On the other hand, we had an over abundance of data for workers from large establishments. Consequently, some workers were "borrowed" from the large establishments and separated into small establishments.

Then, within each establishment the appropriate number of occupations for each MOG x Level (MOGL) cell were also determined by using sampling weights from NCS. At the time the artificial MSA was created, we did not have enough data for some MOGLs. Consequently, workers were "borrowed" from MOGLs where there was an over abundance of workers, and the wages were transformed so that a worker's wage was typical of the MOGL to which the worker was moved. For more information regarding the creation of the artificial MSA, see Springer, Walker, Paben, and Dorfman (SWPD, 1999).

After the creation of the artificial MSA, 100 samples were drawn using the same sampling methodology as a typical NCS survey of "medium" size with the exception that there were no non-respondents. We determined that it was too burdensome to duplicate the NCS non-response adjustment procedures for each of the 100 samples for the believed small impact it would have on the results. Next, the variance for each of the domains; including all workers, MOGs, job levels, and MOGLs, were calculated using a linearized Taylor series method and using three different methods of replication. The variance results of each method were then compared to the true variance for the 100 simulated samples.

Taylor Series

The hourly mean wage for a particular domain of interest (i.e., MOG x level, occupation x level, etc.) is calculated as the ratio of the weighted total annual wages paid to the weighted total annual hours worked.

$$\hat{\bar{Y}}_D = \frac{\sum_i \hat{Y}_{iD}}{\sum_i \hat{X}_{iD}}, \quad (1)$$

where $\hat{Y}_{iD}$ is the total annual wages and $\hat{X}_{iD}$ denotes the total number of annual hours worked in domain *D* in industry stratum *i*. In order to calculate the weekly and annual mean wage, the denominator of (1) would simply have to be changed to the total number of annual weeks worked and the total number of employees, respectively.

One method of estimating the variance of a nonlinear estimator, such as the ratio in (1), is to approximate the estimator by a linear function of the observations using a first-order Taylor series expansion. Higher-order approximations are possible by extending and retaining the additional terms of the Taylor Series expansion. However, it has been shown for large, complex surveys that the first-order approximation usually yields satisfactory results. Then, variance formulae appropriate to the sampling design are applied to the linear approximation. This produces a biased, but typically consistent, estimator of the variance (Wolter, 1985).

The linearized variance formulae for NCS have two components, a non-certainty establishment component and a certainty establishment component. The component of variance for the noncertainty establishments is estimated by a pps with replacement formula reflecting the fact that the first stage of sampling in a PSU is a pps sample of establishments. The component of variance for the certainty establishments is estimated by a simple random sample with replacement formula, since the sample of occupations within an establishment is typically obtained through a systematic sample of employees and for certainty establishments the first stage of sampling in a PSU is the sample of occupations (Tehonica, Ernst, and Ponikowski, 1997). Since the subsampling is actually without replacement, both the non-certainty and certainty variance components of the linearized form generally should be overestimated.

The first-order Taylor series approximation for (1) within a constant is

$$\left(\frac{1}{\hat{X}_D}\right)\left(\sum_i \left[\left(\hat{Y}_{iDS} - \bar{Y}_D \hat{X}_{iDS}\right) + \left(\hat{Y}_{iDC} - \bar{Y}_D \hat{X}_{iDC}\right)\right]\right), \quad (2)$$

where $\hat{X}_D$ denotes the total number of annual hours worked in domain *D*, $E(\bar{Y}_D) = \hat{\bar{Y}}_D$, and the subscripts *S* and C respectively indicate the noncertainty and certainty portions of the variance.

From the above Taylor series approximation, it follows that the estimated total variance for noncertainty establishments is

$$\hat{V}(\hat{\bar{Y}}_D) = \sum_i \frac{n_i}{n_i - 1}\left[\sum_j \left(\hat{Z}_{ijD}\right)^2 - \frac{1}{n_i}\left(\sum_j \hat{Z}_{ijD}\right)^2\right], \quad (3)$$

where the summation of $\hat{Z}_{ijD}$ is over all non-certainty establishments *j*, and where

$$\hat{Z}_{ijD} = \left(\frac{1}{\hat{X}_D}\right)\left[W_{ij}\left(\hat{Y}_{ijD} - \hat{\bar{Y}}_D \hat{X}_{ijD}\right)\right]. \quad (4)$$

In equations (3) and (4), n_i is the number of non-certainty establishments in stratum *i*, $\hat{Y}_{ijD}$ is an estimate of total annual wages in domain *D* for the j^{th}

establishment in the i^{th} stratum, $\hat{X}_{ijD}$ is an estimate of the total annual hours worked (for hourly mean wage estimates) in domain D for establishment ij, and W_{ij} is the sampling weight for establishment ij.

From (2), it also follows that the variance of the certainty establishments is equation (5)

$$\hat{V}(\hat{\bar{Y}}_D) = \sum_{i,j} \frac{m_{ij}}{m_{ij}-1}\left[\sum_{q\in D} W_{ijq}\left(\hat{Z}_{ijqD}\right)^2 - \frac{1}{m_{ij}}\left(\sum_{q\in D} W_{ijq}\hat{Z}_{ijqD}\right)^2\right]$$

where the summation of Z_{ijqD} is over all certainty establishments j, and where

$$\hat{Z}_{ijqD} = \left(\frac{1}{\hat{X}_D}\right)\left[\left(Y_{ijqD} - \hat{\bar{Y}}_D X_{ijqD}\right)\right]. \qquad (6)$$

In equations (5) and (6), m_{ij} is the number of quotes for the j^{th} certainty establishment in the i^{th} stratum, W_{ijq} is the product of the sampling weight for establishment ij and the sampling weight for quote ijq, Y_{ijqD} is the total annual wages for all employees in domain D for the q^{th} quote in certainty establishment ij, X_{ijqD} is the total annual hours worked for all employees (for hourly mean wage estimates) in domain D for the q^{th} quote in certainty establishment ij.

Replication

Another method for estimating the variance of a nonlinear estimator, such as the ratio in (1), is replication. Like the Taylor linearization method, replication methods generally produce a biased, but consistent estimator of the variance for nonlinear estimators (Wolter, 1985). The basic theory behind replication is to calculate the estimate of interest from the full sample as well as a number of subsamples. The variation among the subsample estimates is used to estimate the variance for the full sample. One advantage replication has over the variance approach in the previous section is that there is usually no need to linearize a nonlinear estimator before calculating its variance. This has been demonstrated many times empirically over the years, and was first shown to hold asymptotically as the number of strata increases by Krewski and Rao (1981).

There are many different ways of creating the subsamples in replication. One approach is balanced repeated replication (BRR). The standard BRR design assumes that a population of PSUs are able to be grouped into G strata with two PSUs selected from each stratum using with replacement sampling. Then, h replicate half-sample estimates are formed by selecting one of the two PSUs from each stratum based on a Hadamard matrix and then using only the selected PSU to estimate the parameter of interest. The weights for the selected units are doubled to form the weights for the replicate estimate. In order to obtain a balanced set of replicates the number of replicates used needs to be a multiple of four greater than or equal to the number of strata.

Since BRR requires two PSUs per stratum and the NCS design has more than two PSUs, the two PSUs were artifically created by assigning the design-based PSUs to one of two variance PSUs. For the non-certainty establishments, the first stage of selection is a pps sample of establishments within industry strata with employment as the measure of size. For the certainty establishments, the first stage of selection is equivalent to a pps sample of detailed jobs or quotes. Therefore, the variance PSUs for the noncertainty establishments are created at the establishment level, while the variance PSUs for the certainty establishments are formed at the quote level.

The variance estimator for $\hat{\bar{Y}}_D$ using BRR is then

$$V(\hat{\bar{Y}}_D) = c\sum_{h=1}^{G}(\hat{\bar{Y}}_{D(h)} - \hat{\bar{Y}}_D)^2\,, \qquad (7)$$

where $\hat{\bar{Y}}_D$ is the estimate of hourly mean wage for domain D based on the full sample, $\hat{\bar{Y}}_{D(h)}$ is the estimate of hourly mean wage for domain D based on the h-th replicate half-sample, G is the number of replicates and $c = 1/G$.

Another method of replication investigated in this study was Fay's method. Fay's method was motivated by the observation that the standard half-sample variance estimator runs into difficulty when the denominators are zero for some replicates (Judkins, 1990). This method is a variant of BRR, where the basic idea is to modify the sample weights less than in BRR by using both half-samples in each replicate. In each replicate, one half of the sample is weighted down by a factor K and the remaining half is weighted up by a compensating of factor of 2 - K. For example, if K = .70, then the weights decrease by 30 percent in one half-sample and increase in the other half-sample by 30 percent. When using Fay's method, the variance of the replicates from the full sample estimate becomes too small by a factor of $(1 - K)^2$ (Judkins, 1990). Therefore, the constant c in (8) becomes $1/G(1- K)^2$. In this study, Fay's method was used with $K = 0.5$.

The final method of replication investigated in this study is a sample jackknife method. In general, the jackknife method consists of splitting the total sample into G disjoint and exhaustive PSUs, then dropping out a specified number of PSUs in turn, and estimating the parameter of interest from the remaining units each time. The variability among these estimates is then used to estimate the variance of the full-sample estimator. A sample jackknife method consists of dropping out only a sample of the PSUs from each replicate.

A "general" sample jackknife variance estimator for an estimator, $\hat{\mu}$, is

$$V(\hat{\mu}) = \sum_{h=1}^{H} \frac{(l_h - 1)}{g_h} \sum_{i=1}^{g_h} (\hat{\mu}_{(ih)} - \hat{\mu})^2, \qquad (8)$$

where $\hat{\mu}_{(ih)}$ is the estimate of $\hat{\mu}$ with the i^{th} subsample dropped out in stratum h, H is the number of strata, l_h is the number of subsamples in stratum h, and g_h is the number of subsamples dropped out in stratum h, giving $G = \sum_{h=1}^{H} g_h$ replicates.

The particular sample jackknife approach used in this study is the second method given in WesVarPC®. This method uses the same sampling design as BRR, that is, two variance PSUs per stratum made with replacement. Next, the weights for one variance PSU in a stratum are doubled, while the other variance PSU in that stratum drops out. The other weights for the remaining strata remain unchanged. This process is done separately for each stratum, and the number of replicates in this method equals the number of strata. Therefore, the variance of $\hat{\bar{Y}}_D$ for this method of the jackknife with $l_h = 2, g_h = 1$, and $H = G$ in (8) is of the same form as (7) with $c = 1$.

<u>ANALYSIS</u>

Evaluative statistics were calculated to compare the different variance estimation methods for the variance of average hourly wage in each domain D, that is each MOG, level, and MOGL. In order to simplify the explanation of the evaluative statistics, the following definitions are given:

(A) The mean square error, the estimated variance plus the square of the bias, for $\bar{Y}$ in domain D is

$$M(\bar{Y}_D) = \sum_{r=1}^{R} (\hat{\bar{Y}}_{rD} - \hat{\bar{Y}}_D)^2 / R + (\hat{\bar{Y}}_D - \bar{Y}_{pD})^2 \ , \text{ where}$$

R = 100 simulated samples, $\hat{\bar{Y}}_{rD}$ is the average hourly wage for the r-th simulated sample, $\hat{\bar{Y}}_D$ is the average hourly wage for all 100 samples, and $\bar{Y}_{pD}$ is the average hourly wage of the population.

(B) The average standard error for $\hat{\bar{Y}}_D$ is

$$\overline{SE} = \sqrt{\sum_{r=1}^{R} \hat{V}_r(\hat{\bar{Y}}_D) / R} \ , \text{ where}$$

$\hat{V}_r(\hat{\bar{Y}}_D)$ is the estimated variance of $\hat{\bar{Y}}_D$ calculated using one of the four different variance estimation methods.

(C) The mean square error of the estimated variances from $M(\bar{Y}_D)$ is

$$M(\hat{V}_r, M(\bar{Y}_D)) = \sum_{r=1}^{R} (\hat{V}_r - M(\bar{Y}_D))^2 / R \ .$$

The two evaluative statistics calculated were the "bias" and "stability". The bias, in this case, was a relative measure. It was defined to be the average standard error defined in (B) minus the square root of the MSE defined in (A) relative to the square root of the MSE. The stability was defined to be the ratio of the square root of (C) to (A). This ratio was calculated relative to (A) as a means of standardization. For stability, a smaller ratio implies a more stable variance estimate.

The bias of each variance estimation method by MOG (Table 1A) shows no dramatic differences between the variance methods. These estimates vary more across MOGs than they do across methods except for machine operators, which is underestimated more in the replication methods than in the Taylor linearization method. The bias of each variance estimation method by job level (Table 1B) shows the lower levels (1-6) which have lower SEs tend to overestimate the root MSE, while the higher levels (13-15) with greater root MSEs tend to be underestimated by the Taylor linearization method and Fay's method, overestimated by BRR, while the Jackknife is fairly close to one.

One possible explanation for the Taylor linearization method tending to underestimate the variance for the higher levels is that there are a smaller number of observations generating the estimate (as can be seen by the average number of occupational selections for the 100 samples, *mean occs* in Table 1B). When the sample size is a large, a first-order Taylor series approximation yields satisfactory results. However, for small sample sizes or for less prevalent domain estimates, omitting the higher-order terms could cause the variance to be underestimated. BRR may tend to overestimate the variance for domain estimates with a small number of observations, because of large weight differences between replicates or because some replicates may possibly be zero. However, Fay's method and the jackknife should not be as greatly affected by a small number of observations, because the perturbation of the weights is much less in those methods than in BRR. Although, Fay's method underestimated the root MSE almost as much as BRR overestimated the root MSE for levels 13 and 15. In general, it can be shown analytically that variance estimates for ratio estimators calculated using Fay's method are a decreasing function of K, and hence maximize when K=0, which is BRR.

Table 1A. Bias and Stability of Variance Estimation Methods by Major Occupational Group

	Root	Taylor Series		BRR		Fay's Method		Jackknife	
MOG	MSE	bias	stability	bias	stability	bias	stability	bias	stability
Professional	0.809	-0.04	0.43	-0.07	0.71	-0.07	0.71	-0.07	0.72
Technical	1.067	-0.18	1.00	-0.13	1.14	-0.16	1.07	-0.16	1.05
Exec., Admin., Mgr.	1.246	0.06	0.72	0.10	0.82	0.07	0.77	0.08	0.78
Sales	2.014	0.00	2.17	0.02	2.33	0.00	2.24	0.01	2.38
Admin. Support	0.330	-0.10	0.29	-0.06	0.41	-0.07	0.40	-0.07	0.40
Precision, Production	0.758	0.00	0.54	0.04	0.68	0.02	0.65	0.02	0.66
Machine Operators	0.613	-0.11	0.34	-0.24	0.61	-0.25	0.61	-0.24	0.63
Transportation	0.877	-0.01	0.61	0.01	0.82	-0.03	0.76	-0.02	0.81
Handlers, Laborers	0.527	0.07	0.47	0.08	0.75	0.06	0.71	0.06	0.74
Service	0.630	-0.01	0.28	-0.01	1.05	-0.01	1.04	-0.02	1.00
All Workers	0.340	-0.03	0.26	-0.03	0.57	-0.03	0.57	-0.03	0.57

Table 1B. Bias and Stability of Variance Estimation Methods by Job Level

	Mean	Root	Taylor Series		BRR		Fay's Method		Jackknife	
Level	Occs	MSE	bias	stability	bias	stability	bias	stability	bias	stability
1	207.0	0.184	0.16	0.63	0.20	0.99	0.18	0.93	0.18	0.93
2	212.2	0.301	0.03	0.57	0.06	0.74	0.04	0.71	0.04	0.69
3	327.0	0.278	0.05	0.29	0.01	0.53	0.00	0.51	0.00	0.52
4	290.8	0.283	0.31	0.83	0.29	1.07	0.28	1.03	0.27	1.03
5	217.9	0.349	0.11	0.54	0.09	0.60	0.08	0.58	0.07	0.57
6	134.8	0.602	0.02	0.63	0.09	0.92	0.07	0.88	0.07	0.96
7	251.2	0.697	-0.07	1.69	-0.05	1.72	-0.06	1.71	-0.06	1.66
8	214.4	0.923	-0.03	0.32	-0.05	1.19	-0.06	1.18	-0.07	1.14
9	217.6	0.786	0.01	0.76	0.07	1.08	0.04	1.01	0.09	1.24
10	53.9	2.188	-0.17	1.07	-0.06	1.53	-0.13	1.26	-0.07	2.00
11	105.4	1.815	-0.02	1.15	0.02	1.58	-0.01	1.30	0.00	1.53
12	69.7	1.657	-0.12	0.64	-0.10	0.67	-0.13	0.65	-0.11	0.76
13	32.8	1.885	-0.04	0.48	0.04	0.85	-0.04	0.72	-0.02	0.77
14	21.1	2.933	-0.11	0.64	0.13	1.29	-0.01	0.97	0.01	0.99
15	5.1	11.863	-0.21	1.62	0.18	2.67	-0.16	1.77	-0.07	1.71
JNL	48.6	4.159	0.03	0.60	0.07	0.82	0.03	0.75	0.02	0.79

*JNL = Job not able to be leveled

Table 2. Average Bias and Stability of Variance Estimation Methods by Domain

	No. of	Mean	Root	Taylor Series		BRR		Fay's Method		Jackknife	
Domain	Estimates	Occs	MSE	bias	stability	bias	stability	bias	stability	bias	stability
MOGs	10	240.9	0.792	-0.04	0.56	-0.03	0.84	-0.05	0.81	-0.05	0.82
Levels	16	150.6	0.988	-0.01	0.74	0.06	1.10	0.00	0.99	0.02	1.06
MOGLs	76	30.7	1.282	-0.11	0.76	0.07	1.21	-0.09	0.91	-0.01	1.25

The stability of the variance estimation methods by MOG (Table 1A) and by job level (Table 1B) suggests that the Taylor linearization method is more stable than the replication methods. This is consistent with the work of Kish and Frankel (1974). There are no dramatic differences in relative stability among the replication methods, although the stability of Fay's method was always less than or equal to the stability of BRR. It is unclear, however, if this must always be the case.

The average bias and stability of each variance estimation method (Table 2) for the 10 MOGs, 16 job levels, and 76 MOGLs (the MOGLs present in all 100 samples) summarizes the observed trends of Tables 1A and 1B. The average of the two ratios for the different domains was calculated as a geometric mean. The Taylor linearization method tends to underestimate the variance for less populous domains such as the MOGLs, while BRR tends to overestimate the variance for the less populous domains. Fay's method tended to underestimate the variance for the MOGLs. The jackknife method seemed to be the most consistent in terms of relative bias. As for the stability of the variance estimation methods, the Taylor linearization method was the most stable for all methods, while Fay's method was the most stable of the replication methods.

CONCLUSION

The estimates of most interest from NCS are locality estimates of mean wages for all workers, MOGs, job levels, and MOGLs. The basic idea of this study was to compare different variance estimation methods for a ratio estimator with many domains and sub-domains. The results showed that for prevalent domains there is little difference in terms of bias for the Taylor linearization method, BRR, Fay's method, and the jackknife method. This is reasonable, since all of the methods are valid asymptotically. For less prevalent domains, the jackknife method seemed to be the closest to the true value of the variance, while the Taylor linearization method tended to underestimate the variance and BRR overestimate the variance. The Taylor linearization method produced the most stable estimates, while Fay's method was the most stable of the replication methods.

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

REFERENCES

Brick, J. M., Broene, P., James, P., and Severynse, J. (1997). A User's Guide to WesVarPC®. Version 2.1. Westat, Inc.

Cochran, W.G. (1977). Sampling Techniques. Third Edition. New York: Wiley.

Ernst, L.R. and Williams, T.R. (1987). Some Aspects of Estimating Variances by Half-Sample Replication in CPS. Bureau of the Census Statistical Research Division Report Series, Report Number: Census/SRD/RR-87/16.

Judkins, D. R. (1990). Fay's Method for Variance Estimation. Journal of Official Statistics, 6, 223-239.

Kish, L. and Frankel, M.R. (1974). Inference from Complex Samples. Journal of the Royal Statistical Society, Series B, 36, 1-37.

Krewski, D. and Rao, J.N.K. (1981). Inference from Stratified Samples: Properties of the Linearization, Jackknife, and Balanced Repeated Replication Methods, Annals of Statistics, 9(5), 1010-1019.

Rao, J.N.K. and Wu, C.F.J. (1985). Inference from Sample Surveys: Second Order Analysis of Three Methods for Nonlinear Statistics. Journal of the American Statistical Association, 80, 620-630.

Rust, K. (1985). Variance Estimation for Complex Estimators in Sample Surveys. Journal of Official Statistics, 1, 381-397.

Springer, G., Walker, M., Paben,S., Dorfman, A. (1999). Evaluation of Confidence Interval Estimation Methods for the National Compensation Survey. Proceedings of the American Statistical Association, Section on Survey Methods Research Methods, forthcoming.

Tehonica, J., Ernst, L. R., and Ponikowski, C. H. (1997), Summary of Estimation and Variance Specifications for the NCS. Bureau of Labor Statistics memorandum to The Record, dated July 31.

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