Investigation of Alternative Variance Estimators for the Quarterly Financial Report

Andromache R. Howe and Katherine J. Thompson[i]
U.S. Census Bureau, Room 3108-4, Washington, DC 20233
katherine.j.thompson@census.gov

Keywords: Delete-a-group jackknife, panel survey

1. Introduction

The Quarterly Financial Report (QFR) is a sample survey of large companies from the mining, wholesale trade, retail trade, and manufacturing sectors. The QFR sample is divided into panels that are rotated into and out of the survey, and each non-certainty sampled company is interviewed for eight consecutive business quarters. For any given quarter, eight panels selected from up to three different frame years are in the survey. Each year, a new sample of corporate tax returns is selected from the most recent tax-year data. Often, the sampling fractions are non-trivial (greater than 0.20). This new sample is split into four panels. Each quarter, one of the four new panels is introduced, and the panel that has completed all eight interviews is dropped from the survey.

It is possible for a QFR company to conduct business in a different industry than indicated by the sampling frame. QFR estimates are tabulated by the company-reported industry (the enumerated industry), not the sample (frame) industry. Estimates of quarterly totals are unweighted means multiplied by an estimate of population size for the enumerated industry/size-classification. This population estimate incorporates both industry changes (from the sampling frame) and the rotation scheme.

Currently, the QFR uses an approximate sampling formula variance estimator that treats the enumerated industry as if it were the sampling industry in all calculations, including applying an estimated finite population correction for the **enumerated** industry. The validity of this variance estimator relies on modeling assumptions about the industry reclassification procedure. Use of replication to estimate the variance of the QFR estimators does not make such assumptions and incorporates the original sample finite population correction factors. Our evaluation considers the delete-a-group jackknife variance estimator, which is employed in a variety of business and household surveys: see Thompson, Sigman, Goodwin (2002), Kott (2001), Kott (1998), Smith (2001), and Bell (2000).

The delete-a-group jackknife is usually applied to survey designs with negligible sampling fractions and more sampled units per stratum than random groups. When these conditions are not met, then the delete-a-group jackknife variance estimator is positively biased. Kott (2001) proposes the extended delete-a-group jackknife variance estimator as a reduced-bias variance estimator for stratified designs where there are several strata that contain fewer sampled units than random groups.

This paper compares the statistical properties of two versions of the delete-a-group and extended delete-a-group jackknife variance estimators to the current approximate sampling formula variance estimates for several QFR estimators.

2. Background

The Quarterly Financial Report (QFR) is a quarterly survey of mining, wholesale trade, and retail trade corporations with total assets of $50 million or more and manufacturing corporations with total assets $250 thousand or more. The QFR collects income statement (e.g. sales, net income, depreciation, etc.) and balance sheet (cash, inventories, current assets, long term debt, retained earnings, total liabilities, etc.) data from each surveyed company. From this data, the QFR publishes several key economic statistics, including quarter-to-quarter percentage change in sales (CHANGE). Other key QFR estimates include estimates of total quarterly sales and total net income after taxes (NIAT), and the quarterly ratio of NIAT/sales (RATIO).

The sampling frame for the QFR survey is developed from the file of United States Internal Revenue System (IRS) corporate tax returns. Every year, the Census Bureau receives a list of corporate tax returns for the previous year from the IRS and classifies all the companies by reported industry (sample industry) and total assets. Companies that have total assets of $250 million or more are included with certainty and are in the survey indefinitely. The remaining companies are stratified within sample industry. Units in the manufacturing sectors are further stratified within sample industry code by size; the within-industry size strata are referred to as the asset classes. The other sectors have one non-certainty stratum per sample industry.

This QFR sample is randomly split into four panels, each of which is introduced in a given quarter. The first panel from this new sample is introduced in the fourth quarter of the sampling year. At this point, companies in the four panels from the previous sample (selected from IRS returns two years prior) are mailed a questionnaire, as are three of the four panels from the previous previous sample. In each quarter, as a new sample panel is introduced, the oldest sample panel (which has completed eight questionnaires) is dropped. So for any given quarter, there are up to three different sampling frames represented. At best, the QFR sample is drawn from sampling frames that are one and two years old. Thus, the QFR sample is subject to coverage bias because of eligible cases not included on the sampling frame.

A QFR company may conduct business in a different industry than indicated on the sampling frame. Classification changes are determined via a nature of business questionnaire, administered **after** sample selection and generally completed by the first interview. The asset classification is rarely changed as a result of survey data. Subject-matter experts refer to enumerated industry "types" as "high mover," "medium mover," or "low mover," depending on the proportion of reclassified sample units. The industry reclassification adds variability to the QFR estimates, since sample sizes in the enumerated industries/asset classes (n_{ki}) are random variables. The original sampling industry/asset class sample sizes (n_{hi}) are fixed values determined by optimal allocation.

The QFR does not use a Horvitz-Thompson estimator to produce estimates of quarterly totals (LEVELS). The formula for a QFR LEVEL estimate of item X in enumerated industry k and asset class i at time t is given by

$$\tilde{X}_{kit} = \left[\frac{\frac{(4-b_{kit})}{Q_{kit}}\hat{N}_{kit}^{(-2)} + \frac{4}{Q_{kit}}\hat{N}_{kit}^{(-1)} + \frac{b_{kit}}{Q_{kit}}\hat{N}_{kit}^{(0)}}{n_{kit}^{(-2)} + n_{kit}^{(-1)} + n_{kit}^{(0)}}\right]\left[\sum_{hi\in ki}\sum_{j\in hki} I_{hkijt}\, x_{hkijt}\right]$$

$$= \left[\frac{\hat{N}_{kit}}{n_{kit}}\right]\left[x_{kit}\right] = \tilde{W}_{kit}\, x_{kit}$$

where $\hat{N}_{kit}^{(0)}$, $\hat{N}_{kit}^{(-1)}$, and $\hat{N}_{kit}^{(-2)}$, are the estimated population sizes at time t in enumerated industry k and asset class i for the sample from the current year frame (0), sample from prior year's frame (-1), and sample from prior-prior year's frame (-2); $n_{kit}^{(0)}$, $n_{kit}^{(-1)}$, and $n_{kit}^{(-2)}$ are the number of sampled cases in currently-interviewed panels at time t in enumerated industry k and asset class i from the (up to) three eligible sample frame years; Q_{kit} is the number of active panels at time t in enumerated industry k and asset class i (usually 8); b_{kit} is the number of active panels in the sample from the corresponding sample years; I_{hkijt} is an indicator variable indicating that company j that was sampled in sampling industry h and enumerated in industry k/asset class i at time t ; and x_{hkijt} is the current data.

The enumerated industry level "weight" $(\tilde{W}_{kit})$ approximates a sampling interval, using a weighted average of population estimates in the numerator and the actual sampled cases in the denominator. The population estimates for the *year-1* and *year-2* samples are Horvitz-Thompson (HT) estimates; the population estimate from the most recent sample frame year are frame totals from the sampling industry and asset class adjusted with survey estimates of in-movers (companies in an enumerated industry that were sampled from a different industry) and out-movers (companies sampled in a different industry than enumerated). An in-mover in one industry is by definition an out-mover in another. The latter estimate also includes an adjustment for number of active panels. We refer to $\tilde{W}_{kit}$ as a "variable weight," and the QFR estimator of LEVELS as a variable-weight estimator, denoted by a tilde (~). The variable weight estimator can also be written as $\hat{N}_{kit}\bar{x}_{kit}$, where $\hat{N}_{kit}$ is the weighted-average population estimate defined above and $\bar{x}_{kit}$ is the unweighted cell mean at time t.

The QFR variable-weight estimates are further adjusted for non-response in the enumerated industry and asset class, using unweighted inverse response rates as advocated by Vartivarian and Little (2002). The response rates for QFR in the large company strata are generally quite high (near 1) and are not discussed further.

The QFR variable-weight estimator has been the subject of several different studies: see Chapman and Biemer (1985) Chapman (1993), Kott (1992), and Caldwell et al (2005). The latter paper uses a Monte Carlo simulation to compare the QFR estimator to a variety of other estimators and concludes that this estimator has the best statistical properties of the considered methods. Caldwell et al (2005) develops reasonable simulated populations that assess the effects of population size-change on the QFR variable weight estimator. This paper uses data from two "low mover" industries, four "medium mover" industries, and one "high" mover industry in the following simulated populations:

- Population 1 – monotone increasing population size by 8% per year within enumerated industry
- Population 2 – monotone decreasing population size by 8% per year within enumerated industry
- Population 3 – monotone increasing population size for eight quarters (8% per year, 16% increase total), followed by monotone decreasing population size ("see-saw") for eight quarters (again, 8% per year, 16% increase total)
- Population 4 – no change in population size within enumerated industry

We compute estimates in **16** of the simulated 60 quarters of data for LEVEL (variable-weight estimator) and RATIO estimates and 15 quarters for the CHANGE estimates. The population models were created with **unrealistically** large increases and decreases in population size to exaggerate the differences between alternative estimators for the LEVEL estimates. Consequently, estimates from these populations should be more biased than expected in the QFR sample data.

The QFR currently uses the following approximate sampling formula variance estimator (S^2 estimator) to estimate LEVEL variances:

$$v_{s^2}(\hat{X}_{kit}) = \hat{N}_{kit}^2\left(1 - \frac{n_{kit}}{\hat{N}_{kit}}\right)\frac{S_{kit}^2}{n_{kit}}$$

where $\hat{N}_{kit}$ and n_{kit} are the enumerated industry/asset class level estimates of population and sample size defined above, and

$$S_{kit}^2 = \frac{\left(\sum_{j=1}^{r_{kit}} x_{kijt}^2 \frac{a_{kit}^2}{r_{kit}^2}\right) - \frac{1}{n_{kit}}\left[\sum_{j=1}^{r_{kit}} x_{kijt}\frac{a_{kit}}{r_{kit}}\right]^2}{n_{kit} - 1}$$

where a_{kit}/r_{kit} is the unweighted inverse response rate of enumerated industry k and asset class i (the a_{kit} are the number of active sampled cases and r_{kit} are the number of respondent cases), and x_{kijt} are the unweighted sample data totals. The original sampling fractions are **not** used in this variance estimator, and it does not include a component for the variability caused by industry reclassification. Moreover, the approximated enumeration industry/asset class sampling fractions in v_{s^2} do not in any way approximate the true sampling fractions. Thus, this S^2 estimator was a known underestimate, although the degree of underestimation had not been previously investigated.

The QFR employs the following Taylor linearization approximation to estimate variances of the non-linear CHANGE and RATIO estimators.

$$\mathrm{var}\left[\frac{\tilde{X}_{kt}}{\tilde{Y}_{kt}}\right] \approx \left[\frac{\tilde{X}_{kt}}{\tilde{Y}_{kt}}\right]^2 \left[\frac{\mathrm{var}(\tilde{X}_{kt})}{\tilde{X}_{kt}^2} + \frac{\mathrm{var}(\tilde{Y}_{kt})}{\tilde{Y}_{kt}^2} - 2\frac{Cov(\tilde{X}_{kt}, \tilde{Y}_{kit})}{\tilde{X}_{kt}\tilde{Y}_{kt}}\right]$$

$$= \left[\frac{\tilde{X}_{kt}}{\tilde{Y}_{kt}}\right]^2 \left[cv^2(\tilde{X}_{kt}) + cv^2(\tilde{Y}_{kt}) - 2\rho_{X,Y} cv(\tilde{X}_{kt}) cv(\tilde{Y}_{kt})\right]$$

This approximation is extremely **conservative**, using an assumed level of autocorrelation of ($\rho_{\mathrm{SALES},1}$) = -1 in the CHANGE estimates and a correlation of zero for the within-quarter RATIO of NIAT/SALES.

3. Methodology

3.1. Replicate Variance Estimators

The delete-a-group jackknife variance estimation method is an appealing replication method for both theoretical and operational reasons. From a theoretical perspective, the correctly applied delete-a-group jackknife variance estimator produces "nearly unbiased" variance estimators for a variety of estimators (Kott, 2001). The computational advantages are in the number of replicates: usually 15 or 16, compared to one replicate per sample unit with the stratified jackknife.

To correctly perform delete-a-group jackknife replication (hereafter referred to as DAG replication), the non-certainty portion of parent sample is divided into G random groups using the same sampling methodology used to select the parent sample (Wolter, 1985, pp. 31-32). A jackknife replicate estimate is computed for each replicate g by removing the g^{th} random group from the full sample and reweighting the remaining units to represent the full sample, either by simply multiplying the replicate by $G/(G-1)$ or by developing replicate weights for each unit and using these weights in subsequent estimation [Note: G sets of replicate weights are assigned to each sample unit j, where the g^{th} replicate weight is zero when unit j is in random group g]. We refer to this method of constructing replicate weights as **simple** delete-a-group jackknife (DAGS) replication. All certainty cases are included in each replicate with their full-sample weight (equal to 1 without non-response adjustment). Certainty totals for each enumerated industry k for characteristic X at time t are denoted X_{kt}^{C}. The sample estimation procedure is then applied to each of the replicate weights (e.g., non-response adjustments, industry reclassification) or to the replicate estimates (e.g., expansion or ratio estimates).

Our applications use fifteen random groups. We attempted to make sure that all fifteen random groups were represented in each panel in each sample industry and asset class cell by using the random group assignments in the rotated-out panel to inform the random group assignments in the incoming panel as panels rotate out of sample. This did not strictly follow the procedures outlined in Wolter (1985), but yielded far more stable replicate estimates.

The QFR variable weight estimator of quarterly totals (LEVELS) is a ratio estimator. The sample size term (n_{kit}) in the variable weight denominator is a random variable. The DAGS replicate **variable** weight g for enumerated industry k and asset class i ($\tilde{W}_{kit}^{g}$) uses replicate-factor adjusted survey weights to estimate the numerator's population estimates and estimating the replicate enumerated industry sample size $\left(n_{kit}^{g}\right)$ as $q \sum_{j\in ki,\, j\notin g} I_{hkijt}$ where $q_g = G/(G-1)$.

Kott (2001) states that the DAGS method is "reasonable" under two conditions: (1) sampling fractions are all less than 0.20 and (2) all sample stratum sizes and G are large. The following paragraphs address the condition (2), assuming condition (1). In this context, the DAGS estimator is unbiased if units from each sample stratum are represented in each replicate (i.e., if $n_{hi} \geq G$ for all hi combinations). Kott (2001) develops the **extended** delete-a-group jackknife (DAGE) to account for the situation where condition (1) is true and condition (2) is not, specifically where $n_{hig} = 0$ in several strata (a frequent occurrence with the QFR, whose panel design would require a minimum of 60 sampled elements per stratum for complete representation in each random group and panel). Let n_{hi} be the number of sampled units in sample industry h and asset class i, w_{hij} be the sampling weight associated with unit j in sample industry h and asset class i, and S_{hig} be the set of n_{hig} sample units in stratum hi and random group g. For the QFR design, the extended delete-a-group jackknife (DAGE) weights are

$$w_{hi}^{g} = \begin{cases} \mathrm{w}_{\mathrm{hij}} & \text{when } \mathrm{S}_{\mathrm{hig}} \text{ is empty} \\ \mathrm{w}_{\mathrm{hij}}(1-[\mathrm{n}_{\mathrm{hi}}-1]\mathrm{Z}) & \text{when } \mathrm{j}\in \mathrm{S}_{\mathrm{hig}} \\ \mathrm{w}_{\mathrm{hij}}(1+\mathrm{Z}) & \text{otherwise} \end{cases}$$

where $Z^2 = G/[(G-1)n_{hi}(n_{hi}-1)]$.

The DAGE replicate estimate for an expansion estimate of the form $\hat{X} = \sum_{hi}\sum_{j}^{n_{hi}} w_{hi}x_{hij}$ is given by

$$\hat{X}^{g} = \sum_{hi\in HI^{(g)}}\sum_{j=1}^{n_{hi}} w_{hij}x_{hij} + \sum_{hi\in HI^{g}} w_{hij}\left\{x_{hij*}(1-[n_{hi}-1]Z) + \sum_{j\neq j*} w_{hij}x_{hij}(1+Z)\right\}$$

where $\mathrm{HI}^{(g)}$ is the set of strata with empty $\mathrm{S}_{\mathrm{hig}}$, HI^{g} is the set of strata with at least one sample element in $\mathrm{S}_{\mathrm{hig}}$, and $j*$ are sample units that are not in random group g. We use the DAGE estimator specified in (3.2) to compute the Horvitz-Thompson component estimates of $\hat{N}_{ki,}^{(0)}\ \hat{N}_{ki,}^{(-1)}$ and $\hat{N}_{ki}^{(-2)}$ in the numerator of the variable weight. To obtain the replicate denominator sample sizes, we dropped the survey weight (w_{hij}) from the DAGE computations; we obtain the unweighted enumerated industry/asset class cell totals for characteristic x in the same way. Notice that these estimates are computed within enumerated industry and asset class and are domain estimates, not the sampling stratum estimates.

Both variations of the delete-a-group jackknife estimators assume negligible sampling fractions in all strata. In many of the non-certainty strata, the QFR optimal allocation program selects more than 20 percent of the stratified units. Thus, any variance estimate that does not incorporate finite population correction factors (fpcs) from without-replacement samples is an overestimate. Correctly applying the sampling fpc to the enumerated industry replicate estimates is difficult. Following Wolter (1985, p.43), we incorporate the sample-industry fpc's into the replicate estimates by applying the square-root of the fpc ($\sqrt{1-n_{hi}/N_{hi}}$) to the replicate totals for each **unweighted** characteristic before summing to the enumerated-industry and asset-class level, then applying the appropriate replicate variable weights. The DAGS and DAGE variable weight estimates for enumerated industry k are

$$\tilde{X}_{kt}^{g,m} = X_{kt}^{C} + \sum_{i \in k} \left(\tilde{W}_{kit}^{g,m} \right) \left(\sum_{h \in ki} \sum_{j \in hki} \sqrt{1-n_{hi}/N_{hi}}\, x_{hkijt} \right)$$

where m denotes the variance estimation method (DAGS or DAGE).

Finally, the delete-a-group (and extended delete-a-group) jackknife variance for the variable weight estimate $\hat{X}_{kt}$ is

$$v_m(\hat{X}_{kt}) = \frac{G-1}{G} \sum_{g=1}^{G} \left(\tilde{X}_{kt}^{g,m} - \tilde{X}_{0t} \right)^2$$

Here, $\tilde{X}_{0t}$ is the fpc-adjusted full-sample estimate in the non-certainty strata so that $\tilde{X}_{kt} \neq X_{kt}^{C} + \hat{X}_{0t}$.

The delete-a-group jackknife variance estimator has *G-1* degrees of freedom, where G = 15 for QFR. Bell (2000) proposes increasing the available degrees of freedom by further dividing the replicates into disjoint groups ("zones"), computing replicate variances at the "zoned" level, then summing these zoned replicate variances. If estimates from each zone are independent, then the zoned estimator should have decreased bias and be less variable than its unzoned counterpart. The independence condition is not necessarily true for the QFR estimates because they are computed by asset class within enumeration (not sampling) industry. We use asset class (size strata originally assigned within sampling industry) as zone in the evaluation discussed in Section 4.

Bell (2000) notes that zoned jackknife variance estimators must be restricted to estimates of totals, stating that functions of totals must use linearized variance estimates. Moreover, Thompson, Sigman, and Goodwin (2002) demonstrate that directly replicating ratio-type estimators with the fpc-adjustment **overestimates** the variance. Consequently, we use the Taylor linearization formula with replicate variance and covariance estimates and full-survey variable weight estimates in the relative variance and covariance terms for our "replicate" CHANGE and RATIO variance estimates.

3.2. Evaluation Statistics

To examine the statistical properties of the five different variance estimation methods, we selected 2,500 stratified random samples apiece from our four simulated populations using the QFR stratification and sampling design. We used these 2,500 random samples to construct the empirical variance of each estimate $\hat{\theta}_{pkt}$ in population p in enumerated industry k at time $t\left(V(\hat{\theta}_{pkt})\right)$. This is our "gold standard" (truth).

In 1,000 of the 2,500 samples, we assigned sample units to random groups. Then, in each sample s, we computed five variance estimates per estimate $v_{ms}(\hat{\theta}_{pkts})$. For the remainder of this paper, we denote the currently-used variance estimator as S^2, the simple delete-a-group jackknife variance estimate as DAGS, the zoned simple delete-a-group jackknife variance estimate as ZDAGS, the extended delete-a-group jackknife variance estimate as DAGE, and the zoned extended delete-a-group jackknife variance estimate as ZDAGE.

We compared these quarterly variance estimates within population and enumerated industry in terms of

Relative Bias for variance estimation method m

$$\frac{\sum_{s=1}^{1000} v_{ms}(\hat{\theta}_{pkts})/1000}{V(\hat{\theta}_{pkt})} - 1$$

Stability for variance estimation method m

$$\frac{\sqrt{\sum_{s=1}^{1000} (v_{ms}(\hat{\theta}_{pkts}) - V(\hat{\theta}_{pkt}))^2 /1000}}{V(\hat{\theta}_{pkt})}$$

Confidence Interval Coverage: Percentage of 90% confidence intervals using standard errors that contain θ_{pkt}, the true population value for the estimator. The S^2 confidence intervals use z-statistics; the replicate variance confidence intervals use t_{14} statistics.

The optimal variance estimator will have relative bias and relative stability values near zero and will have coverage rates equal to 90%. We rely primarily on relative bias as our measurement of accuracy. The S^2 estimator approximates the **conditional** variance (conditioned on the actual enumeration industries) of the total estimates of the enumerated (reclassified) industries and does not account for the variance component due to industry reclassification. This missing component can be quite large, especially in certain industries. Including it will worsen the stability (i.e., increase the variance of the variance).

The analysis of coverage rates is somewhat complicated by known properties of the variable weight estimator. The QFR variable weight estimates (LEVELS) are not unbiased. First, they are (average) combined ratio estimates. Second, they have known negative coverage bias in non-decreasing populations, caused by eligible businesses that came into existence after the construction of the sampling frame (Caldwell et al 2005). Finally, all QFR estimates are subject to non-response bias, although the degree of this bias is often negligible due to the high survey response rate. Caldwell et al (2005) showed that the combined biases of the QFR LEVEL estimator tend to be negative; there is a canceling effect of coverage bias in the CHANGE and RATIO estimates.

Cochran (1977, p.12 – 15) discusses the effect of bias in the estimator on confidence interval coverage, assuming an **unbiased** estimate of standard error, providing a working rule that "the effect of bias on the accuracy of the estimate is negligible if the bias is less than one tenth of the standard deviation of the estimate." This explanation assumes an unbiased variance estimate. If the QFR variable weight estimator were unbiased, then the coverage rates would be strongly related to the relative bias properties of the alternative variance estimators (where m = S^2, DAGS, DAGE, ZDAGS, ZDAGE), and we could derive a mathematical expression relating the degree of bias in the standard error estimate to the coverage. Intuitively, we would expect the confidence intervals constructed with **negatively** biased variance estimates to be too narrow (anti-conservative) with less-than-nominal coverage rates. Conversely, we would expect **positively** biased variance estimators to yield overly wide (conservative) confidence intervals with greater-than-nominal coverage rates. However, the QFR variable weight estimator as well as all considered variance estimates are biased, so confidence interval coverage is dependent on both the proximity of the estimate to the true value and the bias of the variance estimate $\hat{\sigma}_m$. Consequently, we cannot directly relate the bias of the variance estimates to their coverage rates: we might expect close to nominal coverage in industries with little industry reclassification (low-mover industries), but have no way to gauge variance estimator bias effects otherwise. On the other hand, because the combined biases generally cancel in CHANGE and ratio estimates, we expect close to nominal coverage in these estimates if their variance estimates have low bias.

4. Results

4.1. Effect of Zoning on Replicate Variance Estimation

When viewing time-series plots of the relative biases and stabilities of each estimator, we noticed that – regardless of population and industry – the DAGS and ZDAGS time-series plots were indistinguishable, as were the DAGE and ZDAGE time-series plots. We hypothesized that the zoned/not zoned variance estimates were not statistically different in the majority of applicable industries. We tested this hypothesis with a simple ANOVA approach, using the repeated measures model $v_{ms}(\hat{\theta}_{pkts}) = \mu_{pkt} + \xi_{pkts}$ in the manufacturing enumerated industries. After verifying the omnibus hypothesis (at least one effect is significant), we tested individual contrasts between DAGS/ZDAGS and DAGE/ZDAGE effects within quarter.

We were unable to find any evidence of a zoning effect for the simple delete-a-group jackknife variance estimators, regardless of estimator (CHANGE, LEVEL, or RATIO), enumerated industry, or population. We did find minor evidence of a zoning effect for the extended delete-a-group jackknife variance estimators. Neither variation of extended delete-a-group jackknife variance estimator showed any advantage over the other variance estimation methods in terms of relative bias and relative stability (see Section 4.2). This is not entirely dissimilar from the simulation results in Bell (2000), who found that the zoned jackknife variance estimator had less variability compared to the standard delete-a-group jackknife variance estimator only when the post-strata and strata were very similar, although the bias of the variance estimator increased regardless.

The main advantage of using a zoned estimator would be to increase the degrees of freedom in variance estimation and in confidence interval construction. An application of the zoned estimator to the manufacturing industries in QFR does increase the available degrees of freedom, since these the companies in these sampling industries are stratified into five non-certainty strata. The same benefits do not apply to industries from other sectors in the QFR design: these sectors have one non-certainty stratum per sampling industry. We found no variance estimation benefits from the zoned estimates and could not consistently incorporate the additional degrees of freedom into our confidence intervals. Consequently, we dropped both zoned variance estimators from our analysis. As an aside, in all industries and scenarios, the contrasts between corresponding simple and extended delete-a-group jackknife were always significant.

4.2. Comparison of the Current Variance Estimation Method to the Replication Variance Estimation Methods

This section compares the statistical properties of the replicate variances (DAGS and DAGE) to the corresponding currently-used variance estimator (S^2), considering each type of **estimator** (CHANGE, LEVEL, RATIO) separately. In the discussion below, for simplicity we use the same notation for all estimators, recalling that the CHANGE and RATIO estimators use Taylor linearization with relative variances and covariances of input LEVEL estimates computed via the referenced method. For brevity, we do not include computed summary statistics; they are available from the authors upon request.

CHANGE The major difference between the replicate and non-replicate variance estimates is in the covariance term of the Taylor Linearization. The currently-used linearization assumes an autocorrelation term of ρ = -1, thus maximizing the covariance term of the variance estimate. This assumption is unrealistic, since three-quarters of the QFR sample do not change from quarter to quarter because of the rotating panel design. Our replicate estimates computed large positive correlations (usually between 0.60 to 0.80), reducing the estimated variances. We found that

- In all scenarios, the simple-delete-a-group jackknife variance estimates for CHANGE are the least biased, the most stable, and have the best coverage (closest to nominal), and the S^2 estimates performed the worst in all three measures. This improvement in precision from non-replicate to replicate estimates is primarily due to the improved estimates of covariance used in the linearization formula [Note: CHANGE variances constructed with replicate variance estimates and same assumed covariance estimates as current method have approximately the same bias properties as the current method].
- The bias of all three CHANGE variance estimators is **positive**, regardless of scenario or industry. The relative biases for all S^2 variance estimates are well over 200%, regardless of industry or population scenario. The two

sets of relative biases for the replicate estimates are quite close in the low-mover industries and are generally around 20%. The DAGE relative biases are considerably larger than their DAGS counterparts in the medium- and high-mover industries, although the degree of bias in the DAGE estimators is still considerably smaller than in the corresponding S^2 variance estimates.

- The instability in the S^2 variance estimates may be attributable to the treatment of the variance estimates as independent in the Taylor Linearization. The replicate estimates have induced covariance by design because of the balanced method used for assigning sample units to random groups. The DAGS and DAGE stabilities are very close in the low- and medium-mover industries; the DAGE stability is much larger (more variable) than the DAGS stability in the high-mover industry.
- Coverage rates are closer to the nominal 90% for the replicate methods; coverage rates are close to 100% for the S^2 variance estimators (due to the large positive bias).

LEVELS Recall that LEVEL estimates are combined ratio estimates and are consequently **not** unbiased. We found that

- The S^2 variance estimator is always **negatively** biased, whereas the replicate estimators are all **positively** biased. This result is expected: the S^2 estimator does **not** account for the industry reclassification and the replicate variance estimators do. In the low-mover industries, the DAGS and DAGE relative biases are approximately the same for both NIAT and SALES; in the medium- and high-mover industries, the DAGE biases are considerably larger than their DAGS counterparts (this effect is more pronounced for SALES than NIAT, although this may be an artifact of the different modeling assumptions for the two variables). Since the DAGE method is designed to reduce the bias of the replicate variance estimator, this result was somewhat unexpected.
- Because they do not account for variability due to industry reclassification, the S^2 variance estimators are the most stable. The same pattern of stability seen in the replicate biases holds here, namely equivalent DAGE and DAGS stabilities in the low-mover industries and worsening DAGE stabilities in the medium- and high-mover industries.
- Coverage is affected by both magnitude of relative bias in the variance estimates and the bias in the variable weight estimates. The replicate variance estimators are more biased (in magnitude) than the S^2 variance estimators. In addition, the replicate confidence intervals are constructed using t-statistics instead of z-statistics, making them slightly wider than corresponding intervals constructed from the S^2 variance estimators.

Regardless of variance estimator, the coverage rates for SALES are much smaller than expected; coverage rates are extremely poor in most scenarios and industries, although the DAGE and DAGS confidence intervals are always closer to nominal than the S^2 confidence intervals. Coverage rates for NIAT constructed with replicate variance estimates and t-statistics are, however, often close to nominal in all populations and industry-types. As discussed in Section 2, the variable weight estimates of SALES are negatively biased in non-decreasing populations, whereas the variable weight estimates of NIAT are essentially unbiased in all populations. Recall that the degree of bias in the SALES estimates is greatly exaggerated by design. For SALES, the high negative bias in the estimator is consistently offsetting the positive bias in the variance estimates, leading to poor coverage. In contrast, confidence interval coverage for NIAT is more dependent on the bias in the variance estimator (see Section 3.2).

RATIO As with CHANGE, the variance estimates for RATIO use Taylor Linearization. In this application, the replicate correlations ($\rho_{\text{NIAT,SALES}}$) are all approximately equal to zero, the assumed covariance level in the current estimator. Thus, the differences in the three sets of approximations are due primarily to the different variance estimation methods. The patterns in the measurement statistics are very similar to those for LEVELS. NIAT is a much more variable item than SALES, and consequently the estimated relative variance of NIAT used in the Taylor Linearization is the dominant term. We found that

- As with the LEVEL estimates, the relative biases of the S^2 variance estimators are all negative, and the replicate estimates are always positively biased. In the medium- and high- mover industries, the DAGE relative biases are much larger than the DAGS statistics
- The increase in stability (in variability) from the S^2 variance estimators to the DAGS estimators is about 0.01 in most populations and industries. The increase in stability from DAGS to DAGE is much larger, although this effect appears to be more pronounced in the medium- and high-mover industries and may be dependent on population scenario.
- Again, the coverage rates are closer to nominal with the DAGS and DAGE estimates. The proximity to nominal coverage rates depends on industry type (low-mover industries have nominal coverage; others have less than nominal coverage).

In all cases, the DAGS estimators yield closer to nominal coverage rates than the corresponding S^2 estimators. Moreover, for CHANGE estimates, the DAGS variance estimates show substantial improvements over the S^2 variance estimates in terms of **magnitude** of the relative bias and stability statistics. For LEVEL and RATIO estimates, the **magnitude** of the DAGS and S^2 variance estimates' biases are generally comparable, as are the computed stability statistics. For these two statistics, however, the DAGS biases are always positive (overestimates) and the S^2 biases are always negative. Regardless of estimator (CHANGE, ratio, level), the DAGE variance estimates have larger relative bias and stability than the corresponding DAGS variance estimates. The improved coverage rates using DAGE variance estimates over the two other variance estimators is due to the larger bias in the DAGE variance estimator. The often marginal improvement in coverage rates constructed with DAGE variance estimators did not offset the losses in statistical precision, so we eliminated the DAGE method from consideration for QFR.

Subject-matter experts had some initial concerns about the "practical" impact of replacing the negatively biased S^2 variance estimates with the positively biased DAGS estimates. To assess this, we computed coefficient of variation (c.v.) values for all estimators using both sets of variance estimates. For RATIO and LEVEL estimates, the two sets of c.v.'s were virtually identical (equal to the second or third decimal place). For CHANGE estimates, the c.v.'s computed from DAGS variance estimates were much smaller than corresponding S^2 variance estimate c.v.'s.

5. Discussion

Prior to this study, it was known that the currently used S^2 variance formula for LEVEL estimates used by QFR underestimated the variance. Conversely, it was known that the Taylor Linearization estimates for CHANGE – at least – were overestimates, due to the assumed correlation.

The primary goal of our research was to provide compelling evidence to replace the current method of estimating LEVEL variances with replication, specifically with a form of delete-a-group replication. We believe that given the very complicated design and estimation procedure used, replication is simply more justifiable from a statistical perspective. These methods directly account the industry reclassification and ratio estimation in the variable weights, include some non-response variance, and use the true sampling fpcs. As expected, the replicate methods did not exhibit any gains in statistical properties in the low-mover industries, since the industry reclassification component is negligible, and the replicate estimates are essentially equivalent to the S^2 estimates. Any gains would be expected in medium and high-mover industries, where the industry reclassification effects can be quite large.

Designing balanced delete-a-group jackknife replicates with a panel survey was quite challenging. Despite our best efforts, not all strata from all panels and frames are represented in each random group. We hoped to reduce the resultant bias in the replicate variance estimates by using the extended delete-a-group jackknife variance estimator. This estimator has optimal properties for survey samples selected with negligible fpcs in all strata where "number of sample units per first-phase stratum (is) large in all strata." (Kott, 2001). In this context, "large" means that all strata are represented in each random group. Neither of these conditions is true for the QFR design, which may explain why using the extended delete-a-group jackknife instead of the simple delete-a-group jackknife did not reduce the bias. Furthermore, Kott (2001) does not discuss the properties of the DAGE variance estimator for panel surveys (or how to best assign units to random groups in a panel survey setting), nor does he discuss the effects of reclassification of sample units. We suspect that the combination of all of these factors explains the poor performance of the DAGE method for the QFR variable weight estimates of SALES and NIAT.

The results in Section 4.2 demonstrate marked improvements in all measurements for the DAGS CHANGE variances over the S^2 estimates. At first (and possibly second) glance, the decreased performance in the relative bias and stability measures for SALES LEVEL estimates with the DAGS over S^2 estimates is unsettling. More unsettling to our subject-matter experts was the poor confidence interval coverage for SALES LEVEL estimates, regardless of variance estimator. Since, however, the corresponding measures of relative bias and stability for NIAT LEVEL estimates are fairly close for the DAGS and S^2 estimates, and the DAGS NIAT confidence interval coverage is much closer to nominal, we suspect these results are an effect of the exaggerated negative bias in the sales estimates and are not necessarily present in QFR data, especially since the relative rankings of the statistics are the same within enumerated industry and population.

Finally, we were unable to find any evidence of a "population size change" effect for relative bias or stability in either the replicate or non-replicate variance estimators. This is a quite different result from the study in **estimator** properties presented in Caldwell et al (2005), which showed a negative bias for LEVEL estimates in non-decreasing populations and trivial bias in strictly decreasing populations. Confidence interval coverage rates are, however, closer to nominal in the decreasing population scenario (Population 2). This population is much like the "ideal" situation described in Section 3, with an unbiased estimator and a biased variance estimate and the coverage rate comparisons are not confounded (as they are in the other populations).

Our simulation study results demonstrated strong statistical advantages of using variance estimates for CHANGE computed with directly replicated DAGS variances and autocovariances over the current method. Recall that the quarter-to-quarter change in sales (CHANGE) is a key economic statistic. The improved results in both precision and confidence interval coverage for this statistic alone almost justify our recommendation to replace the current method with simple delete-a-group jackknife variance estimation. The comparable results for the other estimators reinforce this conclusion, as does the computational simplicity. Areas for future research include comparisons of the simple and extended delete-a-group jackknife variance estimators on a panel survey with **negligible** fpc's and industry reclassification, empirical comparisons of the discussed variance estimation methods on QFR historic data, and an assessment of the effect of replicating non-response adjustment on QFR variance estimates.

Acknowledgements

The authors thank Samson Adeshiyan, Carol Caldwell, Patrick Cantwell, Michael Ikeda, Donald Luery, Brett Moore, Broderick Oliver, Rita Petroni, Mark Sands, Richard Sigman, and John Slanta for their comments on earlier drafts of this paper.

References

Bell, P. (2000). Weighting and Standard Error Estimation for ABS Household Surveys. Unpublished Australian Bureau of Statistics Memorandum.

Caldwell et al (2005). Investigation Of Alternative Estimators For The Quarterly Financial Report. Unpublished internal U.S. Census Bureau memorandum, Company Statistics Division.

Chapman, D.W. and Biemer, P. (1985). A Comparison of Two Estimators for the Quarterly Financial Survey. Unpublished internal U.S. Census Bureau memorandum, Statistical Research Division Report Series Census/SRD/RR-85-20, available from the Statistical Research Division.

Chapman, D.E. (1993). QFR Research: Investigation of the Properties of the Current QFR Estimator. Unpublished U.S. Census Bureau memorandum.

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

Kott, P. (1992). Post-Stratifying With Sample Data Only: Can Using the Obvious Model Help? *Proceedings of the Section of Survey Research Methods*, American Statistical Association, pp. 632-636.

Kott, P. (1998). Using the Delete-A-Group Jackknife Variance Estimator in NASS Surveys. RDResearch 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.

Luery, D. (2005). Data Modeling for a Simulation Study of the Quarterly Financial Report Estimator, *Proceedings of the Section of Survey Research Methods*, American Statistical Association, forthcoming.

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.

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

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

[i] This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.