An Empirical Investigation Into The Effects Of Replicate Reweighting On Variance Estimates For The Annual Capital Expenditures Survey

Katherine J. Thompson[1]

U.S. Census Bureau
Room 3108-4,Washington, D.C. 20233, U.S.A/ katherine.j.thompson@census.gov

Introduction

Many of the business surveys conducted by the U.S. Census Bureau use a one-stage stratified simple random sample without replacement (SRS-WOR) design. When the sampling unit is an establishment or even an EIN (tax-reporting entity), the sampling rates are generally negligible (e.g., less than 20 percent in all strata). However, if the ultimate sampling unit is the company, the size of the universe is much smaller, the number of companies in a strata decrease, and often sampling fractions can no longer be ignored in computation of variance estimates. The Annual Capital Expenditures Survey (ACES) conducted by the U.S. Census Bureau is one such survey: sampling rates in approximately one-fifth of all non-certainty strata are greater than 0.20.

The ACES estimation procedure is fairly straightforward. Sample weights are adjusted for unit non-response, and these non-response adjusted weights are used for subsequent estimation of (annual) expansion estimates and year-to-year change estimates. The ACES variance estimation procedure is equally straightforward: ACES uses a "simple" delete-a-group jackknife variance estimator (Kott 2001) to produce variance estimates of totals and uses Taylor linearization methods to produce variances of change estimates. The "simple" refers to the replicate weights, which are constructed from the **full-sample** non-response adjusted weights; currently ACES does not replicate their non-response adjustment procedure. This variance estimator is computationally quick and easy and does not require excessive overhead in terms of computer storage, storing only 16 replicate weights per sample unit.

The decision to use simple delete-a-group jackknife variance estimation for ACES was based on research presented in Thompson, Sigman, and Goodwin (2003). Somewhat surprisingly, that study found few statistical benefits in **replicating** the non-response adjustment procedures with those data. I hypothesized that our unexpected results could be caused by a variety of factors, such as high unit response in large company strata, non-negligible sampling fractions in large company strata, an inappropriate choice of variance estimator (with a highly stratified survey, the stratified jackknife might be more appropriate), the method of non-response weight adjustment, or characteristics specific to capital expenditures data.

This paper considers all of the factors above while examining the effects of replicating non-response weight adjustment procedures (the fully replicated procedure) on variance estimates in comparison with corresponding "simple" replicated variance estimates ("shortcut" procedure variances). I first provide background on the ACES design and estimation methodology and briefly introduce a model-assisted interpretation of unit non-response in ACES. The section that follows describes the capital expenditures data characteristics. After discussing the survey design and data characteristics, I provide definitions of the delete-a-group jackknife and the stratified jackknife variance estimates, then present and discuss empirical variance estimates of capital expenditures statistics from three years' of ACES data: the first two data sets (from survey years 2002 and 2003) are the full collection of final tabulated ACES data and the third data set (survey year 2003) contains a mid-survey collection of data. Following the empirical comparisons, I provide some simulation study results that pursue some of the possible issues uncovered in the empirical comparisons, considering **only** the delete-a-group jackknife variance estimator. The paper concludes with a few specific comments about ACES applications and some general observations.

---

[1] 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.

Annual Capital Expenditures Survey (Design And Estimation)

The ACES survey collects data about the nature and level of capital expenditures in non-farm businesses operating within the United States. Respondents report capital expenditures for the calendar year in all subsidiaries and divisions for all operations within the United States. ACES respondents report total capital expenditures, broken down by type (expenditures on Structures and expenditures on Equipment). In all subsequent sections and tables, I refer to these characteristics as Total, Structures, and Equipment.

The ACES universe contains two sub-populations: employer companies and non-employer companies[2]. Different forms are mailed to sample units depending on whether they are 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). In the ACE-1 nomenclature, the stratum label indicates the relative size of the companies, with stratum 2A containing the **largest** non-certainty companies in the industry, decreasing to stratum 2D, which contains the **smallest** non-certainty companies in the industry. Stratum 10 units are certainty units and are not included in variance estimation (see Section 3). Sampling fractions in the large size non-certainty ACE-1 (2A) strata can be quite high, as shown in Table 1. In each survey year, at least sixty percent of the size-class 2A strata have non-negligible sampling fractions. Consequently, the sample units from the large-unit non-certainty strata contribute relatively little to the variance estimates (See Section 3).

Table 1: Distribution of Sampling Fractions ($f_h = n_h/N_h$) in ACE-1 Non-Certainty Strata

		Number of Industry-Level Strata				
Survey Year	Size Strata	$f_h < .20$	$(.20 <= f_h < .5)$	$(.50 <= f_h < 1)$	$f_h \equiv 1$	Total Strata
2001	2A	39	32	28	32	131
	2B	103	18	7	0	128
	2C	123	3	0	0	126
	2D	121	0	0	0	121
2002	2A	50	28	23	29	130
	2B	101	18	10	0	129
	2C	121	5	0	0	126
	2D	123	0	0	0	123
2003	2A	44	34	25	27	130
	2B	103	15	11	0	129
	2C	124	2	0	0	126
	2D	122	0	0	0	122

In contrast to the ACE-1 design, sampling fractions in all ACE-2 strata are quite low (all less than 0.01).

Table 2 presents the size strata percentage contribution to ACE-1 capital expenditures estimates over all sampled industries, both overall and as a median percentage within industry size-stratum. Regardless of survey year, the certainty cases – which are excluded from the variance estimation – contribute the most to each estimate; the contributions from the non-certainty size strata are relatively equal within industry for both total capital expenditures and equipment.

---

[2] A nonemployer business is one that has no paid employees, has annual business receipts of $1,000 or more ($1 or more in the construction industries), and is subject to federal income taxes. Most nonemployers are self-employed individuals operating very small unincorporated businesses, which may or may not be the owner's principal source of income.

Table 2: Percentage Contribution to ACE-1 Survey Estimates by Size Stratum

Survey Year	Size Strata	Over All Industries			Median Percentage within Industry Size-Stratum		
		Total	Structures	Equipment	Total	Structures	Equipment
2001	10	0.74	0.69	0.77	0.66	0.72	0.63
	2A	0.07	0.09	0.06	0.08	0.08	0.09
	2B	0.07	0.08	0.07	0.10	0.07	0.10
	2C	0.07	0.08	0.06	0.09	0.04	0.07
	2D	0.05	0.05	0.05	0.03	0.01	0.04
2002	10	0.73	0.68	0.76	0.65	0.72	0.64
	2A	0.08	0.11	0.06	0.09	0.08	0.08
	2B	0.07	0.09	0.07	0.08	0.06	0.09
	2C	0.07	0.06	0.07	0.08	0.04	0.08
	2D	0.06	0.06	0.05	0.04	0.01	0.04
2003	10	0.67	0.38	0.68	0.60	0.48	0.57
	2A	0.09	0.20	0.09	0.10	0.03	0.09
	2B	0.07	0.09	0.07	0.09	0.00	0.09
	2C	0.09	0.22	0.08	0.08	0.00	0.09
	2D	0.08	0.11	0.08	0.06	0.00	0.07

ACES uses adjustment-to-sample models described in Kalton and Flores-Cervantes (2002) to reduce bias due to unit non-response, 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 administrative payroll data for units in the weighting cell divided by the sum of the sample-weighted administrative payroll data for all responding units in the weighting cell. Under complete non-response in an industry's certainty stratum or in the large company (2A) 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. In general, stratum collapsing is a very rare occurrence and is hereafter ignored in this paper. I refer to the ACE-1 weight adjustment procedure as the **ratio-adjustment** procedure.

Ignoring collapsing, the non-response adjusted ACE-1 estimator is a separate ratio estimator. Let y_{hj} be the value of the survey-data item for sampled unit j in strata h, x_{hj} be the corresponding auxiliary-variable data (known for all sample units). With unit non-response, the sampled units subdivide into two disjoint sets: respondents (complete data observed) and non-respondents (only auxiliary data observed). In a sample of n units, there are r respondents and $nr = (n\text{-}r)$ non-respondents. The ACE-1 ratio estimator assumes the following model

$$y_{hj} = x_{hj}\beta_h + \varepsilon_{hj},\ \ \varepsilon_{hj} \sim (0,\sigma_h) \tag{2.1}$$

where x_{hj} is the value of payroll for sampled unit j. The estimator of characteristic $\hat{Y}_{ACE-1}$ is given by

$$\sum_h \left(\frac{\sum_{j\in h} W_h x_{hj}}{\sum_{j\in h} W_h x_{hj} I_{hj}} \right) \sum_{j\in h} W_h y_{hj} I_{hj} = \sum_h \left(\frac{\sum_{j\in h} W_h x_{hj}}{\sum_{r\in h} W_h x_{hr}} \right) \sum_{r\in h} W_h y_{hr} = \sum_h \left(\frac{N_h}{n_h}\right)\left(\sum_{j\in h} x_{hj}\right)\left(\frac{\bar{y}_{hr}}{\bar{x}_{hr}}\right) = \sum_h N_h \left(\frac{\bar{y}_{hr}}{\bar{x}_{hr}}\right)\bar{x}_h \tag{2.2}$$

where I_{hj} is a unit-response indicator variable for unit j, $W_h = N_h/n_h$, and an "hr" index indicates the summation for all r respondents in a given strata.

The ACE-1 estimator is equivalent to the simple regression imputation estimator:

$$\hat{Y}_{ACE-1} = \sum_h N_h \left(\frac{\bar{y}_{hr}}{\bar{x}_{hr}}\right)\bar{x}_h = \left[\sum_h W_h x_{hr} \left(\frac{\bar{y}_{hr}}{\bar{x}_{hr}}\right)\right] + \left[\sum_h W_h x_{h,nr} \left(\frac{\bar{y}_{hr}}{\bar{x}_{hr}}\right)\right] \tag{2.3}$$

where $\bar{y}_{hr}$ and $\bar{x}_{hr}$ are the sample means of the **respondent** and auxiliary data from the respondent units, x_{hr} is the total auxiliary data for respondent units, and $x_{,h,nr}$ is the total auxiliary data for non-respondent units. Thus, the variance of the ACE-1 capital expenditures estimates has two components, one for the respondent data and one for the imputed portion.

The ACE-2 estimates are controlled to sample counts within strata; that is, the non-response weighting adjustment factor for a weighting cell is calculated as the number of sampled units in the weighting cell divided by the number of responding units in the weighting cell, i.e., the unweighted cell inverse response rate, as recommended by Vartivarian and Little (2002). I refer to the ACE-2 weight adjustment procedure as the **count adjustment**. The estimator of characteristic $\hat{Y}_{ACE-2}$ is given by

$$\sum_h \left(\frac{\sum_{j \in h} I_{hj}}{\sum_{r \in h} I_{hr}} \right) \sum_{r \in h} W_h y_{hr} = \sum_h \left(\frac{N_h}{n_h} \right) n_h \left(\frac{y_{hr}}{r_h} \right) = \sum_h N_h \bar{y}_{hr} \tag{2.4}$$

where I_{hj} is a sample-inclusion indicator variable and I_{hr} is the unit-response indicator variable defined above. Again, this estimator has a response data component and an imputed data component:

$$\hat{Y}_{ACE-2} = \sum_h N_h \bar{y}_{hr} = \left[\sum_h W_h y_{hr} \right] + \left[\sum_h W_h (n_h - r_h) \bar{y}_{hr} \right] \tag{2.5}$$

See Caldwell (1999b) for more details on the ACE-1 and ACE-2 non-response weighting adjustment procedures.

A model-assisted perspective on non-response considers the "selection" of non-responding units as another phase of sample selection from *g* disjoint response homogeneity groups (RHG)[3], where the true response distribution exists but is unknown (Särndal, Swenssen, Wretmen, 1992, Ch.15 and Kott 1994). The literature refers to this model as the quasi-randomization model since the second phase of sample selection depends on an unknown response distribution. Under quasi-randomization models, the ACE-1 and ACE-2 stratified SRS-WOR samples are the first stage of sample selection, the ACE-1 and ACE-2 strata are the response homogeneity groups (RHG), and the respondents are Bernouilli samples within RHG. In this framework, the final-weighted ACE-1 estimates are two-phase sample regression (ratio) estimators and the final-weighted ACE-2 estimates are two-phase sample estimates, with sampling weight equal to N_h / r_h ($=N_h/n_h \times n_h/r_h$), i.e., are Horvitz-Thompson estimates based on sample respondents. Kott and Stukel (1997) provide evidence that the stratified jackknife variance estimator yields conditionally unbiased variance estimates (with respect to the assumed model) from a two-phase sample regression estimator (c.f. the ACE-1 ratio adjustment estimator) if the sampling is performed with replacement at the first stage, but **not** for the two-phase sample estimator (c.f. the ACE-2 count adjustment estimator). A design-based perspective for these same estimators requires that the variance estimate include components for the variance from responding units, variance due to imputation, and covariance of the two terms.

Wolter (1985, Ch.2, pp. 31-32) provides guidelines for dividing the parent sample of either a two-stage or two-phase sampling into random groups for variance estimation, advocating allocation of **entire** first stage units (not allocation of ultimate sampling units or respondents) to random groups. The applications described follow these guidelines.

Characteristics Of Capital Expenditures Data

Capital expenditures data are fairly atypical business data. First, they often are characterized by low year-to-year correlation for the same reporting unit: for example, a company that spends a large amount of capital expenditures on structural (building) improvements one year is unlikely to invest much in structural improvements in the following year. Second, they are often poorly correlated with auxiliary data such as payroll or receipts, especially for small companies. Finally, company purchasing patterns are more correlated with the size of the company than the sample industry. For example, in some industries, capital expenditures on structures and equipments are negatively correlated for small companies and positively correlated for large companies. For these reasons, the ACES survey does **not** perform imputation for **item** non-response.

---

[3] Units in RHG groups have the same response probability, with data missing at random within RHG group.

In addition, while the ACE-1 and ACE-2 data are characterized by fairly high unit respondent rates, a reported zero value for most data items is generally legitimate, and the item reported-zero rate can be quite high. Table 3 presents median unit response rates within industry size-class strata (ACE-1) or strata (ACE-2), along with median unweighted item reported-zero rates for the three variables of interest. Note that the item reported-zero rate for total capital expenditures is often much smaller than the corresponding rate for the other two items: a company may not have the expenditures data available in the requested breakdown and may report only a total value.

Table 3: Median Unit Response Rate and Item Reported Zero Rates

Survey Year	Population	Size Strata	Unit Response Rate	Item Reported-Zero Rates		
				Total	Structures	Equipment
2001	ACE-1	10	0.88	4.41	20.42	5.08
		2A	0.83	4.84	25.11	6.13
		2B	0.84	5.19	90.89	5.27
		2C	0.80	9.94	43.7	10.61
		2D	0.71	10.81	51.72	12.15
	ACE-2	All Strata	0.80	10.84	97.69	12.13
2002	ACE-1	10	0.87	16.61	63.26	18.28
		2A	0.84	18.08	72.39	18.92
		2B	0.83	17.68	100.00	20.36
		2C	0.78	31.64	84.72	32.29
		2D	0.71	31.48	85.86	31.71
	ACE-2	All Strata	0.72	31.06	100.00	33.84
2003	ACE-1	10	0.92	57.85	96.48	59.02
		2A	0.74	59.00	95.62	60.05
		2B	0.75	55.96	100.00	58.99
		2C	0.71	59.45	91.81	61.44
		2D	0.66	71.01	97.19	72.94
	ACE-2	All Strata	0.58	66.55	91.15	69.16

For ACE-1, the item reported-zero rate tends to increase as the size of the company decreases, particularly for structures. ACE-2 companies rarely report capital expenditures on structures. Notice that all of the item reported-zero rates increase each year, and that the item reported-zero rates are extremely high for all items and strata in the 2003 survey year. Recall that the 2003 ACES data are an incomplete collection (the 2001 and 2002 data sets comprise the complete survey collection), and consequently, the 2003 rates reported in Table 3 are very different from those in the final data sets.

Variance Estimation Methodology

I consider two variance estimators: the delete-a-group jackknife variance estimator and the stratified jackknife variance estimator. To perform delete-a-group jackknife variance estimation, the non-certainty portion of the survey sample is divided into *K* random groups. The delete-a-group jackknife replicate estimate is then computed for each replicate *k* by removing the k^{th} random group from the full sample and multiplying each replicate estimate by *K/(K-1)*. Note that these replicate factors differ from those presented in Kott (2001), which recommends developing stratum-specific replicate factors from sampled units in stratum *h* assigned to delete-a-group jackknife replicate *k*. The delete-a-group jackknife variance estimator used in this paper is equivalent to the random group variance estimator for Horvitz-Thompson estimates (total estimates) and is an unconditionally unbiased estimate of the true variance, given a systematic assignment of sample units to random group (see Appendix). 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 (which may not be equal to 1 with a non-response adjustment). 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 themselves. The delete-a-group variance for any estimate $\hat{\theta}_i$ is

$$\hat{v}_{DAG}(\hat{\theta}_i) = \frac{K-1}{K}\sum_k \left(\hat{\theta}_{ki(DAG)} - \hat{\theta}_{0i}\right)^2 \tag{4.1}$$

where $\hat{\theta}_{ki(DAG)}$ is the random group k replicate estimate and $\hat{\theta}_{0i}$ is the full-sample estimate. ACES uses $K = 15$ random groups. To account for the non-negligible sampling fractions, I multiply all replicate weights by $\sqrt{1-f_h}$ as proposed by Wolter (1985, Ch. 2, pp. 36). Thus in my applications, the $\hat{\theta}_{0i}$ and $\hat{\theta}_i$ (the tabulated full-sample estimate) are not equivalent.

Kott (2001) states that appropriate use of the delete-a-group jackknife requires that the number of sample units per first-phase stratum be large in **all** strata. This requirement is always met in the ACE-2 sample. However, it is sometimes violated in the ACE-1 strata; for example, in the 2001 survey year design, 11 of the 506 designated non-certainty ACE-1 strata contain less than 15 sampled units. ACES does not employ the extended delete-a-group jackknife procedure proposed for these small strata described in Kott (2001).

The stratified jackknife variance estimator uses considerably more replicates than the delete-a-group jackknife variance estimator. Replicate estimates for the stratified jackknife are constructed by dropping one observation at a time from the strata and multiplying the remaining observations by $n_h/(n_h - 1)$. Thus, the stratified jackknife procedure constructs one replicate estimate per sample unit. The stratified jackknife variance estimator is given by

$$\hat{v}_{STRJ}(\hat{\theta}_i) = \sum_{h=1}^{H}(1-f_h)\left(\frac{n_h - 1}{n_h}\right)\sum_{r=1}^{n_h}(\hat{\theta}_{rhi} - \hat{\theta}_{hi})^2 \tag{4.2}$$

The purpose of this study was to investigate the effect of replicating the non-response adjustment procedure on ACE-1 and ACE-2 variance estimates. I expected that the "shortcut" approach used by ACES should **underestimate** the true variance. There are, however, some arguments against this. The shortcut variance estimator is **not** the naive variance estimator, which treats imputed values as though they were reported values and which has been shown repeatedly to underestimate the variance. The naive variance estimator replaces the missing item responses with imputed values, yielding a dataset with no (visible) item non-response, and the replicate values do not contain any missing values. My replicate assignment procedures assign **sample** units to replicates. Consequently, replicate estimates contain both responding and non-responding units, thus incorporating second-phase effects (model-assisted interpretation) or non-sampling error effects (design-based interpretation). The difference between the shortcut and fully replicated procedures are that the shortcut method estimates **conditional** variance (conditioned on the sample respondents), whereas the fully replicated procedure estimates an unconditional variance for the model-assisted interpretation of two-phase sampling. The fully replicated weight adjustment procedure has varying strata weight adjustment factors by replicate and is conditionally model unbiased (Kott and Stukel, 1997). A derivation of the conditional expectation of the shortcut procedure variance is beyond the scope of this paper.

Intuition notwithstanding, the shortcut procedure could overestimate the variability induced by the ACE-1 ratio estimation procedure, since it treats the ACE-1 estimate as an expansion estimator and does not include covariance between numerator and denominator in calculations (the fully replicated procedure accounts for this). Note that Yung and Rao (1996) found that using a shortcut procedure with a post-stratified estimator from a multi-stage stratified design with the stratified jackknife severely **overestimated** the variance.

There is limited support in the literature for this shortcut approach. Wolter (1985, pp. 83-84) cites results from two studies that showed the slight improvements in random group variance estimates using full replication 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. Finally, Canty and Davison

(1999) found better statistical properties with a partially replicated procedure on a post-stratified estimator (one iteration of a multi-stage ratio raking procedure) using the stratified jackknife than with the fully replicated procedure.

In the following section, I use "DAG" to indicate delete-a-group jackknife and "STRJ" to indicate stratified jackknife, combined with "S" (simple/shortcut) and "R" (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) as are the non-certainty cases in size-strata 2A (within industry) that have sampling rates equal to 1.

Empirical Data Results

This section presents empirical comparisons of standard error estimates using the delete-a-group jackknife and the stratified jackknife with and without replicating the non-response adjustments. Within variance estimation method, Table 4 compares the fully replicated standard errors to shortcut standard errors. Although the ACE-1 component requires a variance estimation adjustment due to non-negligible sampling fractions in many strata, I present results both with and without the fpc-adjustment to determine whether the fpc-adjustments have impact on the effects of replicating the weight adjustment procedure.

Table 4: Ratios of Reweighted/Simple Standard Errors Within Variance Estimation Method

Survey Year	Data Item	ACE-1				ACE-2			
		DAGR/DAGS		STRJR/STRJS		DAGR/DAGS		STRJR/STRJS	
		FPC	No FPC	FPC	No FPC	FPC	No FPC	FPC	No FPC
2001	Total	0.97	1.00	0.98	0.98	1.04	1.04	0.99	0.99
	Structures	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00
	Equipment	0.92	0.92	0.98	0.98	1.04	1.04	0.99	0.99
2002	Total	0.99	1.00	0.97	0.97	1.03	1.03	0.99	0.99
	Structures	1.00	1.01	0.98	0.98	0.99	0.99	1.00	1.00
	Equipment	0.90	0.91	0.96	0.96	1.08	1.08	0.99	0.99
2003	Total	1.02	1.02	0.99	0.98	0.97	0.97	1.00	1.00
	Structures	1.04	1.04	1.00	1.00	1.02	1.02	1.00	1.00
	Equipment	0.99	1.00	0.99	0.98	0.98	0.98	1.00	1.00

Within replication method (delete-a-group jackknife or stratified jackknife), the full replication effects are generally the same regardless of fpc correction. This result is expected: most of the ACE-1 unit non-response occurs in the smaller company size strata, which have negligible sampling fractions; and all of the ACE-2 sampling fractions are negligible. In fact, the fpc corrections to the stratified jackknife standard error estimates are generally ignorable. Notice that the full replicate standard error estimates are very close to the shortcut standard error estimates for capital expenditures on structures, the characteristic with the highest proportion of reported zero values in all years. Also, the delete-a-group and stratified jackknife ACE-1 standard error estimates of total capital expenditures and equipment **decrease** when the full replication procedure is used for the 2001 and 2002 survey years (the two high **unit**-response years). This pattern continues for the stratified jackknife standard error estimates with the 2003 data, but changes for the delete-a-group jackknife standard error estimates, perhaps further illustrating the effect of high item non-response.

In contrast to the ACE-1 estimates, the effects of full replication on ACE-2 standard error estimates are quite different for the two variance estimators: the stratified jackknife standard errors are virtually unchanged, whereas almost all of the 2001 and 2002 delete-a-group jackknife standard errors increase slightly with full replication. Given the relatively low unit response rates coupled with the high reported-zero rate for the ACE-2 samples, it seems unlikely that the full replication would have little effect on the standard error estimates, c.f., the stratified jackknife estimates.

Table 5 assesses the effect of the fpc-correction on standard error estimates by presenting standard error ratios within variance estimation method. [Note: ACE-2 sampling fractions are all less than 0.01, so all standard error ratios are 1, as expected].

Table 5: Standard Error Ratios (Without FPC-Correction/With FPC Correction)

Survey Year	Data Item	ACE-1				ACE-2			
		Delete-a-Group Jackknife		Stratified Jackknife		Delete-a-Group Jackknife		Stratified Jackknife	
		Simple	Reweighted	Simple	Reweighted	Simple	Reweighted	Simple	Reweighted
2001	Total	1.05	1.07	1.14	1.14	1.00	1.00	1.00	1.00
	Structures	1.03	1.03	1.19	1.18	1.00	1.00	1.00	1.00
	Equipment	1.09	1.08	1.10	1.10	1.00	1.00	1.00	1.00
2002	Total	1.02	1.03	1.03	1.03	1.00	1.00	1.00	1.00
	Structures	1.03	1.03	1.03	1.03	1.00	1.00	1.00	1.00
	Equipment	1.06	1.07	1.04	1.04	1.00	1.00	1.00	1.00
2003	Total	1.27	1.26	1.23	1.23	1.00	1.00	1.00	1.00
	Structures	1.11	1.11	1.10	1.10	1.00	1.00	1.00	1.00
	Equipment	1.33	1.33	1.30	1.30	1.00	1.00	1.00	1.00

Regardless of variance estimation method, the fpc-correction reduces the ACE-1 standard error estimates. This effect is most pronounced with total capital expenditures and equipment. This is in line with the data characteristics and the fpcs, i.e., large fpcs in strata with positive (non-zero) capital expenditures estimates and small fpcs in strata with high item non-response rates (small or approximately zero expenditures estimates). Table 6 compares the size of corresponding stratified jackknife and delete-a-group jackknife standard error estimates.

Table 6: Ratios of Stratified Jackknife Standard Error estimates/Delete-a-Group Jackknife Standard Error Estimates

Survey Year	Data Item	ACE-1				ACE-2			
		STRJS/DAGS		STRJR/DAGR		STRJS/DAGS		STRJR/DAGR	
		FPC	No FPC	FPC	No FPC	FPC	No FPC	FPC	No FPC
2001	Total	1.07	1.17	1.08	1.15	1.00	1.00	0.96	0.96
	Structures	1.18	1.36	1.17	1.34	0.98	0.98	0.98	0.98
	Equipment	0.82	0.83	0.87	0.89	0.83	0.83	0.79	0.79
2002	Total	1.11	1.12	1.09	1.09	1.00	1.00	0.97	0.97
	Structures	1.28	1.28	1.25	1.24	1.01	1.01	1.01	1.01
	Equipment	1.04	1.02	1.11	1.07	0.99	0.99	0.90	0.90
2003	Total	0.97	0.94	0.94	0.92	1.09	1.09	1.11	1.11
	Structures	0.87	0.87	0.84	0.84	1.33	1.33	1.30	1.30
	Equipment	1.01	0.99	1.00	0.98	0.73	0.73	0.74	0.74

Again, the 2001/2002 survey year results are quite different from the 2003 survey year results for ACE-1, with the stratified jackknife standard error estimates generally larger in the first two years (complete data collections), and smaller or the same in 2003 (incomplete data collection). The ACE-2 ratios are quite variable, again perhaps showing the effect of the high item reported-zero rates on the different variance estimators.

To gauge the effect of the non-response weight adjustment procedure on the ACE-1 standard error estimates, I computed ACE-1 variances using the count adjustment procedure (the ACE-2 non-response adjustment procedure). Table 7 presents these results. Again, there is very little difference between shortcut and full replication estimates within the same variance estimation method. When the count adjustment procedure is used instead of the ratio adjustment procedure, the stratified jackknife standard error estimates are much smaller than the corresponding delete-a-group jackknife standard error estimates.

So far, the empirical results show the following:

- The effects of full replication are not affected by the fpc-adjustment, for either ACE-1 or ACE-2;

- Regardless of variance estimation method, shortcut procedure standard errors are as likely to be **smaller** than corresponding fully replicated standard errors as **larger** (neither procedure yields consistently larger or smaller standard error estimates);
- Corresponding shortcut and full replication standard error estimates are very close in value;
- For characteristics with high item response rates, the ACE-1 ratio adjustment procedure (for unit non-response) generally yields larger stratified jackknife standard error estimates than the corresponding delete-a-group jackknife standard error estimates. The stratified jackknife standard error estimates may more correctly reflect the variability due to the **ratio** imputation. However, the stratified jackknife standard error estimates for ACE-1 are much smaller than the corresponding delete-a-group jackknife variance estimation when the count adjustment procedure (**mean** imputation) is used;
- Finally, there appears to be an interaction between the effects of replicating the non-response weight adjustment procedure and the unit response rates. Because ACES has high unit response rates in all strata, it is impossible to assess the effect of unit response rate on variance estimates from these data sets.

Table 7: ACE-1 Standard Error Ratios Using Ratio-to-Sample Count Non-Response Adjustment

Survey Year	Data Item	Reweighted/Simple (Same Variance Estimator)				Stratified Jackknife/Delete-a-Group Jackknife			
		DAGR/DAGS		STRJR/STRJS		STRJS/DAGS		STRJR/DAGR	
		FPC	No FPC	FPC	No FPC	FPC	No FPC	FPC	No FPC
2001	Total	0.98	1.02	1.00	1.00	0.59	0.65	0.59	0.64
	Structures	1.00	1.00	1.00	1.00	0.77	0.90	0.77	0.90
	Equipment	0.96	0.98	0.99	0.99	0.47	0.47	0.48	0.48
2002	Total	1.00	1.00	1.00	1.00	0.65	0.67	0.65	0.67
	Structures	1.00	1.01	1.00	1.00	0.70	0.72	0.70	0.72
	Equipment	0.92	0.93	0.98	1.02	0.74	0.73	0.79	0.77
2003	Total	1.04	1.03	1.06	1.06	0.63	0.63	0.64	0.65
	Structures	1.02	1.02	1.03	1.03	0.62	0.62	0.63	0.63
	Equipment	1.02	1.02	1.06	1.07	0.64	0.65	0.66	0.68

ACES uses the delete-a-group jackknife variance estimator, as does a variety of other business surveys administered by the Census Bureau. The empirical results shown above provide some evidence that the stratified jackknife variance estimator might be more appropriate for the ACE-1 component of the survey, **if** the current non-response weight adjustment procedure is retained. On the other hand, there is very little to recommend the stratified jackknife variance estimator for ACE-2, or for ACE-1 if the non-response weight adjustment procedure is changed to an inverse response-rate procedure. Furthermore, the confusing results vis-à-vis the fully replicated versus shortcut standard errors are not ameliorated by the choice of variance estimator.

The Census Bureau's Standardized Economic Processing System (StEPS) software offers delete-a-group jackknife variance estimation, but does not include the stratified jackknife. Additional testing of computing resources would be necessary before adding this variance estimator to StEPS. In the meantime, there are several studies that report excellent results with the delete-a-group jackknife on a variety of sample designs (including stratified SRS-WOR) for expansion, ratio, and restricted regression estimators (e.g., Kott (1998), Kott (2001) and Smith (2001)).

The empirical results point to factors that might account for the differences in the shortcut and fully replicated standard errors, namely percentage unit non-response, item reported-zero rates, and choice of non-response weight adjustment procedure. The next section explores some of these factors via simulation studies, using only the delete-a-group jackknife variance estimator.

Simulation Study Results

The simulation studies described below assume a two-phase design, where the first phase is a SRS-WOR of units from a frame, and the second phase uses a Bernouilli sampling mechanism to assign response status.

The empirical results indicated a possible interaction effect between unit and item response rates and full replication on standard error estimates. To assess this, I simulated twelve different populations: six normally distributed populations (similar to income measurements) and six gamma distributed populations (similar to payroll or employment data). Within distribution, three of the six populations had "small" population c.v.'s (less than 0.50) and three had "large" population c.v.'s (greater than 1.4), and each of the three populations within c.v. category had different levels of **item** reported-zero rate (90%, 70%, and 50%). Then, I selected **three** sets of 5,000 SRS-WOR samples from each population using a 1-in-20 sampling rate, and in each set, I randomly assigned **unit** non-response within sample using a missing-at-random model with the following response proportions – 90%, 70%, and 50%. In 1,000 of the 5,000 samples, I assigned sample units to 15 random groups and computed fully replicated and shortcut procedure variance estimates using the count adjustment (inverse response rate) procedure for unit non-response. To summarize, I ran 18 simulations per population, systematically varying patterns of unit response and item reported-zero rates.

To examine the statistical properties of the shortcut and full replication procedures, I used my 5,000 random samples within population *i* and response proportion *j* to construct empirical variances as

$$V(\hat{X}_{ij}) = \frac{\sum_{r=1}^{5,000}(\hat{X}_{rij} - \overline{X}_{ij})^2}{5,000} \tag{6.1}$$

where $\hat{X}_{rij}$ is the estimate from sample *r*, and $\overline{X}_{ij}$ is the mean of the $\hat{X}_{rij}$.

Then, I calculated two variance estimates (v_{meth}) per sample *r* from 1,000 of the 5,000 samples. I compared these variance estimates in terms of

Relative Bias

$$\frac{\frac{1}{1,000}\sum_{r=1}^{1,000} v_{meth}(\hat{X}_{rij})}{V(\hat{X}_{ij})} - 1 \tag{6.2}$$

MSE

$$\frac{1}{1000}[\sum_{r=1}^{1000}(v_{meth}(\hat{X}_{rij}) - V(\hat{X}_{ij}))^2] + (V(\hat{X}_{ij}) - \sigma_{ij}^2)^2 \tag{6.3}$$

where σ_{ij}^2 is the population variance. Table 8 presents the results from this first analysis. For consistency with the earlier sections, fully replicated delete-a-group jackknife results are labeled "DAGR" and shortcut procedure delete-a-group jackknife results are labeled "DAGS."

Both sets of biases are generally quite comparable, as are the MSEs (in fact, the two sets of MSEs are generally indistinguishable). Within the gamma-distribution generated populations, the fully replicated variances are usually slightly **less biased** than the shortcut procedure variances (in 15 out of 18 simulations). Within the normal-distribution generated populations, in two-thirds of the simulations, this pattern is repeated. Here, the shortcut standard errors appear to be less biased when the unit response rates are low; population c.v.'s and item response rates do not seem to affect the outcome. Capital expenditures data resemble the simulated data in the gamma-distributed (plus item response) populations, so these results provide some supporting evidence for the reduction in variance levels with full replication.

Table 8: Combined Effects of Unit Non-response, Item Reported-zero Rate, and Population Distribution

Population Distribution	CV Category	Unit Response Rate	Item Reported-Zero Rate	Relative Bias		Ratio DAGS /DAGR MSE
				DAGS	DAGR	
Normal	Large	90	90	0.0215	0.012	1.00
			70	0.0383	0.0276	1.00
			50	-0.1302	-0.145	1.00
		70	90	0.1551	0.1399	0.99
			70	0.0797	0.07	0.99

			50	-0.1379	-0.1404	1.00
		50	90	0.0467	0.0244	1.00
			70	0.0469	0.0314	1.00
			50	-0.0872	-0.0905	1.00
	Small	90	90	0.0817	0.0795	1.00
			70	0.0544	0.0523	1.00
			50	-0.0327	-0.0348	1.00
		70	90	0.058	0.0534	1.00
			70	-0.0082	-0.0117	1.00
			50	0.0825	0.0803	1.00
		50	90	0.0298	0.0237	1.00
			70	-0.0568	-0.0595	1.00
			50	0.0102	0.0054	1.00
Gamma	Large	90	90	-0.0619	-0.0775	1.00
			70	0.0893	0.0788	1.00
			50	0.0716	0.0647	0.99
		70	90	-0.0273	-0.0481	1.00
			70	0.1338	0.1171	0.99
			50	0.0382	0.0311	1.00
		50	90	0.0505	0.0175	0.99
			70	0.0557	0.0363	0.99
			50	0.0138	0.0033	0.99
	Small	90	90	-0.0635	-0.0672	1.00
			70	0.0821	0.0796	1.00
			50	0.0319	0.0306	1.00
		70	90	0.0573	0.053	1.00
			70	0.0565	0.0522	1.00
			50	0.0938	0.0914	1.00
		50	90	0.0715	0.0662	1.00
			70	0.0759	0.0719	1.00
			50	0.0389	0.0367	1.00

With the empirical comparisons, the effects of full replication for the ACE-1 variance estimates differed depending on the choice of non-response weight adjustment procedure. Unable to draw a direct link between unit non-response, item reported-zero rate, and the statistical properties of the shortcut and full replicated procedures, I hypothesized that the differences could be a combined effect of the weight adjustment procedure and the unit response rate. In addition, the empirical results obtained using a ratio non-response adjustment procedure could depend on the correlation between payroll and the capital expenditures characteristic. My second simulation study examines the effect of correlation between characteristic and auxiliary variable in conjunction with unit response rates. These simulations use three bivariate lognormal populations with differing levels of correlation between the two variables (ρ =90%, ρ = 70%, and ρ = 60%). As in the other simulation study, I selected three sets of SRS-WOR samples from the three populations and assigned MAR unit response in each sample, then computed shortcut and fully replicated variance estimates using both non-response weight adjustment procedures. Table 9 presents these results.

Table 9: Comparison of Variance Properties Considering Correlation Between Item and Auxiliary Data and Unit Response Rate

Population	Unit Response Rate	Weight Adjustment	Bias of Shortcut SEs	Bias of Fully Replicated SEs	Ratio of Shortcut MSE to Fully Replicated MSE
ρ = 90%	90	Ratio	0.013	-0.021	1.01
		Count	0.0122	-0.0211	1.01
	70	Ratio	0.101	0.0556	1.02
		Count	0.0986	0.0556	1.02
	50	Ratio	0.1574	0.078	1.05
		Count	0.1532	0.078	1.04
ρ = 70%	90	Ratio	0.0181	-0.0197	1.01
		Count	0.0174	-0.0197	1.01
	70	Ratio	0.0905	0.04	1.02
		Count	0.0883	0.0401	1.02
	50	Ratio	0.1377	0.0709	1.03
		Count	0.1337	0.0707	1.03
ρ = 60%	90	Ratio	0.1129	0.0706	1.02
		Count	0.1123	0.0706	1.02
	70	Ratio	0.073	0.0276	1.03
		Count	0.0714	0.0276	1.02
	50	Ratio	0.1159	0.0462	1.03
		Count	0.1131	0.0461	1.03

In all three populations, the biases of the ratio-adjusted and count-adjusted variance estimates are very close within replication procedure (full versus shortcut) and item response rate category. In all three, the bias of the shortcut procedure variance estimates is generally at least twice as large as that of the fully replicated variance estimates when the response rate is less than 90%. In the high response rate scenarios (response rates = 90%) in high and medium correlation populations, the fully replicated procedure variance estimates are **negatively** biased and the shortcut procedure variance estimates are **positively** biased, and the magnitude of the two sets of biases is about the same. In the low correlation population, both procedures overestimate the variance, but the shortcut biases are usually twice as large as the fully replicated procedure (consistent with the results presented in Table 8). Here again, the shortcut MSEs are always larger than the corresponding fully replicated MSEs.

Both simulation studies presented show one consistent relationship between the bias of the shortcut procedure variance estimates and the unit response rate, namely that the shortcut variance estimates are **overestimates** when the unit response rate is low (50%), but are often comparable to the fully replicated procedure variance estimates when the unit response rates are high (70% or greater). The two sets of one-strata sample simulation study results show that the fully replicated procedure variances are often (slightly) less biased than corresponding shortcut procedure variance estimates, but that this frequency is highly dependent on the underlying population distribution. The effect of underlying population distribution is further illustrated by the second set of simulation results (with stratified sampling), where the different strata population distributions clearly have an aggregated effect on the level of the biases for the shortcut and fully replicated procedures.

Conclusion

This study examines in detail differences between variance estimates that fully replicate a non-response weight adjustment procedure and those that use a shortcut procedure. The empirical comparison considers the delete-a-group jackknife variance estimator and the stratified jackknife. With both variance estimators, I found very few differences between the fully replicated and shortcut procedure variance estimates, regardless of weight adjustment method. Somewhat surprisingly, I also found that the fully replicated standard errors were often smaller than the corresponding shortcut procedure standard errors. This finding was reinforced by my simulation study results, where the fully replicated variance estimates were often less biased than those from the corresponding shortcut procedure. Moreover, the MSE of the fully replicated variance estimates was almost always less than that of the shortcut procedure variance estimates, providing further evidence of the "closer to unbiasedness" properties of the fully replicated variance estimates.

The empirical results indicated a relationship between the combined percentage of unit and item reported-zero rates and the size of the full replicated and shortcut procedure variance estimates. The simulation study results confirmed this, demonstrating underestimation with the shortcut procedure under very low unit response rates and overestimation with the shortcut procedure otherwise. It is worth noting that when response rates are high, the biases of the shortcut and fully replicated standard errors were comparable.

In interpreting the empirical results, recall that capital expenditures data are somewhat atypical from other business data in their proportion of legitimate zero responses and that consequently, small differences in variance estimate level may appear as "large" ratio differences (and may yield equivalent coefficients of variation with all standard error estimates). Moreover, the ACES unit response rates are quite high in all strata, which may explain the generally small differences between corresponding shortcut and fully replicated standard errors.

The simulation study results attempt to quantify the empirical results using a model-assisted interpretation of unit non-response. To guard against this model assumption, I use a variety of response propensities. However, the view of unit non-response as a second stage of sampling is not necessarily realistic for a voluntary economic survey: it is equally likely that several non-respondents are **fixed** in the population, so that the unit non-response is an estimation-bias problem, not a variance estimation problem.

If there is a moral to this research, it is that our intuition can often lead us astray. Without evaluation, I assumed that the shortcut procedure variance estimates **must** always be smaller than their fully replicated counterparts. Upon reflection, there is no reason to assume that this must be true, especially when unit response rates are high. In this case, while there are clearly theoretical advantages to fully replicating the weight adjustment procedure, there may be little or no practical advantage. If time is of the essence or computing resources are scarce, the shortcut procedure variances – which are very close to the fully replicated variances for the ACES data – are a quite viable alternative.

Finally, my original suspicion that the effects of full replication were hidden by the fpc-adjustment turned out to be unfounded since the strata used for variance estimation generally have negligible fpc's. This brings up an unaddressed issue. If we truly believe that the unit non-response is another stage of sampling, then we should consider including a "certainty" component (for the certainty strata with unit non-response) in our variance estimates. Figuring out how to do this correctly is, however, the topic for future research.

Acknowledgements

The author thanks Amy Newman-Smith, Rita Petroni, and Richard Sigman for their contributions to this research and for their useful comments on an earlier version of this manuscript, and Phillip Kott for his excellent discussion of this research.

References

Caldwell, C. (1999a). Response Characteristics for the 1997 ACES. Unpublished internal memorandum. Washington, DC: U.S. Bureau of the Census.

Caldwell, C. (1999b). Non-response Weighting 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.

Kalton, G. and Flores-Cervantes, I. (2003). Weighting Methods. *JOS*, **19**, pp. 81-97.

Kott, P. (1994). A Note on Handling Nonresponse in Sample Surveys. *Journal of the American Statistical Association*, **89**, pp. 693 –696.

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.

Kott, P. and Stukel, D. (1997). Can the Jackknife Be Used With a Two-Phase Sample? *Survey Methodology,* **23**, pp. 81-89.

Särndal, C., Swensson, B., and Wretman, J. (1992). Model Assisted Survey Sampling. New York: Springer Verlag.

Schindler, E. (2002). Simplified Variance Estimation for Complex Surveys. *Proceedings of the Section on Survey Research Methods,* American Statistical Association, pp. 3095-3100.

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.

Thompson, K., Sigman, R. and Goodwin, R. (2003). Investigation of Alternative Replicate Variance Estimators for the Annual Capital Expenditures Survey. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 4226-4233.

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

Vartivarian, S. and Little, R.J. (2002). On the Formation of Weighting Adjustment Cells for Unit Non-response. *Proceedings of the Section of Survey Research Methods*, American Statistical Association, pp. 3553-3558.

Yung, W. and Rao, J.N.K. (1996). Jackknife Linearization Variance Estimators Under Stratified Multi-Stage Sampling. *Survey Methodology*, **22**, pp. 23-31.

Appendix

Postulate 1: Given an unbiased systematic assignment of sampled units to K random groups within <u>all</u> strata and delete-a-group jackknife replicate factors of (K/K-1), the delete-a-group jackknife variance estimator is an unbiased estimator of the variance of an estimate of a total.

Let $\hat{\theta} = \sum_h \sum_i w_{hi} y_{hi}$ = a full sample estimate, $\hat{\theta}_{(k)} = \left(\frac{K}{K-1}\right) \sum_h \sum_{i \notin k} w_{hi} y_{hi}$ = the delete-a-group jackknife replicate k estimate, $\breve{\theta}_k = \sum_h \sum_{i \in k} w_{hi} y_{hi}$ = the random-group k estimate, and $\hat{\theta}_k = K\breve{\theta}_k$ = the random-group **replicate** k estimate.

$$
\begin{aligned}
\hat{V}_{DAG}(\hat{\theta}) &= \left(\frac{K-1}{K}\right) \sum_{k=1}^{K} (\hat{\theta}_{(k)} - \hat{\theta})^2 \\
&= \left(\frac{K-1}{K}\right) \sum_{k=1}^{K} \left[\left(\frac{K}{K-1}\right)(\hat{\theta} - \breve{\theta}_k) - \hat{\theta} \right]^2 \\
&= \left(\frac{K-1}{K}\right) \left(\frac{1}{K-1}\right)^2 \sum_{k=1}^{K} (\hat{\theta} - \hat{\theta}_k)^2 \\
&= \left(\frac{1}{K(K-1)}\right) \left(\sum_{k=1}^{K} (\hat{\theta}_k - \hat{\theta})^2 \right) \\
&= \hat{V}_{RG}(\hat{\theta})
\end{aligned}
$$

Wolter (1985), pp. 21, proves that $\hat{V}_{RG}(\hat{\theta})$ is an unbiased estimator of $V(\hat{\theta})$.

Postulate 2: There are two valid approaches to replicating the non-response adjustment procedure for the count estimator.

Consider the full-sample estimate in stratum h for a stratified SRS-WOR design with the count adjustment applied for unit non-response: $\hat{Y}_h = \left(\frac{N_h}{n_h}\right)\left(\frac{n_h}{r_h}\right) \sum_{i \in h} y_{hi} = \left(\frac{N_h}{r_h}\right) \sum_{i \in h} y_{hi} = N_h \bar{y}_{rh}$, where r_h are the respondent units in stratum h.

Approach 1 (used in this paper):

Use replicate factors of (K/K-1) to produce replicate estimates, and replicate non-response adjustment as $\frac{n_{(g)h}}{r_{(g)h}}$.

Advantage:

- Unconditionally unbiased variance estimates, even if weighting cells are collapsed.

Disadvantage:

- Form of replicate estimates is different from full-sample estimator (replicate estimates are **not** Horvitz-Thompson estimates based on sample respondents).

Approach 2 (advocated in Kott 2001)

Use strata-specific replicate factors of $n_h/(n_h - n_{hg})$. In the case of no collapsing of weighting cells, the replicate estimate k is given by

$$\hat{Y}_{h(g)} = \left(\frac{N_h}{n_h}\right)\left(\frac{n_h}{n_h - n_{hg}}\right)\left(\frac{n_h - n_{hg}}{r_h - r_{hg}}\right)\sum_{i \in h \text{ and } i \notin g} y_{hi} = N_h \bar{y}_{r(g)}.$$

Advantage:

- Replicate estimates are Horvitz-Thompson estimates based on sample respondents (in the event of no weighting cell collapse), leading to a "pure" replication of the non-response adjustment procedure.

Disadvantage:

- Variance estimates are not unbiased (follows from postulate 1), although they are conditionally unbiased given the realized sample;
- If weighting cells are collapsed, the form of the replicate estimate is no longer the same as the full-sample estimate.