INVESTIGATION OF ALTERNATIVE REPLICATE VARIANCE ESTIMATORS FOR THE ANNUAL CAPITAL EXPENDITURES SURVEY

Katherine J. Thompson[1], Richard S. Sigman, Roger L. Goodwin
U.S. Bureau of the Census

1. INTRODUCTION

The Annual Capital Expenditures Survey (ACES) collects information about the nature and level of capital expenditures in non-farm companies, organizations, and associations in the United States. ACES uses a one-stage stratified simple random sample without replacement (SRS-WOR) design, and sampling fractions in several strata are larger than 0.20. ACES performs weight adjustment for unit non-response and does not perform imputation for item non-response. Section 2 describes the ACES sample design and estimation methodology.

From the collection year 2000 data onward, ACES began using the U.S. Census Bureau's Standard Economic Processing System (StEPS) as its post-data-collection system (Ahmed and Tasky, 2000). While the existing StEPS estimation module software easily accommodated the ACES estimators, variance estimation enhancements were required. In prior collections periods, ACES used a sampling theory formula (S^2) variance estimator with non-response adjusted weights in place of sampling weights, so that the ratio-adjusted weights were treated as constants in the production variance estimates. StEPS is a generalized system, which lends itself more to replication variance estimation methods. Thus, the primary purpose of this study was to determine whether replicate variance estimation could be used to estimate ACES variances.

By 2000, StEPS included variance estimation software for the method of random groups. This variance estimation method is quite popular with many of the U.S. Census' business surveys for theoretical and for operational reasons. Theoretically, random group variance estimates of expansion estimators are nearly unbiased for stratified SRS-WOR samples with small sampling fractions [the most-commonly used design for our non-manufacturing business surveys]. Operationally, surveys that incorporate births (new businesses) into ongoing samples can easily and correctly include the new units in the variance estimations by assigning new units to random groups as they are selected. Moreover, random group estimation requires fewer computer resources than other more popular methods such as the stratified jackknife: K replicate estimates or replicate weights, where K is the number of random groups versus one replicate estimate or replicate weight per **respondent** for the stratified jackknife.

Random group variance estimation has two drawbacks. First, it can be unpredictable when applied to SRS-WOR samples because the random group estimator "tends to estimate the variance as if the sample were selected with replacement" (Wolter, 1985, p.43). The second drawback is the instability of the random group variance estimates, especially when the number of sampled observations in each random group is small or when there is a high rate of unit non-response. Because of these drawbacks, we investigated the delete-a-group jackknife variance estimator. This method can be applied to the same types of sample designs as the random group method and should yield more stable variance estimates, since replicates are constructed from more sample units. Moreover, since StEPS already had random group estimation capability, we knew that that the number of delete-a-group replicates or replicate weights would not pose an operational problem. Kott (2001 and 1998) reports excellent results using the delete-a-group jackknife for several of the National Agricultural Statistics Service (NASS) programs with a variety of sample designs (including stratified SRS-WOR) for expansion, ratio, and restricted regression estimators. Smith (2001) also reports some success with the delete-a-group jackknife variance estimator for New Zealand's labor force survey.

Our investigation specifically examined how to modify random group and delete-a-group jackknife variance estimators for without replacement samples with non-negligible sampling fractions and whether the non-response adjustment procedure should be repeated in each replicate. The first issue is discussed in Sections 3 and 6, and the second issue is discussed in Sections 4 and 5. Section 4 provides our empirical results for three key capital expenditures characteristics using 1999 ACES data. The results from this empirical estimation motivated the simulation study described in Section 5. Section 6 describes an appropriate way to use these replication methods to obtain variance estimates for combined ratio or trend estimates for survey designs with non-negligible sampling fractions. Section 7 provides our conclusions.

2. ACES SAMPLE DESIGN AND ESTIMATION METHODOLOGY

The ACES universe contains two sub-populations: employer companies and non-employer companies. Different forms are mailed to sample units depending on whether they are

[1] The authors thank William C. Davie, Jr., Jacklyn Jonas, Andromache Howe, and Quatracia Williams for their useful comments on earlier versions of this manuscript. This paper reports the results of research and analysis undertaken by the U.S. 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.

employer (ACE-1) companies or non-employer (ACE-2) companies. New ACE-1 and ACE-2 samples are selected each year, both with stratified SRS-WOR designs. The ACE-1 sample comprises approximately seventy-five percent of the ACES sample (roughly 45,000 companies selected per year for ACE-1, and 15,000 selected per year for ACE-2).

The ACE-1 frame is stratified first by primary industry activity found in the Census Bureau's Business Register. Five separate strata are formed within industry: one certainty stratum consisting of companies with 500 or more employees, and four non-certainty strata determined using a modified Lavallee-Hidiriglou method with payroll as a measure of size (Slanta and Krenzke, 1996). Sampling fractions in the noncertainty ACE-1 strata can be quite high: in the 1999 design, 113 of the 514 strata had sampling fractions greater than 0.20.

Unlike the ACE-1 design, sampling fractions in all ACE-2 strata are quite low (all less than 0.01 in the 1999 design). Two of the ACE-2 strata are post-stratified using updated information from the Business Register after data collection.

The ACE-1 and ACE-2 non-response weight procedures follow the adjustment-to-sample models described in Kalton and Kasprzyk (1986), i.e., all sampling weights in a weighting class *l* are multiplied by a factor derived from data corresponding to sample units. The ACE-1 non-response adjustment procedure controls sampling weights to independently obtained estimates of payroll; that is, the non-response weighting adjustment factor for a weighting cell *l* is the sum of the sample-weighted payroll estimates for units in the weighting cell divided by the sum of the sample-weighted payroll estimates for all responding units in the weighting cell. Under complete non-response in a certainty stratum or complete non-response in the large company stratum, the two strata are combined into one weighting cell (within the sample industry). Presently, there is no collapsing procedure in place for complete non-response in the three remaining non-certainty strata. The post-stratified ACE-2 estimates are controlled to sample counts within strata; that is, the non-response weighting adjustment factor for a weighting cell *l* is calculated as the number of sampled units in the weighting cell divided by the number of responding units in the weighting cell [Note: since ACE-2 performs non-response adjustment within strata, the sample weights cancel out]. Thus, the final weight for survey *v* ($W_{h(adj)v}$) is given by

$$W_{h(adj),v} = (W_h)\left(\frac{S_l}{R_l}\right) \text{ for all } h \in l$$

$$S_l = \begin{cases} \sum_{h \in l} \sum_{j \in h} p_{hj} w_h, & l \in \text{ACE}-1 \\ \sum_{h \in l} n_h, & l \in \text{ACE}-2 \end{cases}$$

$$R_l = \begin{cases} \sum_{h \in l} \sum_{j \in h} p_{hj} r_{hj} w_h, & l \in \text{ACE}-1 \\ \sum_{h \in l} \sum_{j \in h} r_{hj}, & l \in \text{ACE}-2 \end{cases}$$

where *h* indexes the stratum, *v* indexes the survey, *j* indexes the sample unit, p_h is the estimated payroll in stratum *h*, n_h is the sample size of stratum *h*, $W_h = N_h/n_h$ (the sampling weight), and r_{hj} is a variable indicating the response status of sample unit *j*. See Caldwell (1999a) for more details on the ACE-1 and ACE-2 non-response weighting adjustment procedures

ACES publishes expansion estimates for all characteristics. Technically, these estimates are non-linear because of the ratio non-response weight adjustment procedure. Additionally, ACES publishes year-to-year trend estimates.

3. VARIANCE ESTIMATION METHODOLOGY

This section describes our considered variance estimation procedures. Our non-replicate variance estimator for characteristic $\hat{X}_i (= \sum_{v=1}^{2} \sum_h W_{h(adj)v} x_{hij})$ is an approximate sampling-formula (S^2) variance given by

$$\hat{\text{var}}_{s^2}(\hat{X}_i) = \sum_{v=1}^{2} \hat{\text{var}}_{s^2}(\hat{X}_{i,ACE_v}) \quad (3.1)$$

$$\hat{\text{var}}_{s^2}(\hat{X}_{i,ACE_v}) = \sum_{v=1}^{2} \sum_{h}^{NSR\,Strata \in v} N_h (W_{h(adj)v} - 1) s_{hv}^2 \quad (3.2)$$

where $N_h \approx n_h \bullet W_h$, the approximated population size of stratum *h* and

$$s_h^2 = \begin{cases} x_{hij}^2 \; if \, r_h = 1, \; (\text{r}_\text{h} = \sum_j^{n_h} r_{hj}) \\ \dfrac{\sum_{j \in h} x_{hij}^2}{r_h - 1} - \dfrac{\left(\sum_{j \in h} x_{hij}\right)^2}{r_h (r_h - 1)}, \text{ otherwise} \end{cases}$$

See Caldwell (1999b). Using x_{hij}^2 to estimate s_h^2 when a stratum contains one respondent allows the stratum to contribute to the variance computation (however poorly). Its use is not theoretically justified.

Notice that formula 3.2 directly incorporates the finite population correction (fpc). Wolter (1985, Ch.2) provides modifications for random group and delete-a-group jackknife estimators for stratified samples with non-negligible sampling fractions, specifically suggesting using $W_h^* = \sqrt{1 - f_h} W_h$ in place of W_h, where W_h is the stratum final weight and $f_h = n_h/N_h$ is the stratum sampling fraction. We used this adjustment in all of the replicate procedures described below.

The random group method begins by splitting the non-certainty portion of the survey sample into *K* random groups using the survey's sampling methodology (Wolter, 1985, pp. 31-32). Each random group's sample is then reweighted to represent the full sample, either by simply multiplying the random group estimate by *K* (simple reweighting) or by developing *K* replicate weighting factors within each stratum (strata-specific reweighting). The strata-specific replicate factors for stratum *h* and random group *k* are n_h / m_h^k, where m_h^k is the number of sample units in stratum *h* assigned to random group *k*. Developing strata-specific replicate weighting factors yields replicate estimates that are conditionally unbiased. Such replicate weighting factors may, however, increase the variance of the estimated variances since they differ by strata. Certainty units are included in each random group. These cases (or their associated replicate weights) are not multiplied by *K* or any adjustment factor. Thus, *K* replicate weights are assigned to each sample unit *j*. If unit *j* is in a non-certainty stratum, the k^{th} replicate weight is zero unless unit *j* is in random group *k*. In a certainty stratum, all *K* replicate weights are equal to the sampling or final weight.

The full sample estimation procedure is then applied to each of the replicate weights (e.g., non-response adjustments, post-stratification) or to the replicate estimates. The random group variance for any estimate $\hat{\theta}_i$ is

$$\hat{\text{var}}_{RG}(\hat{\theta}_i) = \frac{\sum_k (\hat{\theta}_{ki(RG)} - \hat{\theta}_{0i})^2}{K(K-1)} \qquad (3.3)$$

where $\hat{\theta}_{ki(RG)}$ is the random group *k* replicate estimate and $\hat{\theta}_{0i}$ is the full-sample estimate. If the replicate weights used incorporate the fpc adjustments described above, then $\hat{\theta}_{0i}$ and $\hat{\theta}_i$ (the tabulated full-sample estimate) will not be equivalent.

For delete-a-group jackknife variance estimation, again the non-certainty portion of survey sample is divided into *K* random groups. However, the delete-a-group jackknife replicate estimate is computed for each replicate *k* by removing the k^{th} random group from the full sample. Replicates are obtained either by multiplying each replicate estimate by *K/(K-1)* or by developing stratum-specific replicate weights. The strata-specific replicate factors for stratum *h* and delete-a-group jackknife replicate *k* are $n_h / (m_h - m_h^k)$. Certainty units are included in each delete-a-group jackknife replicate estimate. Thus, for delete-a-group jackknife replication, *K* replicate weights are assigned to each sample unit *j*. If unit *j* is in a non-certainty stratum, the k^{th} replicate weight is zero when unit *j* is in random group *k*. In a certainty stratum, all *K* replicate weights are equal to the sampling or final weight.

Since jackknife replicate sample sizes are larger than the corresponding random group replicate sample sizes, delete-a-group jackknife variance estimates are often more stable, at least for smooth statistics. The delete-a-group jackknife variance for an estimate $\hat{\theta}_i$ is

$$\hat{\text{var}}_{DAG}(\hat{\theta}_i) = \frac{K-1}{K} \sum_k \left(\hat{\theta}_{ki(DAG)} - \hat{\theta}_{0i}\right)^2 \qquad (3.4)$$

where $\hat{\theta}_{ki(DAG)}$ is the replicate *k* delete-a-group jackknife estimate.

As stated earlier, we also wanted to investigate the statistical properties of replicating the non-response adjustment procedure (i.e., independently performing the non-response adjustment procedure on each set of replicate weights). This can be quite time-consuming and computer resource-intensive, so we considered a "shortcut" approach, using the full sample non-response adjusted weights in all replicates. Institutional intuition held that both the sampling formula variance procedure and shortcut approach would **underestimate** the true variance by failing to explicitly account for non-response variance. This intuition is somewhat supported in the literature: for example, Canty and Davison (1999) found replicating the calibrated weighting procedure reduced the degree of relative bias in their stratified jackknife variance estimates (from those using the "shortcut approach") for a similar design. On the other hand, Wolter (1985, pp. 83-84) cites results from two studies that showed the slight improvements in random group variance estimates using full replicate reweighting versus the shortcut approach did not offset the additional computing costs. In a similar vein, Schindler (2002) found trivial differences between the variance computed with a fully-reweighted stratified jackknife procedure versus those obtained with a simple jackknife that used final weights in all estimates (shortcut procedure) for selected dual system estimates from the Census 2000 Accuracy and Coverage Enumeration Survey.

We considered three different replicate weighting variations per replication method:

Simple *Construct replicate weights from the full sample's* ***non-response adjusted weights*** *(the shortcut). Random group estimation uses K as the replicate adjustment factor; delete-a-group jackknife estimation uses K/K-1.*

Simple Reweighted *Construct replicate weights from the sample weights, then perform the non-response adjustment procedure on each set of replicate weights. Random group estimation uses K as the replicate adjustment factor; delete-a-group jackknife estimation uses K/K-1.*

Stratified Reweighted *Construct replicate weights from the sample weights, where non-certainty units' adjusted weights in a given replicate are multiplied by strata-specific replicate factors. Perform non-response adjustment on each set of replicate weights.*

In subsequent sections, we use "RG" to indicate random group and "DAG" to indicate delete-a-group jackknife, combined with "S" (simple), "SR" (simple reweighted), and "STR" (stratified reweighted). Certainty cases are excluded from all of the discussed replicate variance estimates via the fpc-adjustment (all certainty cases have replicate weights of zero). None of the replicate variance methods described account for the variance contribution due to non-responding certainty units. This is consistent with ACES current production method.

We used 15 random groups in all applications. For a stratified SRS-WOR design, Kott (2001) proves that the delete-a-group jackknife variance estimator is approximately unbiased for the true variance when the sample size in each stratum is larger than the number of random groups and all sampling fractions are negligible (less than or equal to 1/5) and is biased upwards otherwise. Thus, the 1999 design ACE-2 delete-a-group jackknife variance estimates are approximately unbiased. This is not the case with ACE-1: there were twelve (of 514) ACE-1 strata that did not have sample in all fifteen random groups. Moreover, the proportion of strata that are not represented in any random group is actually higher due to unit non-response. Consequently, the ACE-1 DAGSTR estimates are biased upwards. The bound on this bias given by Kott (2001) – that is, $(14/15)\min_h\{1/(n_h-1)\}$ – is probably not applicable because such a high proportion of the ACE-1 sampling fractions are quite large.

4. <u>EMPIRICAL DATA RESULTS</u>

Initially, we compared the six replicate variance estimators to the sampling formula approximation for three capital expenditure statistics (**Total** Capital Expenditures; Capital Expenditures on **Structures**; Capital Expenditures on **Equipment**); using 1999 ACES data.

We found several interesting patterns. First, the sampling formula (S^2) standard errors were generally larger than corresponding replicate estimates. We found this perplexing, having assumed that the S^2 and simple replicate variance methods would consistently underestimate the variance since they do not explicitly account for the non-response adjustment. Second, performing non-response adjustment in each delete-a-group jackknife replicate – the simple reweighted (SR) or stratified reweighted (STR) methods -- usually **reduced** the estimated standard error from the corresponding simple replication estimate (again, a counter-intuitive result). There is no consistent pattern with the random group estimates. Finally, the simple reweighted jackknife estimates were less than or equal to the corresponding stratified reweighted jackknife estimates. This was reasonable, since the variable replicate factors used for stratified reweighting should increase the variability among the replicate estimates. In contrast, using strata-specific random group adjustment factors reduced the estimated standard errors for all ACE-1 characteristics and for three ACE-2 characteristics. This did not seem reasonable. This last difference could partially account for the variability in replicate factors for the RGSTR and DAGSTR methods. As Table 1 shows, the strata-level random group adjustment factors were quite variable and were (on the average) quite different from their expected value, unlike the corresponding delete-a-group jackknife adjustment factors.

Table 1: Strata-Level Adjustment Factors for 1999 ACES Data

Frame	Method	Expected Mean	Sample Mean	Standard Deviation	Minimum	Maximum
ACE-1	RGSTR	15.00	15.59	3.38	1.00	29.00
	DAGSTR	1.07	1.07	0.03	1.03	2.00
ACE-2	RGSTR	15.00	15.12	1.42	12.5	20.00
	DAGSTR	1.07	1.07	0.01	1.05	1.09

The inconsistent empirical results for the ACE-1 random group estimates are partially explained by the replicated non-response adjustment procedures. Table 2 presents the number of weight adjustment cells with complete non-response by replication method for ACE-1 [Note: the results are equivalent for the simple and stratified reweighted methods]. "Total 2A" refers to the large-size non-certainty within-industry strata, which are collapsed with industry certainty strata under complete non-response for weighting adjustments. In all other ACE-1 strata, the weighting cell l is equivalent to the stratum h.

With the random group methods, an overly high proportion of ACE-1 strata have no respondents in a replicate weighting cell. This poses two problems. The first is mechanical: except for ACE-1 Stratum 2A, ACES does not have a collapsing mechanism in place for non-certainty strata. Second, and far more important, the random group variance estimation method is not mimicking the full-sample estimation procedure. In contrast, in all but one replicate, the delete-a-group jackknife replicates use the same weighting cells as the full sample.

Table 2: ACE-1 Weighting Cells with Complete Non-Response

Method		Full Sample	Replicate														
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RG	Total 2A	1	18	20	19	24	22	28	25	22	21	21	23	21	23	17	22
	Other	0	40	38	41	37	42	40	39	34	33	35	30	37	39	34	38
DAG	Total 2A	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1
	Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Since approximately twenty-percent of the ACE-1 strata sampling rates are larger than 0.20, we were concerned that **not** incorporating the fpc into the replicate variance estimates could lead to substantial overestimates of variance. Table 3 presents ratios of unadjusted to fpc-adjusted replicate standard errors for the same three ACE-1 characteristics [Note: ACE-2 sampling fractions are all less than 0.01, so all standard error ratios are 1, as expected].

Table 3: ACE-1 Standard Error Ratios (Without FPC/With FPC)

	RGS	RGSR	RGSTR	DAGS	DAGSR	DAGSTR
Total	1.02	1.01	1.02	1.02	1.02	1.03
Structures	1.00	1.00	1.00	1.00	1.01	1.01
Equipment	1.06	1.06	1.06	1.06	1.06	1.07

Failing to account for the non-negligible fpcs in the variance estimates leads to a six-percent overestimate of standard error for capital expenditures on equipment and consistently overestimates the SE for total capital expenditures (the primary statistic of interest) by approximately two-percent. There is no real difference between standard error estimates for capital expenditures on structures, but this is a fairly small characteristic. The overestimation for the other two characteristics could affect coverage (in particular for capital expenditures on equipment), thus justifying the need for incorporating fpcs in the replication.

The overly variable replicate factors for the stratified reweighted random group method concerned us. The unrealistic non-response adjustment collapsing pattern with the two reweighted methods convinced us not to further pursue random group variance estimation methods with ACES data and to instead concentrate on delete-a-group jackknife variance estimation methods. Of course, our empirical results still left us with four different sets of variance estimates and no "gold standard" against which to measure them. So, we conducted a Monte Carlo simulation study to evaluate the properties of these four different variance estimators.

5. SIMULATION STUDY

5.1 Creation of the Frame and Sample Selection

Capital expenditures data are difficult to model. First, they are often poorly correlated with auxiliary data such as payroll or employment, especially for small companies. Second, purchasing patterns are not necessarily consistent within an industry. For example, in some industries, capital expenditures on structures and equipments are negatively correlated for small companies and positively correlated for large companies. The multivariate correlation structure becomes more complicated when capital expenditures data is further cross-classified by new or used status. Consequently, we developed models only for non-certainty employer companies, using the actual reported sample data for certainty companies in eight sample NAICS industries provided by ACES methodologists. These industries encompassed a variety of professional sectors: Utilities; Manufactures; Wholesale Trade; Retail Trade; Information; Professional, Scientific, and Technical Services; and Administrative Support, Waste Management, and Remediation Services. This simulation study did not include the ACE-2 frame data, which represents approximately 25% of the total ACES sample universe but only approximately seven percent of the total estimated capital expenditures.

In general, each sample industry required three separate sets of models: one for units that reported all capital expenditures on equipment; one for units that reported all capital expenditures on structures; and one for units that reported spending on both. In the latter case, we modeled two of the three characteristics explicitly, deriving the remaining characteristic as the difference of the other characteristics. In each sample industry, we randomly applied the three sets of models to the frame data in the same proportions as in the (respondent) sample data: that is, first, we simulated total capital expenditures data in the same proportion as reported in the industry, then we applied our three different models to the units that had nonzero simulated capital expenditures data.

We stratified this complete frame data using the ACES production programs. After stratification, we used a missing-at-random model to assign response status in which the probabilities of non-response matched the ACES non-response rates by stratum. Thus, we assumed a fixed set of non-respondents on the frame, adding bias to the estimator but allowing for different response patterns by sample.

Finally, we selected 5,000 stratified SRS-WOR samples from this simulated population using the strata sampling rates from the ACES stratification and allocation programs. In 1,000 of the 5,000 samples, we assigned sample units to 15 random groups. Like the empirical study, not all strata contain all random groups: one of the 32 non-certainty strata contains ten sample units.

5.2 Evaluation Criteria

To examine the statistical properties of the four different variance estimation methods, we used our 5,000 stratified random samples to construct the empirical variance of each characteristic *i* in sample industry *u* as

$$V(\hat{X}_{iu}) = \frac{\sum_{r=1}^{5,000} (\hat{X}_{riu} - \overline{X}_{iu})^2}{5,000} \qquad (5.1)$$

where $\hat{X}_{riu}$ is the estimate of characteristic *i* in industry *u* in sample *r*, and $\overline{X}_{iu}$ is the mean of the $\hat{X}_{riu}$.

Next, we calculated four variance estimates (v_{meth}) per characteristic *i* in industry *u* from 1,000 of the 5,000 samples. We compared these variance estimates in terms of

$$\textbf{Relative Bias} = \frac{\frac{1}{1,000}\sum_{r=1}^{1,000} v_{meth}(\hat{X}_{riu})}{V(\hat{X}_{iu})} - 1$$

$$\textbf{c.v.}(\mathbf{v}_{\mathbf{meth}}) = \frac{\sqrt{\frac{1}{1,000}\sum_{r=1}^{1,000}\left[v_{meth}(\hat{X}_{riu}) - V(\hat{X}_{iu})\right]^2}}{V(\hat{X}_{iu})}$$

Relative bias is a measure of the bias of the variance estimate as a proportion of the true variance. The coefficient of variation (c.v.) measures the variance of the variance estimates; this statistic is called the "stability" in other publications (e.g., Rao and Shao, 1996). With an "optimal" variance estimator, both the relative bias and the c.v. will be near zero.

5.3 Results

Table 4 presents the relative biases of the four variance estimation methods. Statistically significant contrasts between biases (α = 0.05) are shaded. To compare contrasts in relative biases between variance estimation methods within industry for each characteristic, we used an ANOVA approach with the repeated measures model $v_{meth}(\hat{X}_{riu}) = \mu_{meth,iu} + \varepsilon_{meth,riu}$.

For each characteristic/industry, we first tested the "omnibus hypothesis" (H_0: $\mu_{S2,iu} = \mu_{DAGS,iu} = \mu_{DAGSR,iu} = \mu_{DAGSTR,iu}$). Since we rejected H_0 for all characteristics in all industries, we tested the contrasts in variance estimate means. Pairwise differences between relative biases are statistically different when the contrasts between the corresponding variance estimate means are significantly different (i.e, $p \leq 0.05$).

Table 4: Relative Biases of Variance Estimators

	Ind.	Relative Bias				Contrasts (Relative Biases)				
		S^2	DAGS	DAGSR	DAGSTR	B_{S2}-B_{DAGS}	B_{S2}-B_{DAGSR}	B_{S2}-B_{DAGSTR}	B_{DAGS}-B_{DAGSR}	B_{DAGSR}-B_{DAGSTR}
Total	2213	-0.07	-0.03	-0.02	-0.02	-0.04	-0.05	-0.05	-0.01	0.00
	3253	0.03	0.04	0.06	0.11	-0.01	-0.03	-0.08	-0.02	-0.05
	3369	-0.20	-0.07	0.06	0.12	-0.13	-0.26	-0.32	-0.13	-0.06
	4220	-0.05	-0.03	-0.01	-0.01	-0.02	-0.04	-0.04	-0.02	0.00
	4480	-0.04	0.05	0.07	0.08	-0.09	-0.11	-0.12	-0.02	-0.01
	5112	-0.20	0.00	0.15	0.20	-0.20	-0.35	-0.40	-0.15	-0.05
	5415	-0.12	-0.12	-0.10	-0.09	0.00	-0.02	-0.03	-0.02	-0.01
	5619	-0.03	0.03	0.06	0.07	-0.06	-0.09	-0.10	-0.03	-0.01
Structures	2213	0.00	0.01	0.03	0.03	-0.01	-0.03	-0.03	-0.02	0.00
	3253	0.07	0.05	0.10	0.11	0.02	-0.03	-0.04	-0.05	-0.01
	3369	-0.21	-0.17	-0.09	-0.06	-0.04	-0.12	-0.15	-0.08	-0.03
	4220	-0.01	-0.02	-0.01	-0.01	0.01	0.00	0.00	-0.01	0.00
	4480	-0.06	-0.06	-0.04	-0.04	0.00	-0.02	-0.02	-0.02	0.00
	5112	-0.06	-0.02	0.05	0.08	-0.04	-0.11	-0.14	-0.07	-0.03
	5415	-0.12	-0.12	-0.11	-0.10	0.00	-0.01	-0.02	-0.01	-0.01
	5619	-0.04	-0.04	-0.02	-0.02	0.00	-0.02	-0.02	-0.02	0.00
Equipment	2213	-0.14	-0.11	-0.09	-0.09	-0.03	-0.05	-0.05	-0.02	0.00
	3253	0.04	0.04	0.06	0.12	0.00	-0.02	-0.08	-0.02	-0.06
	3369	-0.16	-0.03	0.09	0.16	-0.13	-0.25	-0.32	-0.12	-0.07
	4220	-0.11	-0.07	-0.05	-0.03	-0.04	-0.06	-0.08	-0.02	-0.02
	4480	-0.02	0.07	0.09	0.10	-0.09	-0.11	-0.12	-0.02	-0.01
	5112	-0.20	0.00	0.15	0.19	-0.20	-0.35	-0.39	-0.15	-0.04
	5415	-0.12	-0.11	-0.10	-0.09	-0.01	-0.02	-0.03	-0.01	-0.01
	5619	-0.02	0.04	0.07	0.07	-0.06	-0.09	-0.09	-0.03	0.00

There is very little evidence of difference between the two reweighted delete-a-group jackknife variance estimates (DAGSR and DAGSTR). Otherwise, the majority of contrasts are significantly different.

The relative bias results can be summarized as follows:

- S^2 relative biases are negative for all characteristics in all but one industry. On the average, this variance estimation method **underestimates** the true variance;
- For characteristics in industries with significant differences between the S^2 and DAGS relative biases, the DAGS method generally yields variance estimates whose relative bias is closer to zero (4 of 6 for total capital expenditures; 3 of 3 for structures; 4 of 6 for equipment);
- DAGSR relative biases are **always** larger than corresponding DAGS relative biases. Some of this bias increase (between DAGS and DAGSR) could be caused by having one stratum that is **not** represented in all random groups.

To summarize, Table 4 shows clear gains in relative bias using **either** the DAGS or DAGSR method over the S^2 method, but does not identify a clearly superior method in terms of bias. Table 5 presents the c.v.s of the variance estimates for each characteristic for each variance estimation method.

Table 5: C.V.s of the Four Variance Estimation Methods

	Industry	S^2	DAGS	DAGSR	DAGSTR
Total	2213	1.37	1.43	1.45	1.45
	3253	0.93	1.04	1.08	1.15
	3369	0.79	1.09	1.50	1.75
	4220	1.21	1.25	1.27	1.29
	4480	0.76	0.89	0.93	0.99
	5112	0.65	0.85	1.15	1.33
	5415	1.14	1.16	1.16	1.27
	5619	0.93	1.10	1.14	1.17
Structures	2213	1.70	1.75	1.79	1.79
	3253	1.67	1.72	1.85	1.88
	3369	1.59	1.82	2.09	2.10
	4220	1.70	1.67	1.69	1.69
	4480	1.67	1.64	1.67	1.68
	5112	2.05	2.25	2.45	2.52
	5415	1.16	1.18	1.18	1.23
	5619	3.55	3.41	3.45	3.45
Equipment	2213	2.09	2.16	2.19	2.19
	3253	0.97	1.09	1.13	1.19
	3369	0.93	1.29	1.70	1.93
	4220	0.53	0.62	0.66	0.98
	4480	0.86	1.01	1.05	1.07
	5112	0.66	0.86	1.15	1.33
	5415	1.11	1.14	1.14	1.28
	5619	0.92	1.08	1.13	1.15

Across the board, the S^2 variance estimates are the least variable. Of course, the S^2 estimator does not explicitly account for the variance component due to unit non-response and does not incur an additional resampling variance component. Reasonably, the variability of variance estimates increases with replication: all of the delete-a-group jackknife c.v.s are higher than the corresponding S^2 c.v.s. The variability of the variance estimates further increases when replicating the non-response adjustment (c.f. the DAGS to the DAGSR and DAGSTR stabilities), especially in the transportation manufacturing (3369) and software publishers (5112) industries. Finally, using the strata-specific adjustment factors (DAGSTR) instead of constant adjustment factors (DAGSR) increases the variance of the variance. Consequently, with our data sets, there is no advantage to the DAGSTR method: it yields overly variable variance estimates, its variance estimates are not statistically different from the DAGSR method, and it requires the most computer resources.

With these results, we have two almost equally good replicate variance estimators for ACES: DAGS and DAGSR. Both have very similar statistical properties. Moreover, in most samples, the two sets of variance estimates were very close.

We expected the DAGS estimates to consistently **underestimate** the true variance. This did not happen in our sample data: 49-percent of the DAGS variance estimates are larger than the DAGSR estimates for Total Capital Expenditures; 37-percent for Capital Expenditures on Structures; and 49-percent for Capital Expenditures on Equipment. This pattern is consistent with our empirical results.

Here, the slight relative bias improvements of the DAGSR method over the DAGS method do not completely offset the worsening stability measures (i.e., increased cv(v_{meth})) of the DAGSR method. Moreover, the DAGS method is much faster and is less computer-resource intensive than the DAGSR method. Finally, as mentioned in Section 4, the ACE-1 non-response adjustment procedures only provide collapsing criterion for two of the five within-industry strata. By design, complete non-response in the remaining three strata is highly unlikely. However, we cannot guarantee that will always happen if we use DAGSR variance estimation for ACES, and we want to avoid the type of collapsing problem seen in Section 4 with random group estimation. For these reasons, ACES methodologists elected to use the DAGS method.

6. <u>COMPUTING VARIANCES OF COMBINED RATIO AND TREND ESTIMATORS</u>

The literature supports using jackknife-type variance estimates for ratio estimators when sampling fractions can be ignored (e.g., Kott 2001, Rao and Shao 1996). When sampling fractions are large, directly replicating the variance of a combined ratio estimate can yield large overestimates. To see this, consider the simple example of a SRS-WOR design. For **any** estimator $\hat{\theta}_i$,

$$\mathrm{Var}_{\text{SRS-WOR}}(\hat{\theta}_i) \approx (1\text{-}f)\mathrm{Var}_{\text{SRS-WR}}(\hat{\theta}_i) \qquad (6.1)$$

Our DAGS estimator multiplies each replicate weight k by the square-root of each unit's fpc. For expansion estimators, the DAGS estimate is

$$\begin{aligned} Var_{DAG}(\hat{\theta}_i) &= \frac{K-1}{K}\sum_k\left[\sqrt{1-f}\left(\frac{K}{K-1}\right)(W)\sum_{j\notin k}x_j - \sqrt{1-f}(W)\sum_j x_j\right]^2 \\ &= \frac{K-1}{K}\sum_k\left(\sqrt{1-f}\right)^2\left(\hat{\theta}_{(k)i}-\hat{\theta}_i\right)^2 \\ &= (1-f)\frac{K-1}{K}\sum_k\left(\hat{\theta}_{(k)i}-\hat{\theta}_i\right)^2 \\ &= (1-f)Var_{SRS-WR}(\hat{\theta}_i) \end{aligned}$$

With $h > 1$ strata, this fpc-adjustment to the DAGS replicate weights gives unbiased variance estimates for linear estimators (see Kott 2002, pp. 523-524, replacing t_{hj} with $t^*_{hj} = \sqrt{1-f_h}\,t_{hj}$.)

For ratio estimates, however, the DAGS estimate from a SRS-WOR design is

$$\begin{aligned} Var_{DAG}\left(\frac{\hat{X}}{\hat{Y}}\right) &= \frac{K-1}{K}\sum_k\left[\frac{\sqrt{1-f}\left(\frac{K}{K-1}\right)(W)\sum_{j\notin k}x_j}{\sqrt{1-f}\left(\frac{K}{K-1}\right)(W)\sum_{j\notin k}y_j} - \frac{\sqrt{1-f}(W)\sum_j x_j}{\sqrt{1-f}(W)\sum_j y_j}\right]^2 \\ &= \frac{K-1}{K}\sum_k\left[\frac{\sum_{j\notin k}x_j}{\sum_{j\notin k}y_j} - \frac{\sum_j x_j}{\sum_j y_j}\right]^2 \\ &= Var_{SRS-WR}\left(\frac{\hat{X}}{\hat{Y}}\right) \end{aligned}$$

Obviously, such cancellation does not occur with a stratified SRS-WOR design unless it is self-weighting. By extension though, our replicate weighting procedure will overestimate the variance of **combined ratio estimates**. Furthermore, it can be shown that there are only two survey designs for which applying the square-root-fpc correction to **only** the numerator replicate weights will yield correct **combined ratio estimates:** SRS-WOR and self-weighting stratified SRS-WOR. This technique will provide correctly adjusted estimates for **separate ratio estimators** under (unrestricted) stratified random sampling.

The ACES publishes year-to-year trend estimates. Like combined ratio estimators, trend estimators use estimates constructed from the full sample in both the numerator and the denominator (a trend estimate is the difference of the current and prior period estimates of characteristic i divided by the prior period estimate of characteristic i). Direct replication using square-root-fpc adjusted replicate weights is inappropriate, as illustrated at time t for a SRS-WOR design:

$$\begin{aligned} Var_{DAGS}\left(\frac{\hat{X}_{i,t}-\hat{X}_{i,t-1}}{\hat{X}_{i,t-1}}\right) &= \frac{K-1}{K}\sum_k\left[\frac{\sqrt{1-f_1}\left(\frac{K}{K-1}\right)(W_1)\sum_{j\notin k}x_{i,t}}{\sqrt{1-f_2}\left(\frac{K}{K-1}\right)(W_2)\sum_{j\notin k}x_{i,t-1}} - \frac{\sqrt{1-f_1}(W_1)\sum_j x_{i,t}}{\sqrt{1-f_2}(W_2)\sum_j x_{i,t-1}}\right]^2 \quad (6.1) \\ &= \left(\frac{K-1}{K}\right)\left(\frac{1-f_1}{1-f_2}\right)\sum_k\left(\frac{\hat{X}_{(k)i,t}}{\hat{X}_{(k)i,t-1}} - \frac{\hat{X}_{i,t}}{\hat{X}_{i,t-1}}\right)^2 \end{aligned}$$

where f_1 is the sampling fraction for the current sample and f_2 is the sampling fraction for the prior sample. This is an obviously poor approximation: if $f_1 < f_2$, the ratio of the two fpc's is **larger** than one, and consequently the estimated SRS-**WOR** variance is **larger** than the SRS-**WR**, which should be impossible. Moreover, even with this simple design, it is difficult to come up with a strategy that appropriately combines the two fpc adjustments (the geometric mean might be an option).

To avoid this problem, we use Taylor Series methods to estimate trend variances (Wolter, 1985, Ch. 6). The Taylor Series approximation for the variance of the trend estimator is given by

$$Var_{TAYLOR}\left(\frac{\hat{X}_{i,t}-\hat{X}_{i,t-1}}{\hat{X}_{i,t-1}}\right) \approx \left[\frac{\hat{X}_{i,t}}{\hat{X}_{i,t-1}}\right]^2\left[\frac{Var_{DAGS}(\hat{X}_{i,t})}{\hat{X}^2_{i,t}} + \frac{Var_{DAGS}(\hat{X}_{i,t-1})}{\hat{X}^2_{i,t-1}} - 2\frac{Cov_{DAGS}(\hat{X}_{i,t},\hat{X}_{i,t-1})}{\hat{X}_{i,t}\hat{X}_{i,t-1}}\right]$$

where the variance and covariance estimates are appropriately adjusted DAGS variance estimates for the expansion estimates [Note: the choice of replicate method is not particularly important]. With ACES, the covariance term is zero because of the independently-selected samples (Foreman, 1991, p.249). For non-independent samples, the covariance term can be obtained via subtraction.

7. CONCLUSION

We present results of a study comparing the statistical properties of two different replicate variance estimators (random group and delete-a-group jackknife) for a survey that uses a one-stage SRS-WOR design with non-negligible sampling fractions in several strata. We also examine the effects of fully replicating the non-response adjustment procedure versus a shortcut approach of using the full-sample's non-response adjusted weights to construct each replicate. Much of our study focuses on linear estimators, although we discuss applications of our methods to combined ratio and trend estimators. Our empirical data comparisons led us to eliminate the random group variance estimator from consideration for ACES. Our simulation study focused on the benefits of replicating the non-response adjustment procedure in delete-a-group jackknife replicates.

The simulation study results demonstrated some statistical advantages of both the simple delete-a-group-jackknife (DAGS) and the simple reweighted delete-a-group jackknife (DAGSR) methods over the approximate sampling formula method formerly used by ACES. They also provided evidence against using strata-specific replicate weighting factors recommended by Kott (2001): there were few – if any – relative bias improvements with this method over the others, coupled with increased variability of the variance estimates. Ultimately, the choice between the DAGS and DAGSR methods for ACES was not obvious in terms of the studied statistical properties. Thus, administrative considerations such as computer resources and production run time were the deciding factors leading us to recommend using the simple delete-a-group jackknife variance estimator for ACES, at least initially.

After choosing our variance estimator, we examined how to calculate replicate variances of non-linear estimators such as combined ratio estimators or year-to-year trend estimators for stratified SRS-WOR designs with non-negligible sampling fractions. For these estimators, we show that using replicate weights that incorporate the fpc to construct replicate estimates, then directly replicating combined ratio or trend variance tends to overestimate the variance [Note: this also applies to random group estimation under the same conditions]. Using Taylor linearized variance estimates reduces this overestimation.

Prior to this study, we assumed that directly replicating the non-response adjustment procedure was statistically preferable. Our results did not support this hypothesis. We used one sampling design and one non-response adjustment methodology, and we studied a survey that traditionally has a very high unit response rate (approximately 75%). While our results support conclusions cited in Wolter (1985) and Schindler (2002), more variations on both sample design and weight adjustment methodology are required before making any general recommendations.

Delete-a-group jackknife variance estimation is one of a variety of jackknife estimators. This particular estimator was appealing for anecdotal and production reasons. Examining alternative jackknife estimators such as the stratified jackknife for surveys with similar designs is an area of future study.

REFERENCES

Ahmed, S.A. and Tasky, D.L. (2000). An Overview of the Standard Economic Processing System. *Proceedings of the International Conference on Establishment Surveys II,* Alexandria, VA: American Statistical Association, pp. 633-642.

Caldwell, C. (1999a). Non-response Weighting for the 1999 Annual Capital Expenditures Survey (ACES). Unpublished internal memorandum. Washington, DC: U.S. Bureau of the Census.

Caldwell, C. (1999b). Estimation and Variance Estimation for the 1999 Annual Capital Expenditures Survey (ACES). Unpublished internal memorandum. Washington, DC: U.S. Bureau of the Census.

Canty, A.J. and Davison, A.C. (1999). Resampling-based Variance Estimation for Labour Force Surveys. *The Statistician*, **48**, pp. 379-391.

Fay, R.E. (1995). VPLX: Variance Estimation for Complex Surveys. Unpublished program documentation. Washington, DC: U.S. Bureau of the Census.

Foreman, E.K. (1991). *Survey Sampling Principles.* New York: Marcel Decker, Inc.

Kalton, G. and Kasprzyk (1986). The Treatment of Missing Survey Data. *Survey Methodology*, **12**, pp. 1 – 16.

Kott, P. (1998). Using the Delete-A-Group Jackknife Variance Estimator in NASS Surveys. RD Research Report Number RD-98-01. Washington, DC: National Agricultural Statistics Service.

Kott, P. (2001). The Delete-A-Group Jackknife. *Journal of Official Statistics,* **17**, pp. 521-526.

Rao, J.N.K. and Shao, J. (1996). Some Recent Work on Resampling Methods for Complex Surveys. *Survey Methodology,* **18**, pp. 209-217.

Schindler, E. (2002). Simplified Variance Estimation for Complex Surveys. *Proceedings of the Section on Survey Research Methods,* American Statistical Association, forthcoming.

Slanta, J.G. and Krenzke, T.R. (1996). Applying the Lavallee and Hidiroglou Method to Obtain Stratification Boundaries for the Census Bureau's Annual Capital Expenditures Survey. *Survey Methodology*, **22**, pp.65-76.

Smith, H. (2001). Investigation of the Delete-a-Group Jackknife Variance Estimator for the HLFS. Research report, Statistics New Zealand Paper 13, Te Tari Tatau. Wellington, New Zealand.

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