Comparison of Variance Estimates
Using Random Group and Taylor Series Methods in MEPS-IC

Sadeq Chowdhury

The variances of MEPS-IC estimates were computed using the random group (RG) variance estimation method through 2013. Since the RG method is found to be less reliable for some estimates, particularly those based on smaller sample sizes where formation of random groups is not very stable, the variance estimation method in the MEPS-IC was changed to the Taylor series (TS) linearization method starting with 2014 survey. The TS method is also used for variance estimation in the MEPS Household Component (HC). To give an idea about the extent of differences in variance estimates, this document presents a comparison of variances (in terms of relative standard errors) computed using RG and TS methods for a range of MEPS-IC estimates.

A general discussion of RG and TS methods of variance estimation can be found in Wolter (1985). A discussion of the TS method as used in the MEPS-HC can be found in Chowdhury (2013). Also, a brief overview of the two methods as used in the MEPS-IC is given below.

Random Group Method

Under the RG method, as each establishment is selected into the sample after sorting by key characteristics, a number is sequentially assigned such that the final sample can be easily divided into ten "random" groups. The choice of using ten random groups was made because that number of random groups is commonly used for this method.

The random group estimator for the variance of an estimate, $\hat{\theta}$, is then computed as

$$var\,(\hat{\theta}) = \frac{1}{k(k-1)} \sum_{\alpha}^{k} \left(\hat{\theta}_{\alpha} - \hat{\theta}\right)^2$$

where k is the number of random groups, $\hat{\theta}$ is the estimate based on the entire sample, and $\hat{\theta}_{\alpha}$ is the estimate based on the establishments in random group α.

Taylor Series Method

Under the TS method, standard variance estimation formulae available for linear estimators are used for all estimators. For nonlinear estimators, the linear approximation is obtained by using a first-order Taylor series expansion.

For the i-th establishment in stratum h, if x_{hi} is the value of a target variable, w_{hi} is the estimation weight (which can just be the inverse of the selection probability in the absence of any nonresponse or other adjustment), $\theta_{hi} = w_{hi}x_{hi}$ is the corresponding weighted value, and n_h is the sample size in the stratum then the variance of an estimator of total,

$$\hat{\theta} = \sum_{h=1}^{H} \hat{\theta}_h = \sum_{h=1}^{H} \sum_{i=1}^{n_h} w_{hi}x_{hi} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \theta_{hi} \;,$$

is obtained under the TS method as

$$var(\hat{\theta}) = \sum_{h=1}^{H} \frac{n_h}{(n_h - 1)} \sum_{i=1}^{n_h} (\theta_{hi} - \bar{\theta}_h)^2 \quad \text{with} \quad \bar{\theta}_h = \frac{\sum_{i=1}^{n_h} \theta_{hi}}{n_h}.$$

The formulas for variance estimation using the TS method for different nonlinear estimates produced from the MEPS-IC can be found in SAS Stat User's Guide (2012) or SUDAAN Technical Manual (1996).

The standard error (SE) and the relative standard error (RSE) of the estimate $\hat{\theta}$ is defined as

$$SE(\hat{\theta}) = \sqrt{var\,(\hat{\theta})} \quad \text{and} \quad RSE(\hat{\theta}) = SE/\hat{\theta}$$

and RSEs are often expressed in percentages i.e., by multiplying with 100.

Comparison of Variance Estimates

The variances computed using RG and TS methods are compared for a wide range of MEPS-IC estimates using 2013 data. The estimates are compared only for private sector estimates as the variances of estimates in government tables are not expected to differ much between RG and TS methods due to the large number of cases selected with certainty. The estimates are separately compared for totals of continuous variables (such as premiums, contributions, enrollments, etc.) and percentages for categorical variables (such as offer rates, take-up rates, eligibility rates, etc.) both at the U.S. and State levels by firm size, industry group, age of firm, ownership, low wage, union presence and multi/single status. Variance estimates of about 38,000 estimates of totals and about 68,000 estimates of percentages are compared. The RSEs of all estimates using both RG and TS methods are produced and the percentage point difference in RSE (i.e., Diff RSE% = RG RSE% - TS RSE%) is computed for each estimate.

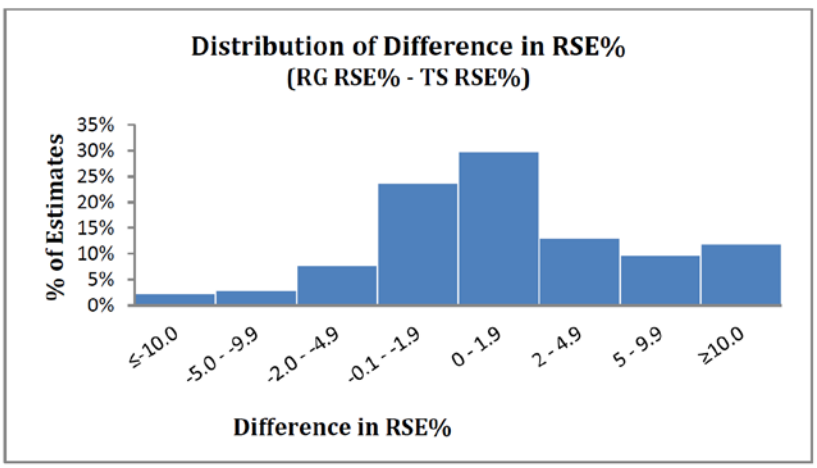
Difference in RSE Estimates of Totals

Table 1 shows the distribution of percentage point difference in estimated RSEs between RG and TS methods and Figure 1 shows the corresponding histogram of such differences for estimates of total. Table 1 shows that the difference is less than ±2 percentage points for 53% of the estimates and less than ±5 percentage points for 74% of the estimates. Sample size appears to be a major factor for larger differences in RSEs. The SE estimate under the RG method is less stable when the sample size is smaller. If the estimates which are based on a sample of size 50 or less are excluded from the comparison then for 69% of the estimates, the difference in RSEs is less than ±2 percentage points and for 91% of the estimates, the difference in RSEs is less than ±5 percentage points. It can also be seen from Table 1 that for the majority of estimates (64%), the RSE estimate under the RG method is higher than the RSE estimate under the TS method.

Table 1: Distribution of differences in RSEs of estimates of totals computed using RG and TS methods

Difference in estimated RSE% (RG-TS)	Number of estimates	Percent of estimates
≤-10.0	817	2.1%
-5.0 – -9.9	1,044	2.7%
-2.0 – -4.9	2,894	7.6%
-0.1 – -1.9	8,973	23.6%
0 – 1.9	11,296	29.7%
2 – 4.9	4,907	12.9%
5 – 9.9	3,645	9.6%
≥10.0	4,482	11.8%
Total	38,058	100.0

Figure 1. Histogram of difference in RSEs of estimates of totals computed using RG and TS methods



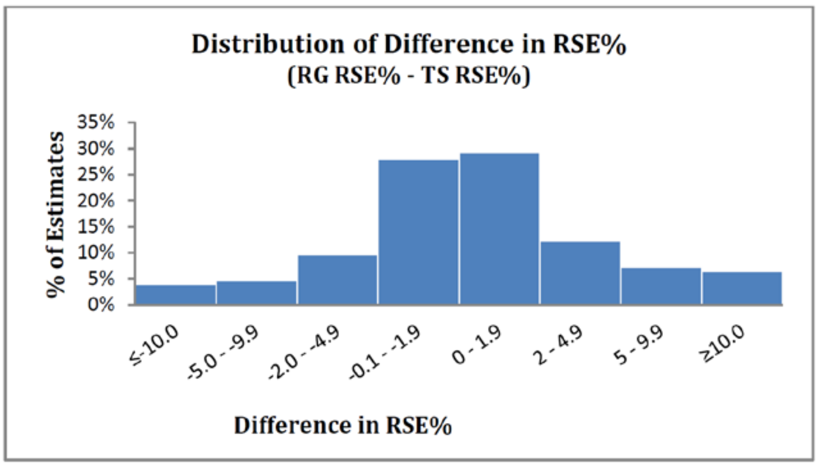
Difference in RSE Estimates of Percentages

Table 2 presents the distribution and Figure 2 presents the corresponding histogram of the differences in estimated RSEs for about 68,000 percentage estimates. For about 57% of the estimates, the difference in RSEs is less than ±2 percentage points, and for about 78% estimates, the difference is less than ±5 percentage points. If the estimates based on a sample size of 50 or less are excluded then for 63.5% of the estimates, the RSE difference is less than ±2 percentage points and for 85.2% of the estimates, the RSE difference is less than ±5 percentage points. For 55% of all estimates, the RSE estimate under the RG method is higher than the RSE estimate under the TS method.

Table 2: Distribution of differences in RSE of estimates of percentages computed using RG and TS methods

Difference in estimated RSE% (RG-TS)	Number of estimates	Percent of estimates
≤-10.0	2,553	3.8%
-5.0 – -9.9	3,069	4.5%
-2.0 – -4.9	6,471	9.5%
-0.1 – -1.9	18,895	27.8%
0 – 1.9	19,775	29.1%
2 – 4.9	8,256	12.1%
5 – 9.9	4,809	7.1%
≥10.0	4,237	6.2%
Total	68,065	100.0

Figure 2. Histogram of difference in RSEs of estimates of percentages computed using RG and TS



References

Chowdhury, S.R. (2013). A Comparison of Taylor Linearization and Balanced Repeated Replication Methods for Variance Estimation in Medical Expenditure Panel Survey. Agency for Healthcare Research and Quality Working Paper No. 13004, July 2013, http://meps.ahrq.gov/data_files/publications/workingpapers/wp_13004.pdf

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