A COMPARISON OF VARIANCE ESTIMATORS FOR THE ADVANCE MONTHLY RETAIL TRADE SURVEY[1]

Brett Moore, Jock Black, Xijian Liu, U.S. Bureau of the Census
Brett Moore, U.S. Bureau of the Census, SSSD, FOB 2754-3, Washington, DC 20233

Keywords: Replication Variance Estimation, Random Groups, Delete-a-Group Jackknife, Monte Carlo Simulation

1. Introduction

This paper compares the random groups method with the delete-a-group jackknife method for estimating the variance of a ratio estimate of consecutive month's sales (M_t/M_{t-1}) from the Advance Monthly Retail Trade Survey (MARTS). For both of these methods, three alternatives are tested: 1) the basic method, 2) a new procedure that attempts to account for additional variation due to non-response from units in the certainty stratum, and 3) a combination of the first two methods. Within each of these six approaches, we investigate the effects of varying the number of random groups. The particular features of the MARTS survey that are addressed are small sample sizes, ratio estimation, and non-response. Methods are compared via a Monte Carlo simulation used to estimate the mean squared error and confidence interval coverage. Additionally, a comparison of the three random groups methods based on historical survey data is presented.

2. Methodology Background

The U.S. Census Bureau conducts two surveys that measure monthly retail sales – the MARTS and the Monthly Retail Trade Survey (MRTS). The MARTS is a quick-turnaround survey that gives an early indication of sales for the most recent month. Costs are incurred both in response and processing to achieve this quickness. Many units cannot respond in time to be included in the survey, only simple edits are performed, and no imputation is done. Thus the MARTS estimates can have substantial sampling and nonsampling error. To remedy these drawbacks, the Bureau conducts the larger MRTS, which has a longer data collection period, extensive followup, thorough editing, and imputation for missing and remaining erroneous data. The MRTS estimates revise the MARTS estimates and are published one month later.

The MARTS sample is a subsample of the MRTS. Transferring MARTS responses to the MRTS alleviates response burden. The MRTS is a single-stage stratified random sample of business units selected without replacement. Kind of business and estimated annual sales define the strata. For the MARTS, the selected MRTS units are grouped into the slightly broader kind-of-business and sales strata and a stratified systematic probability proportional to estimated size (ppes) sample is selected. The estimated size measure used is the reciprocal of the unit's MRTS selection probability. Thus, almost all units in a given MARTS stratum will have an equivalent two-stage selection probability. Units selected with certainty are assigned to a certainty random group (random group g0). Noncertainty units are assigned to one of sixteen nonindependent random groups in a systematic manner.

For MARTS, when estimating the ratio of consecutive month's sales for the most detailed kinds of business, only units that respond in both months are used. Response data is weighted by the reciprocal of the (two-stage) selection probability and summed separately. The ratio of the two monthly sums is then computed. To estimate sales levels, MARTS uses a link-relative estimate that multiplies the previous month's MRTS sales level by the MARTS ratio. For the MRTS, nonresponding units are imputed so that results account for the entire survey universe and M_t/M_{t-1} is estimated as the ratio of consecutive Horvitz-Thomson estimates of sales levels. For broader kinds of businesses in MARTS, levels are estimated as sums of the link-relative estimated detailed levels and month-to-month ratios are computed as the ratio of these levels.

The method of random groups is used for estimating variances in MARTS. In most of the non-certainty strata the sampling fractions (f) can be considered negligible. Therefore, sampling-with-replacement methods can be used (i.e. a finite population correction factor (fpc) is not needed).

3. Description of the Study

Three variations each of two common techniques (see e.g., Wolter 1985) of variance estimation are compared (six estimators total). The variance estimation techniques are random groups and delete-a-group jackknife (hereafter referred to as jackknife). We are looking at modifying the basic techniques to account for additional variation caused by nonresponse in the certainty stratum. A second goal of this investigation is

[1]This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

to examine the relative stability of the variance estimates. The chart below displays the point estimates of the standard error of the month-to-month change from the start of the current sample.

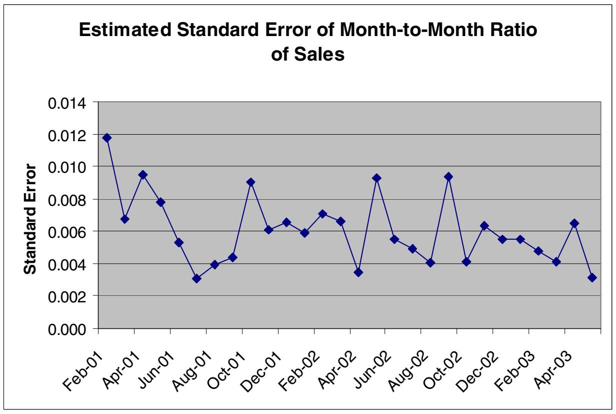


Chart 1

Because only about half of the certainty units respond and no imputation is done to account for the nonresponding certainties, the certainty stratum is effectively contributing to sampling variability (assuming the non-response mechanism is missing at random (MAR)) and has a sampling fraction equal to the unit response rate of the certainties, about 0.5. It is reasonable to try to account for this when estimating variances. As described below, our current variance estimation method (method A) does not account for sampling variance in the certainty stratum.

The variations are: Method A, the current method, which includes a certainty stratum that does not contribute to estimates of variance; Method B, which treats the certainty stratum as though it were sampled with replacement; and Method C, which uses a combination of Methods A and B and adjusts the certainty variance estimate using a finite population correction based on the certainty response rate. The estimators are labeled RA, RB, RC, JA, JB, JC (R = Random Groups, J = Jackknife, {A, B, C} = Method A, B, C).

The comparisons are based on two criteria: mean squared error (MSE) and confidence interval (CI) coverage. MSE is commonly used for comparing estimators; it combines the variance and bias in one number. Confidence interval coverage is used as well, since most measures of variability associated with retail estimates are used primarily for the construction of CIs or statements of statistical significance, which are related.

Within these comparisons, we additionally consider the effect of varying the number of random groups. Ideally, random groups should be assigned so that all strata are represented and balanced in each random group. This is not feasible in the MARTS survey since target stratum sample sizes are often as low as two, and attained stratum sample sizes are occasionally zero due to nonresponse. In addition, units are assigned to random groups nonindependently. It seems reasonable to expect that with these small sample sizes, fewer random groups would produce more accurate estimates because each random group would more closely mimic the sample design.

Because of the complex sample design and estimation methods used for MARTS, it is not possible to give an exact design-based expression for the variance of the estimator. Consequently, it was decided that a simulation would be the best way to compare the variance estimators. Our first step was to create a simulated MARTS universe. This was used to make comparisons among the estimators. Its design is based on modeled MARTS and MRTS data.

The simulation study consists of repeatedly drawing random samples from the simulated universe, simulating nonresponse, and forming estimates of the month-to-month ratio. The arithmetic average of these values is the target estimate. The standard error of these values is the target for the variance estimators examined here. To generate the targets, we used 10,000 runs. The various estimators are compared using 1,000 runs for each method.

A second comparison is made using actual MARTS data. From past MARTS estimates, an estimate of the MSE of the estimate can be made based on the revision of MARTS estimates when the larger MRTS sample is collected. The estimates of MSE obtained using this method are then used as a basis for comparing Methods A, B, and C.

4. Formulas

Methods A and B estimate variances using replicates. They differ in how the certainty units are assigned to random groups.

4.1. Replicate Estimates

Let x_i = Current Month (CM) sales for unit i, y_i = Prior Month (PM) sales for unit i, w_i = sampling weight for unit i, G = number of replicates (random groups), g = group (g = 1…G).

The certainty stratum is denoted by g0. For Method B, all units (including those assigned to g0 for Method A) are assigned systematically to each random group g, g = 1…G.

4.1.1. Method A Random Groups (RA)

$$rep\ (R, A, g) = \frac{G\sum_{i\in g} w_i x_i + \sum_{i\in g0} x_i}{G\sum_{i\in g} w_i y_i + \sum_{i\in g0} y_i}$$

4.1.2. Method A Jackknife (JA)

$$rep(J, A, g) = G\frac{\sum_{i\in g} w_i x_i}{\sum_{i\in g} w_i y_i} - (G-1)\frac{\frac{G}{G-1}\sum_{i\notin g} w_i x_i + \sum_{i\in g0} x_i}{\frac{G}{G-1}\sum_{i\notin g} w_i y_i + \sum_{i\in g0} y_i}$$

4.1.3. Method B Random Groups (RB)

$$rep\ (R, B, g) = \frac{\sum_{i\in g} w_i x_i}{\sum_{i\in g} w_i y_i}$$

4.1.4. Method B Jackknife (JB)

$$rep(J, B, g) = G\frac{\sum_{i\in g} w_i x_i}{\sum_{i\in g} w_i y_i} - (G-1)\frac{\sum_{i\notin g} w_i x_i}{\sum_{i\notin g} w_i y_i}$$

4.2. Variance Estimates

4.2.1. Methods A and B (JA, JB, RA, RB)

$$V\hat{a}r_{\alpha\beta} = \frac{\sum_{g=1}^{G}(rep\ (\alpha, \beta, g) - \overline{rep\ (\alpha, \beta, g)})^2}{G(G-1)}$$

$\alpha\in\{R,J\}$, $\beta\in\{A,B\}$

Method C uses the variance estimates of methods A and B, and does not require computing replicate estimates.

4.2.2. Method C

$$V\hat{a}r_{\alpha C} = V\hat{a}r_{\alpha A} + (1-r)(V\hat{a}r_{\alpha B} - V\hat{a}r_{\alpha A})$$

The term $(V\hat{a}r_{B\beta} - V\hat{a}r_{A\beta})$ in Method C gives an estimate of the additional variance due to nonresponse within the certainty stratum. The term (1-r) is a finite population correction based on the response rate (r) in the certainty stratum.

5. Building the Test Frame

A simulated MARTS frame needed the following variables that are used in the MARTS survey:

- A Measure Of Size (MOS) that estimates annual sales dollar volume,
- Prior Month (PM) sales,
- Current Month (CM) sales.

MOS is used for creating strata and assigning units to strata, CM and PM are used for estimating the month-to-month ratio (M_t/M_{t-1}). Only MOS is available for all units on the real frame.

The simulated frame should also have the following key features typical for MARTS kinds-of-businesses:

- Heavily skewed distribution with many smaller units and relatively few large units.
- Neyman sample allocation: sample size is proportional to stratum size and standard deviation.
- Large certainty stratum (in terms of dollar volume): approximately one half of total sales.

Three sample sizes were used: n = 22, 34, and 70 (small, medium, and large).

The variables MOS and PM sales were available for units in the MRTS sample, and CM and PM sales were available for MARTS respondents. Since real MRTS and MARTS data could only be obtained for the relatively few businesses participating in those surveys, a large number of observations had to be simulated. A frame-size set of MOS values was made by extrapolation from the MOS values in the MRTS survey. Monthly data was then derived as follows:

1. PM sales values were simulated using the model $\log(PM)=\alpha_1+\log(MOS)+\varepsilon_1$ where $\varepsilon_1\sim N(0, \sigma_1)$. Estimates of α_1 and σ_1 were made using MRTS survey data.

2. CM sales were simulated with the model $\log(CM)=\alpha_2+\log(PM)+\varepsilon_2$ where $\varepsilon_2\sim N(0, \sigma_2)$. The parameters α_2 and σ_2 were estimated using data from MARTS respondents.

The final frame size was 18,036 units.

6. Comparison of Variance Estimators

6.1. Simulation Comparison

The variance estimators were first compared by Monte Carlo simulation. 10,000 repetitions were used to determine the target value. 1,000 runs were then used to compare estimates of the standard error of the month-to-month ratio for each method with 4, 8, 12, and 16 random groups.

Non-response was modeled by randomly excluding selected units at a rate of 50% from all strata. This rate was chosen based on overall response rates for MARTS.

Three different sample sizes were used. The target standard errors for each sample size are listed below in Table 1.

Sample Size	N	M_t / M_{t-1}	SE of estimated M_t / M_{t-1}
Small	22	.95	.159
Medium	34	.95	.118
Large	70	.95	.062

Table 1

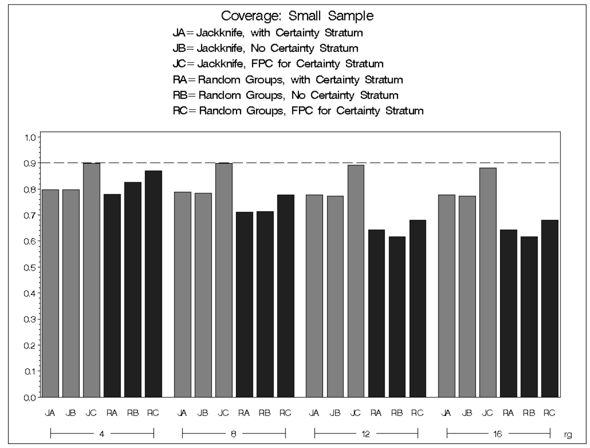
The relative performance of the variance estimators were compared by the coverage percentage of confidence intervals based on the estimates and by their root mean squared error (RMSE).

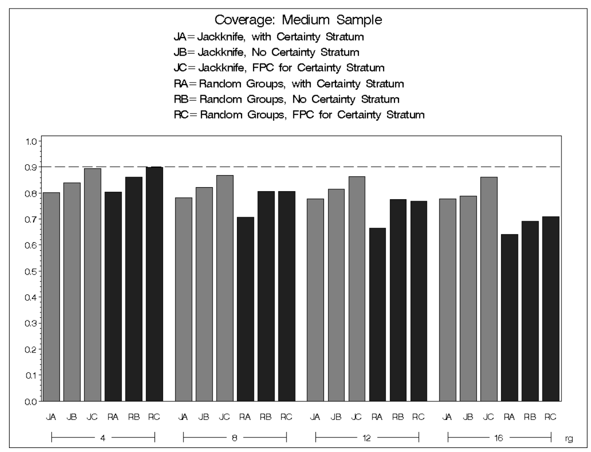
Table 2 shows the estimated coverage rate for 90% confidence intervals for the three sample sizes. The confidence intervals were based on t-statistics with degrees of freedom, df = G - 1. If only G_R groups were non-empty, then $df = G_R - 1$. Each chart presents four groups of six bars. The four groups represent varying the number random groups from 4 to 8 to 12 to 16. The six bars compare the three jackknife methods (A, B, and C) with the three random groups methods. The dashed line at the 90% level shows the target coverage rate of 90%.

Upon examining the results of the coverage comparisons, we conclude that all of the methods except for the jackknife method C have actual coverage percentages that fall far short of the desired coverage percentage. For most numbers of groups the coverage percentages for the random groups methods are less than the coverage percentages for the jackknife methods. Also note that the actual coverage percentage decreases with increasing numbers of random groups. One possible source of this is the decrease in the Student's t value used to compute the confidence intervals. The t-value used for computing confidence intervals based on 4 groups is 2.35; the corresponding value based on 16 groups is 1.75. Thus the confidence intervals based on 4 groups would be expected to be about 34% longer.

Table 3 gives the estimated RMSE of each estimator. The results are displayed in the same manner as the coverage percentages. Here, the target error is, of course, zero. Note in these charts that the RMSE of the random groups estimates is considerably less than the RMSE of the jackknife estimates.

Confidence Interval Coverage





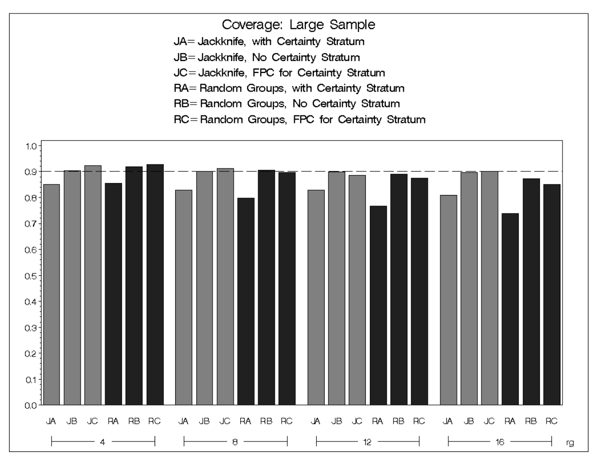
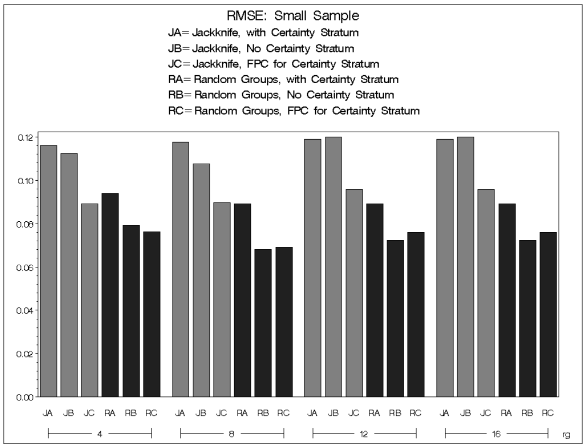
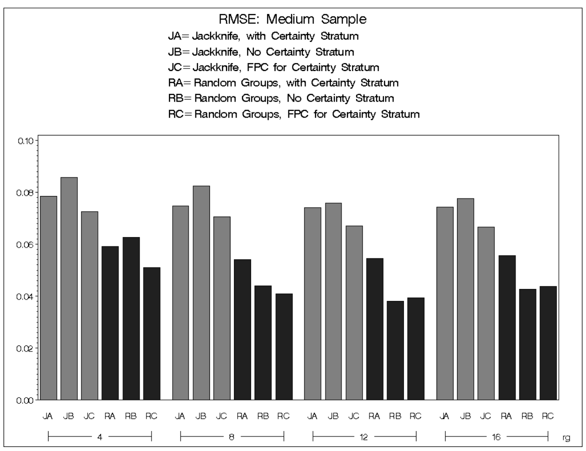


Table 2

RMSE





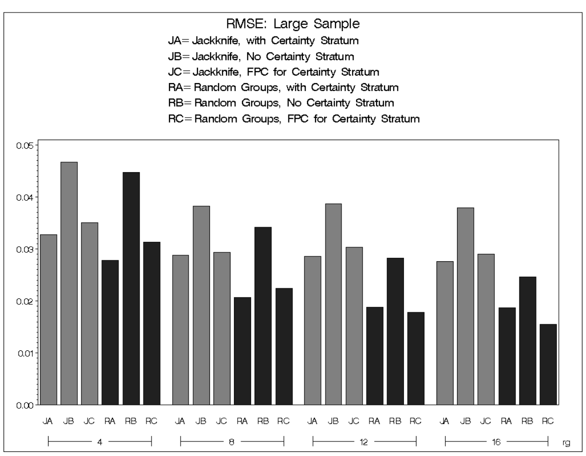


Table 3

6.2. Historical Comparison

Since random group totals were available for both MARTS and MRTS starting in October 2001, it was possible to obtain a separate estimate of the MARTS variance by combining the estimated expected value of MARTS revisions (changes made to the MARTS estimate when the full MRTS sample estimate is released), with the estimated MRTS variance (from the MRTS sample). The historic MARTS variance can be approximated by its MSE. The MSE can be estimated as follows.

Let A = MARTS estimate, R = MRTS estimate, U = population value.

$$\begin{aligned} MSE(A) &= E(A-U)^2 = E((A-R) + (R-U))^2 \\ &= E(A-R)^2 + E(R-U)^2 + 2E((A-R)(R-U)) \\ &= E(Revision)^2 + MSE(R) + 2E((A-R)(R-U)). \end{aligned}$$

The right hand term is approximately zero as can be seen by conditioning on R.

So, $VAR(A) \approx MSE(A) \approx E(Revision)^2 + MSE(R)$.
Note: $E(Revision)^2$ can be estimated from historical revisions. MSE(R) can be estimated by the MRTS random groups variance estimates.

Comparison results for all MARTS industries (based on the North American Industry Classification System (NAICS)) using the random groups technique with methods A, B, and C (16 random groups) are given in Table 4.

NAICS	RMSE Historic	Estimator		
		RA	RB	RC
441	1.47%	1.71%	1.82%	1.77%
442	1.43%	2.06%	2.50%	2.25%
443	22.74%	2.52%	3.70%	3.87%
444	1.51%	1.16%	2.03%	1.62%
445	0.81%	0.36%	0.90%	0.70%
446	1.09%	0.68%	2.50%	1.61%
447	1.44%	0.80%	1.24%	1.02%
448	2.16%	0.88%	2.52%	1.72%
451	5.93%	2.37%	4.76%	3.67%
452	0.55%	0.08%	2.27%	1.44%
453	3.51%	4.66%	7.80%	6.60%
454	5.65%	4.43%	14.84%	11.58%
722	1.21%	1.05%	1.23%	1.14%

Table 4

7. Conclusions

Random Groups vs. Jackknife

In the next several tables, we average across the various sample sizes because we are interested in selecting the single best method and number of random groups that will be used for all the industries covered by the MARTS. Our current processing system will not allow for varying the number of random groups by industry. This may change in the future, however. Table 5 and Table 6 show the overall results of the simulation averaged across sample sizes and number of random groups.

Coverage						
		4	8	12	16	All
All Sample Sizes	RA	0.813	0.739	0.692	0.674	0.729
	RB	0.868	0.808	0.761	0.727	0.791
	RC	0.898	0.827	0.775	0.747	0.812
	JA	0.816	0.800	0.794	0.788	0.800
	JB	0.847	0.835	0.828	0.819	0.832
	JC	0.905	0.893	0.880	0.881	0.890

Table 5

RMSE						
		4	8	12	16	All
All Sample Sizes	RA	0.060	0.055	0.054	0.054	0.056
	RB	0.062	0.049	0.046	0.046	0.051
	RC	0.053	0.044	0.044	0.045	0.047
	JA	0.076	0.074	0.074	0.074	0.074
	JB	0.082	0.076	0.079	0.079	0.079
	JC	0.066	0.063	0.064	0.064	0.064

Table 6

It appears from Table 5 that the jackknife method came closer overall to achieving the desired confidence interval. On the other hand, Table 6 indicates that the random groups method performed better in terms of RMSE. The higher RMSE of the jackknife method may result from a higher bias, or higher standard error. The following tables give the estimated bias and standard error from the simulation.

Bias						
		4	8	12	16	All
All Sample Sizes	RA	-0.022	-0.030	-0.037	-0.039	-0.032
	RB	-0.004	-0.014	-0.021	-0.026	-0.016
	RC	-0.009	-0.019	-0.026	-0.030	-0.021
	JA	-0.006	-0.001	0.000	0.001	-0.002
	JB	0.006	0.008	0.017	0.016	0.012
	JC	0.009	0.011	0.017	0.017	0.013

Table 7

Table 7 shows that the bias of the jackknife estimator was lower in magnitude than the random groups estimator. Looking at Table 8, it is seen that the larger RMSE of the jackknife estimator is due to a higher standard error.

Standard Error						
		4	8	12	16	All
All Sample Sizes	RA	0.055	0.044	0.039	0.038	0.044
	RB	0.058	0.039	0.031	0.029	0.039
	RC	0.049	0.035	0.031	0.029	0.036
	JA	0.075	0.073	0.074	0.073	0.074
	JB	0.080	0.073	0.076	0.076	0.076
	JC	0.064	0.061	0.061	0.061	0.062

Table 8

Method A vs. B vs. C

The simulation results indicate that the finite population correction of Method C seems to work well with both jackknife and random groups variance estimators—either for reduction of RMSE, or obtaining more accurate confidence intervals. The historic series comparisons of Table 4 also show some evidence that the historic estimates of the MSE obtained for MARTS were a closer to the MSEs estimated using method C.

Number of Random Groups

It appears from Table 6 that increasing the number of random groups generally leads to a decrease in RMSE. Table 3 shows that only for the small sample does there appear to be a slight decrease in RMSE with fewer random groups. The confidence interval coverage is markedly better for fewer random groups if the sample size is small or medium.

References

Thompson, K., Sigman, R., Goodwin, R. (2002). Investigation of Alternative Replicate Variance Estimators for the Annual Capital Expenditures Survey. Unpublished Internal Memorandum. Washington DC: U.S. Census Bureau.

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