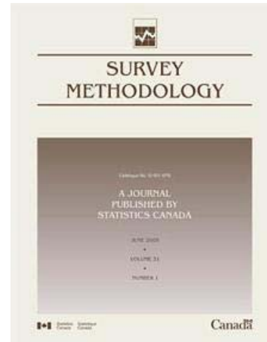
Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Variance estimation in the presence of nonrespondents and certainty strata



by Jun Shao and Katherine J. Thompson

December 2009

Statistics Canada Statistique Canada

Canada

Survey Methodology, december 2009
Vol. 35, No. 2, pp. 215-225
Statistics Canada, Catalogue No. 12-001-X

Variance estimation in the presence of nonrespondents and certainty strata

Jun Shao and Katherine J. Thompson [1]

Abstract

Business surveys often use a one-stage stratified simple random sampling without replacement design with some certainty strata. Although weight adjustment is typically applied for unit nonresponse, the variability due to nonresponse may be omitted in practice when estimating variances. This is problematic especially when there are certainty strata. We derive some variance estimators that are consistent when the number of sampled units in each weighting cell is large, using the jackknife, linearization, and modified jackknife methods. The derived variance estimators are first applied to empirical data from the Annual Capital Expenditures Survey conducted by the U.S. Census Bureau and are then examined in a simulation study.

Key Words: Covariate dependent nonresponse; Jackknife; Linearization; Ratio adjustment; Uniform nonresponse.

1. Introduction

Many business surveys use a one-stage stratified simple random sample without replacement design. Because of the skewness of the sampled populations, these designs generally include both certainty and non-certainty strata. With such designs, the sampling rates in the non-certainty strata are generally negligible (*e.g.*, less than 20 percent in all strata). However, if the ultimate sampling unit is large business entity such as a company, the size of the universe is much smaller and often sampling fractions should not be ignored in computation of variance estimates.

Most surveys have nonresponse. We consider surveys using weighting adjustment for nonresponse. For certainty strata, there is no sampling error and, hence, standard variance formulas do not include any component for certainty strata. When nonresponse is present, however, there is an estimation error even in a certainty stratum, which is often an appreciable component of the total estimation error.

The purpose of this paper is to develop some methods for variance estimation that take into account the weighting adjustment for nonresponse and the existence of certainty strata. After introducing notation and assumptions in Section 2, we show that the jackknife and linearization variance estimators ignoring nonresponse in certainty strata, which are often currently used in many surveys, underestimate the true variance of the weight adjusted estimated population total. By directly deriving an approximate variance formula, we obtain two consistent variance estimators. These variance estimators are also consistent if there are non-certainty strata with large sampling fractions. A modified jackknife variance estimator taking into account the variability due to nonresponse in certainty strata is also derived.

In Section 3, we compare variance estimators using five years' of data from the Annual Capital Expenditures Survey (ACES) conducted by the U.S. Census Bureau. Simulation results are presented in Section 4 using a population generated from 2003 ACES data. Our simulation results show that the variance estimators ignoring certainty strata have large negative biases; the derived consistent variance estimators perform well when stratum sample sizes are all large and perform inconsistently otherwise; and the jackknife variance estimator ignoring all sampling fractions overestimates. Some concluding remarks are given in Section 5.

2. Main results

Consider a stratified sample without replacement from a finite population containing H strata. Let n_h and N_h be the sample and population size of stratum h, respectively, y_{hj} be a variable of interest that may have nonresponse, and x_{hj} be a covariate that takes positive values and does not have nonresponse, where j is the index of population unit and h is the index for stratum. Using the sample-response path considered by Fay (1991) and Shao and Steel (1999), we view the finite population as a census with y, x values and nonrespondents, *i.e.*, each unit j in stratum h of the finite population is associated with an indicator I_{hj} (= 1 if y_{hj} is a respondent and = 0 if y_{hj} is a nonrespondent). Our sample is taken from this finite population, and if unit j in stratum h is in the sample, y_{hj} is a respondent if $I_{hj} = 1$ and a nonrespondent if $I_{hj} = 0$.

Let E_s and V_s be the expectation and variance, respectively, with respect to sampling and E_m, V_m, and P_m be the expectation, variance, and probability, respectively,

1. Jun Shao, University of Wisconsin-Madison and U.S. Census Bureau; Katherine J. Thompson, U.S. Census Bureau. E-mail: shao@stat.wisc.edu.

with respect to the model m specified in one of the following assumptions.

Assumption M. Values of (y_{hj}, x_{hj}, I_{hj}) in the finite population are independently generated from a superpopulation model m. The finite population is divided into P sub-populations such that, within sub-population p, the response probability $P_m(I_{hj} = 1 \mid y_{hj}, x_{hj}) = P_m(I_{hj} = 1 \mid x_{hj}) > 0$, $E_m(y_{hj} \mid x_{hj}) = \beta_p x_{hj}$, and $V_m(y_{hj} \mid x_{hj}) = \sigma_p^2 x_{hj}$, where β_p and σ_p are unknown parameters depending on p.

Assumption P. The finite population is divided into P sub-populations such that, under a superpopulation model, $P_m(I_{hj} = 1 \mid y_{hj}, x_{hj}) = \pi_p > 0$ is constant within sub-population p.

The sub-population in Assumption M or Assumption P is called nonresponse adjustment weighting cell (or weighting cell for short), since we handle nonrespondents by weight adjustment within each weighting cell. (If imputation is applied within each sub-population, then sub-populations are called imputation cells.) In applications, weighting cells may be strata, or unions of strata (strata are collapsed when they have insufficient respondents), or may cut across strata. Assumption M involves a prediction model between y_{hj} and x_{hj} and a covariate-dependent response mechanism within each weighting cell. The response mechanism under Assumption P is the within-weighting-cell uniform response mechanism and is often referred to as the quasi-random response model. Assumption P is stronger than Assumption M in terms of the response mechanism. However, Assumption M requires an explicit model between y_{hj} and x_{hj} within each weighting cell. In this paper we assume either Assumption M or Assumption P. Estimators that can be justified under Assumption P are referred to as the "quasi-randomization" estimators (Oh and Scheuren 1983).

When we study asymptotic consistency of estimators, we consider the limiting process of $k_p \to \infty$ for all p with fixed H and P, where k_p is the sample size in weighting cell p. If weighting cells are the same as strata or unions of strata, then $k_p \to \infty$ is the same as $n_h \to \infty$ for all h.

After the ratio-adjustment for nonresponse, we consider the following estimator of the total of y-values in the finite population:

$$\hat{Y} = \sum_p \sum_h \sum_{j \in s_h} \left(\frac{\hat{X}_p}{\hat{X}_{pr}} w_{hj} \right) \delta_{phj} I_{hj} y_{hj} = \sum_p \frac{\hat{X}_p}{\hat{X}_{pr}} \hat{Y}_{pr}, \quad (1)$$

where p is the index for weighting cell, s_h is the sample in stratum h, δ_{phj} is the indicator for the weighting cell p, and w_{hj} is the survey weight constructed for the stratified sampling,

$$\hat{X}_p = \sum_h \sum_{j \in s_h} w_{hj} \delta_{phj} x_{hj}, \quad \hat{X}_{pr} = \sum_h \sum_{j \in s_h} w_{hj} \delta_{phj} I_{hj} x_{hj},$$

and

$$\hat{Y}_{pr} = \sum_h \sum_{j \in s_h} w_{hj} \delta_{phj} I_{hj} y_{hj}.$$

In the special case where weighting cells are the same as strata,

$$\hat{Y} = \sum_h \frac{\hat{X}_h}{\hat{X}_{hr}} \hat{Y}_{hr}, \quad (2)$$

where

$$\hat{X}_h = \sum_{j \in s_h} w_{hj} x_{hj}, \quad \hat{X}_{hr} = \sum_{j \in s_h} w_{hj} x_{hj} I_{hj},$$

and

$$\hat{Y}_{hr} = \sum_{j \in s_h} w_{hj} y_{hj} I_{hj}.$$

When the covariate $x_{hj} \equiv 1$, $\hat{Y}$ is referred to as the count estimator. The count estimator controls respondent estimates to frame population totals. When the weighting cells are the same as strata, the count estimator uses the unweighted cell response rates, as recommended in Vartivarian and Little (2002).

Under Assumption M or P,

$$E_m E_s (\hat{Y} - Y) = E_s E_m (\hat{Y} - Y) = 0,$$

where Y is the finite population total of y values, and the total variance

$$V_{m,s}(\hat{Y} - Y) = E_m[V_s(\hat{Y})] + V_m[E_s(\hat{Y}) - Y].$$

Let $V_1 = E_m[V_s(\hat{Y})]$ and $V_2 = V_m[E_s(\hat{Y}) - Y]$. To estimate V_1, it suffices to estimate the sampling variance $V_s(\hat{Y})$. Since $\hat{Y}$ defined by (1) is a sum of ratios and each of $\hat{X}_p$, $\hat{X}_{pr}$, and $\hat{Y}_{pr}$ is a weighted total of variables and indicators, we can apply the stratified jackknife variance estimator

$$v_{J1} = \sum_h \left(1 - \frac{n_h}{N_h}\right) \frac{n_h - 1}{n_h} \sum_{j \in s_h} \left(\hat{Y}_{(hj)} - \frac{1}{n_h} \sum_{k \in s_h} \hat{Y}_{(hk)} \right)^2 \quad (3)$$

(see Wolter 1985 or Shao and Tu 1995), where $\hat{Y}_{(hj)}$ is the jackknife analog of $\hat{Y}$ when unit j in stratum h is deleted. Note that sampling fractions are incorporated in this formula. When $k_p \to \infty$ for all weighting cells, the standard result for the complete data case (see, *e.g.*, Krewski and Rao 1981) implies that the jackknife estimator v_{J1} is consistent for the sampling variance $V_s(\hat{Y})$, under Assumption M or P. Since V_1 is the expectation of $V_s(\hat{Y})$, v_{J1} is also consistent for V_1 under some minor conditions.

Since the function in (1) is the sum of ratios and data in different weighting cells are independent, a linearization estimator of $V_s(\hat{Y})$ can be derived using Taylor's expansion.

When weighting cells are the same as strata, for example, $\hat{Y}$ is given by (2) and is a separate ratio estimator whose linearization variance estimator can be obtained using standard techniques. An alternative way to derive a linearization variance estimator is to linearize the jackknife estimator v_{J1} (Thompson and Yung 2006). The resulting estimator is

$$v_{L1} = \sum_h \frac{n_h}{n_h - 1} \sum_{j \in s_h} \left\{ \sum_p \left[\frac{\hat{X}_p}{\hat{X}_{pr}} (\bar{e}_{ph} - w_{hj} e_{phj} I_{hj} \delta_{phj}) + \frac{\hat{Y}_{pr}}{\hat{X}_{pr}} (\bar{x}_{ph} - w_{hj} x_{hj} \delta_{phj}) \right] \right\}^2, \quad (4)$$

where $e_{phj} = y_{hj} - (\hat{Y}_{pr}/\hat{X}_{pr}) x_{hj}$, $\bar{e}_{ph} = n_h^{-1} \sum_{j \in s_h} w_{hj} e_{phj} I_{hj} \delta_{phj}$, and $\bar{x}_{ph} = n_h^{-1} \sum_{j \in s_h} w_{hj} x_{hj} \delta_{phj}$. The estimator in (4) is exactly the same as the standard linearization variance estimator for the separate ratio estimator in (2) when weighting cells are the same as strata. Like v_{J1}, v_{L1} is consistent for V_1 when $k_p \to \infty$ under Assumption M or P, which follows from the standard result for the complete data case (Krewski and Rao 1981).

Since ratio is a smooth function, under Assumption M or P,

$$E_s(\hat{Y}) = \sum_p E_s \left(\frac{\hat{X}_p \hat{Y}_{pr}}{\hat{X}_{pr}} \right) \approx \sum_p \frac{E_s(\hat{X}_p) E_s(\hat{Y}_{pr})}{E_s(\hat{X}_{pr})} = \sum_p \frac{X_p Y_{pr}}{X_{pr}},$$

where

$$X_p = \sum_h \sum_{j \in \mathcal{P}_h} \delta_{phj} x_{hj},$$

$$X_{pr} = \sum_h \sum_{j \in \mathcal{P}_h} \delta_{phj} I_{hj} x_{hj},$$

$$Y_{pr} = \sum_h \sum_{j \in \mathcal{P}_h} \delta_{phj} I_{hj} y_{hj},$$

and $\mathcal{P}_h$ is the finite population in stratum h. Let Y_p be the same as X_p with x_{hj} replaced by y_{hj}. Then

$$V_2 = V_m[E_s(\hat{Y}) - Y] \approx \sum_p V_m \left(\frac{X_p Y_{pr}}{X_{pr}} - Y_p \right).$$

Note that V_2 is small if the nonresponse rate is low ($V_2 = 0$ if there is no nonresponse) or if the model under Assumption M is highly predictive. If the overall sampling fraction, $\sum_h n_h / \sum_h N_h$, converges to 0, then V_2/V_1 converges to 0 and, hence v_{L1} and v_{J1} are consistent estimators of the total variance $V_{m,s}(\hat{Y}) = V_1 + V_2 \approx V_1$. Note that V_1 does not contain the variation from certainty strata due to nonresponse. Because the y-values from certainty strata are influential in the total Y in many surveys, and because in applications it is difficult to tell how small $\sum_h n_h / \sum_h N_h$ has to be for the convergence $V_2/V_1 \to 0$ to take place, it is necessary to estimate V_2.

Under Assumption M, let $\tilde{E}_m$, $\tilde{V}_m$, and $\tilde{C}_m$ be the conditional expectation, variance, and covariance, respectively, given all x-values and response indicators. Since

$$\tilde{E}_m \left(\frac{X_p Y_{pr}}{X_{pr}} - Y_p \right) = 0,$$

we obtain

$$\begin{aligned} V_m \left(\frac{X_p Y_{pr}}{X_{pr}} - Y_p \right) &= E_m \left[\tilde{V}_m \left(\frac{X_p Y_{pr}}{X_{pr}} - Y_p \right) \right] + V_m \left[\tilde{E}_m \left(\frac{X_p Y_{pr}}{X_{pr}} - Y_p \right) \right] \\ &= E_m \left[\tilde{V}_m \left(\frac{X_p Y_{pr}}{X_{pr}} - Y_p \right) \right] \\ &= E_m \left[\frac{X_p^2}{X_{pr}^2} \tilde{V}_m(Y_{pr}) - 2 \frac{X_p}{X_{pr}} \tilde{C}_m(Y_{pr}, Y_p) + \tilde{V}_m(Y_p) \right] \\ &= E_m \left[\frac{X_p^2}{X_{pr}^2} \sigma_p^2 X_{pr} - 2 \frac{X_p}{X_{pr}} \sigma_p^2 X_{pr} + \sigma_p^2 X_p \right] \\ &= \sigma_p^2 E_m \left(\frac{X_p^2}{X_{pr}} - X_p \right). \end{aligned}$$

Under Assumption P, let V_m^I be the variance with respect to I_{hj}'s. Since

$$E_m^I \left(\frac{X_p Y_{pr}}{X_{pr}} - Y_p \right) \approx 0,$$

we obtain

$$\begin{aligned} V_m \left(\frac{X_p Y_{pr}}{X_{pr}} - Y_p \right) &\approx E_m \left[V_m^I \left(\frac{X_p Y_{pr}}{X_{pr}} - Y_p \right) \right] \\ &\approx E_m \left[\frac{1 - \pi_p}{\pi_p} \sum_h \sum_{j \in P_h} \delta_{hj} \left(y_{hj} - \frac{Y_p}{X_p} x_{hj} \right)^2 \right] \\ &\approx E_m \left[\left(\frac{X_p^2}{X_{pr}} - X_p \right) S_p^2 \right], \end{aligned}$$

where

$$S_p^2 = \frac{1}{X_{pr}} \sum_h \sum_{j \in \mathcal{P}_h} \delta_{hj} \left(y_{hj} - \frac{Y_p}{X_p} x_{hj} \right)^2.$$

Since X_p and X_{pr} can be estimated by $\hat{X}_p$ and $\hat{X}_{pr}$, respectively, to estimate V_2 we only need to find an estimator of σ_p^2 or S_p^2. Under Assumption M, a regression estimator of β_p is $\hat{Y}_{pr}/\hat{X}_{pr}$ and a consistent estimator of σ_p^2 based on regression residuals is

$$\hat{\sigma}_p^2 = \frac{1}{\hat{X}_{pr}} \sum_h \sum_{j \in s_h} \delta_{phj} I_{hj} w_{hj} \left(y_{hj} - \frac{\hat{Y}_{pr}}{\hat{X}_{pr}} x_{hj} \right)^2.$$

From the theory of sampling, $\hat{\sigma}_p^2$ is also a consistent estimator of S_p^2 under Assumption P. Hence, under Assumption M or P, a consistent estimator of V_2 is

$$v_{L2} = \sum_p \hat{\sigma}_p^2 \left(\frac{\hat{X}_p^2}{\hat{X}_{pr}} - \hat{X}_p \right). \quad (5)$$

The subscript L indicates that this estimator is based on linearization.

In some applications $\sum_h n_h / \sum_h N_h$ is negligible and non-response in noncertainty strata has negligible contribution to the variance component V_2, *i.e.*,

$$V_2 \approx V_m \left[\sum_p \left(\frac{X_{cp} Y_{cpr}}{X_{cpr}} - Y_{cp} \right) \right], \quad (6)$$

where the subscript c stands for certainty strata,

$$X_{cp} = \sum_{h \in \mathcal{C}} \sum_{j \in \mathcal{P}_h} \delta_{phj} x_{hj}, \quad X_{cpr} = \sum_{h \in \mathcal{C}} \sum_{j \in \mathcal{P}_h} \delta_{phj} I_{hj} x_{hj},$$

$$Y_{cp} = \sum_{h \in \mathcal{C}} \sum_{j \in \mathcal{P}_h} \delta_{phj} y_{hj}, \quad Y_{cpr} = \sum_{h \in \mathcal{C}} \sum_{j \in \mathcal{P}_h} \delta_{phj} I_{hj} y_{hj},$$

and $\mathcal{C}$ is the collection of indices of certainty strata. A consistent jackknife estimator of V_2 can be obtained as follows. Note that X_{cp}, X_{cpr}, and Y_{cpr} are estimators, since $\mathcal{P}_h = s_h$ for $h \in \mathcal{C}$, but Y_{cp} is not an estimator because of nonresponse. Thus, we cannot apply the jackknife to the function $X_{cp} Y_{cpr} / X_{cpr} - Y_{cp}$. From the previous derivation we note that, under Assumption M,

$$V_2 \approx E_m \tilde{V}_m \left[\sum_p \left(\frac{X_{cp} Y_{cpr}}{X_{cpr}} - Y_{cp} \right) \right]$$

$$= E_m \left[\sum_p \left(1 - \frac{X_{cpr}}{X_{cp}} \right) \tilde{V}_m \left(\frac{X_{cp} Y_{cpr}}{X_{cpr}} \right) \right].$$

Similarly, under Assumption P, the result holds with $\tilde{V}_m$ replaced by V_m^I. Hence, we can apply the jackknife to the estimator $X_{cp} Y_{cpr} / X_{cpr}$. Let

$$\breve{Y} = \sum_p \sqrt{1 - \frac{X_{cpr}}{X_{cp}}} \left(\frac{X_{cp} Y_{cpr}}{X_{cpr}} \right)$$

and $\breve{Y}_{(hj)}$ be the jackknife analog of $\breve{Y}$ after unit j in $h \in \mathcal{C}$ is deleted, when we treat $X_{cp} Y_{cpr} / X_{cpr}$ as estimators. Then a jackknife estimator of V_2 is

$$v_{J2} = \sum_{h \in C} \frac{N_h - 1}{N_h} \sum_{j \in \mathcal{P}_h} \left(\breve{Y}_{(hj)} - \frac{1}{N_h} \sum_{k \in P_h} \breve{Y}_{(hk)} \right)^2$$

($n_h = N_h$ and $s_h = \mathcal{P}_h$ when $h \in \mathcal{C}$). The factor $\sqrt{1 - X_{cpr}/X_{cp}}$ in the formula for $\breve{Y}$ makes the appropriate adjustment for nonresponse. Under Assumption P, $X_{cpr}/X_{cp} \approx \pi_p$ is the response rate, which can be view as a "sampling" fraction for certainty strata.

The resulting jackknife estimator of the total variance $V_1 + V_2$ is then $v_{J1} + v_{J2}$. Since $n_h = N_h$ (*i.e.*, $1 - n_h / N_h = 0$) if stratum h is a certainty stratum, it is easy to see that $v_{J1} + v_{J2}$ is equal to

$$v_J = \sum_h \frac{n_h - 1}{n_h} \sum_{j \in s_h} \left(\tilde{Y}_{(hj)} - \frac{1}{n_h} \sum_{k \in s_h} \tilde{Y}_{(hk)} \right)^2, \quad (7)$$

where

$$\tilde{Y}_{(hj)} = \begin{cases} \breve{Y}_{(hj)} & \text{if stratum } h \text{ is a certainty stratum} \\ \hat{Y}_{(hj)} \sqrt{1 - \dfrac{n_h}{N_h}} & \text{if stratum } h \text{ is not a certainty stratum.} \end{cases}$$

Compared with the jackknife variance estimator v_{J1} in (3), v_J in (7) addresses the variability due to nonresponse in certainty strata, whereas v_{J1} does not. Under (6) and Assumption M or P, v_J is consistent.

Finally, the jackknife estimator that ignores all sampling fractions is:

$$\tilde{v}_J = \sum_h \frac{n_h - 1}{n_h} \sum_{j \in s_h} \left(\hat{Y}_{(hj)} - \frac{1}{n_h} \sum_{k \in s_h} \hat{Y}_{(hk)} \right)^2. \quad (8)$$

This estimator seems to be conservative, although it is not theoretically justified.

In summary, we have the following estimators of the total variance $V_{m,s}(\hat{Y})$:

1. The jackknife estimator v_{J1} defined in (3), which underestimates when V_2/V_1 is not negligible.
2. The linearization estimator v_{L1} defined in (4), which is asymptotically equivalent to v_{J1}.
3. $v_L = v_{L1} + v_{L2}$ with v_{L2} is defined in (5), which is consistent.
4. $v_{JL} = v_{J1} + v_{L2}$, which is asymptotically equivalent to v_L.
5. The jackknife variance estimator v_J defined in (7), which is consistent when (6) holds.
6. The jackknife estimator $\tilde{v}_J$.

Under stratified simple random sampling and Assumption P, v_L is approximately the same as the variance estimator obtained by treating the set of respondents as an additional phase of the stratified simple random sample (*i.e.*, a two-phase sample design) and applying standard variance formula (when $x_{hj} \equiv 1$) or the variance formula for calibration estimators (Kott 1994, Särndal, Swensson and Wretman 1992, and Hidiroglou and Särndal 1998). This variance estimator, however, is not consistent when Assumption P does not hold.

3. Empirical comparisons

In this section, we apply the variance estimators described in Section 2 to five years of empirical data from the employer component of the ACES introduced in Section 1. Section 3.1 provides background on the ACES analysis variables, sample design, and estimation procedures. Section 3.2 presents the empirical comparisons.

3.1 Background of ACES

The ACES collects data about the nature and level of capital expenditures in non-farm businesses operating within the United States. Respondents report capital expenditures, broken down by type (expenditures on Structures and expenditures on Equipment) for the calendar year in all subsidiaries and divisions for all operations within the United States.

The ACES universe contains two sub-populations: employer companies (ACE-1) and non-employer (ACE-2) companies. (A nonemployer company 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 un-incorporated businesses, which may or may not be the owner's principal source of income). Different forms are mailed to sample units depending on whether they are ACE-1 companies or ACE-2 companies. New ACE-1 and ACE-2 samples are selected each year, both with stratified simple random sample without replacement designs. The ACE-1 sample comprises approximately seventy-five percent of the ACES sample (roughly 46,000 companies selected per year for ACE-1, and 15,000 selected per year for ACE-2). In the ACE-1 design, units are stratified into size-class strata within each industry on the sampling frame. There are five separate ACE-1 strata in each industry, consisting of one certainty stratum (referred to as stratum 10) and four non-certainty strata defined by company size within industry (denoted by 2A through 2D, ranked from largest to smallest within industry), with approximately 500 non-certainty strata in each year's design. Sampling fractions in the large-size class-within-industry strata (2A) can be fairly high: in most years, approximately 55% of the sample in 2A strata are sampled at rates between 0.5 and 1. Sampling fractions in the other three size class within-industry strata are usually less than 0.20. Design weights range from 1 to 1,000, depending on industry and size-class strata. The ACE-2 component is much less highly stratified, with between a total of six to eight size-class strata used each year, and sampling fractions less than 0.01 in all strata. Our empirical analysis is restricted to the ACE-1 component of the survey, which meets all of the conditions described in the previous section.

The ACES publishes total and year-to-year change estimates. Estimates are published for the entire survey, and by industry code as indicated by the respondent units (not necessarily the industry code on the sampling frame). If there is no nonresponse, variances are estimated using the delete-a-group jackknife variance estimator (Kott 2001). To account for unit nonresponse, the ACE-1 component uses the ratio-adjustment procedure presented in Section 2 with administrative payroll data as the auxiliary variable x. Weighting cells are the design strata, provided that there is at least one respondent in the cell. Cell collapsing is extremely rare and is hereafter ignored in this paper. More details concerning the ACES survey design, methodology, and data limitations are available on-line at http://www.census.gov/csd/ace.

Although the ACE-1 survey design is fairly typical for a business survey, the collected data are not. Smaller companies often report legitimate values of zero for capital expenditures, and consequently the majority of the estimates are often obtained from the certainty and large non-certainty (2A) companies. As the capital expenditures are further cross-classified, the incidence of reported zeros (especially among smaller companies) increases.

3.2 Comparisons

To assess the effect of the unit non-response weight adjustment procedure on the ACE-1 standard errors, we computed variance estimates from unit nonresponse adjusted ACE-1 data using the ratio estimator with payroll as the auxiliary variable, in four industries, each with high sampling rates in the large company non-certainty strata (2A). The selected industries represent a cross-section of the sectors represented in the ACES. These industries and their North American Industrial Classification System (NAICS) codes are: Oil and Gas Extraction (211100), Nonmetallic Mineral Mining and Quarrying (212300), Other Miscellaneous Manufacturing (339900), and Architectural, Engineering, and Related Services (541300). In subsequent tables and discussions, industries are referred to by their NAICS code.

Table 1 presents variance estimate comparisons using five years' of ACE-1 survey data for three characteristics: the total capital expenditures (Total), capital expenditures on structures (Structures), and capital expenditures on equipment (Equipment). For comparison, the variance estimates are presented as a ratio to v_{J1} in Table 1. The estimated totals are also included. (Note that these totals are not the same as the published estimates, since they are computed using the industry classification on the frame, not the industry classification provided by the respondent).

As expected, the jackknife estimator v_{J1} and the linearization jackknife estimator v_{L1} are very close for all variables. The consistent variance estimators (v_L and v_{JL}) are all noticeably larger than their corresponding jackknife counterparts (v_{L1} and v_{J1}). In general, most capital expenditures are reported by certainty or large non-certainty companies, so effect on variance estimation of including non-respondent component in the variance estimator is noticeable. The jackknife estimator v_J, which adjusts for the effect of certainty strata, is generally between v_{J1} and v_{JL}. In some cases, v_J is equal to or very close to v_{J1}, indicating that the variability due to nonresponse mainly comes from non-certainty strata with large sampling fractions. The jackknife estimate $\tilde{v}_J$, which ignores sampling fractions, is much larger than any other estimates.

4. Simulation results

In this section, we present a simulation study using data modeled from the ACE-1 industries presented in the previous section. Section 4.1 describes the simulation settings. Section 4.2 presents and summarizes the results.

4.1 Simulation settings

We modeled our population using respondent data from the 2003 data collection of the three key items collected by the survey (Total, Structures, and Equipment). Frame data for the auxiliary variable (payroll) were available for all units. The complete population data were generated using the SIMDAT algorithm (Thompson 2000) with modeling cells equal to sampling strata and population size equal to the original frame size in each cell. Table 2 provides sampling fractions and correlation coefficients with the payroll for the modeled data in each stratum.

In the simulation, stratified simple random samples were selected from the generated population. We examine the statistical properties of the six variance estimators described in Section 2 over repeated samples under the following two different response mechanisms applied to the sample data:

1. The covariate-dependent response mechanism obtained by randomly applying response propensities modeled from the survey data with payroll as the covariate, which yields very high probabilities of responding to the large units and very small probabilities to the small (non-certainty) units;
2. The within-stratum uniform response mechanism obtained by using the observed survey response rate as the within-stratum response probability.

On the average, response probabilities in the individual stratum within industry were 0.85, 0.76. 0.77, 0.76, and 0.68 for strata 10, 2A, 2B, 2C, and 2D, respectively.

We selected 5,000 samples from the population, computed $\hat{Y}$ in (1) from each sample with nonresponse and weight adjustment, and computed the empirical mean and variance of the 5,000 $\hat{Y}$ values. This was done for each industry and each item, with two adjustment methods: the ratio and count estimators. When $\hat{Y}$ is the ratio estimator using the payroll as the auxiliary variable, the absolute value of the empirical relative bias is under 1.4% and is smaller than 1% in most cases. For the count estimator under the within-stratum uniform response mechanism, its absolute value of the empirical relative bias is under 0.5%. The count estimator is not approximately unbiased in theory under the covariate-dependent response mechanism. In the simulation, however, its absolute value of the empirical relative bias is under 1% in most cases and has a maximum value of 2.7%. The empirical variance of the 5,000 $\hat{Y}$ values was used as the "true value" of the variance of $\hat{Y}$.

4.2 Results

In 2,000 of the 5,000 samples, we computed the six different variance estimates for all three items, four industries, and two weight adjustment methods. We examined the statistical properties of each of variance estimator over repeated samples using the relative bias (RB) defined as

$$\frac{\text{the average of } 2{,}000 \text{ variance estimates}}{\text{the true variance}} - 1,$$

the stability (ST) defined as

$$\frac{\sqrt{\text{the empirical mean squared error of variance estimate}}}{\text{the true variance}},$$

and the error rate (ER) defined as the empirical proportion of the approximate 90% confidence intervals ($\hat{Y} \pm 1.645\sqrt{\text{variance estimate}}$) from 2,000 samples that do not contain the true population total.

Tables 3 and 4 respectively report the simulation results under the two response mechanisms. The results from these tables can be summarized as follows.

1. Two variance estimators ignoring V_2, v_{J1} and v_{L1}, have large negative relative biases in general. The error rates of the related confidence intervals are also large.
2. Two consistent variance estimators, v_L and v_{JL}, have very similar performances and are generally much better than v_{J1} and v_{L1} in terms of the relative bias and the error rate of the related confidence intervals.
3. The jackknife variance estimator v_J performs well in industries 339900 and 541300, but may have large positive relative biases in industries 211000 and 212300. We think that this is a "small sample" effect, since v_J is justified by asymptotic consistency and the sizes of the certainty strata in industries 211000 and 212300 are 26 and 30, respectively (Table 2). The sizes of the certainty

strata for the other two industries are 158 and 160, respectively. In fact, the performance of v_L and v_{JL} is generally better in industries 339900 and 541300.

4. In some cases v_J has more than 10% negative relative biases, which is caused by the fact that some non-certainty strata have large sampling fractions, *i.e.*, the approximation (6) does not hold enough.
5. The jackknife variance estimator $\tilde{v}_J$ ignoring all sampling fractions has very large positive relative biases and is too conservative.

5. Concluding remarks

When nonresponse is present in certainty strata (or strata with large sampling fractions), the jackknife and the linearization variance estimators that ignore certainty strata (or strata with large sampling fractions) are not acceptable because of their large negative biases. We derive two asymptotically unbiased and consistent variance estimators by adding an extra term that accounts the variability from nonresponse in certainty strata (or strata with large sampling fractions). We also derive a modified jackknife estimator that is consistent when the certainty strata are the only strata that contribute to the variance due to nonresponse (*i.e.*, Assumption (6) holds).

Our simulation results show that the three derived variance estimators perform well when stratum sample sizes are all large and perform inconsistently otherwise, and that the jackknife variance estimator that ignores all sampling fractions is very conservative.

Compared with the linearization method, the jackknife requires more computational resources but it has other advantages such as being easy to program, using a single recipe for different problems, and not requiring complicated or separate derivations for different estimators. Our linearization variance estimator given in (4) is in fact obtained by linearizing the jackknife estimator in (3).

Table 1
Variance estimates for $\hat{Y}$ with ratio adjustment in ACE-1 survey

Industry	Item	Year	$\hat{Y}$	v_{J1}	$\frac{v_{L1}}{v_{J1}}$	$\frac{v_L}{v_{J1}}$	$\frac{v_{JL}}{v_{J1}}$	$\frac{v_J}{v_{J1}}$	$\frac{\tilde{v}_J}{v_{J1}}$
211000	Total	2002	1.63E+7	4.63E+11	0.97	1.14	1.17	1.00	17.3
		2003	2.28E+7	6.87E+12	0.95	1.21	1.26	1.00	2.81
		2004	2.30E+7	2.45E+12	0.98	1.23	1.25	1.00	4.77
		2005	3.08E+7	4.29E+12	0.98	1.22	1.24	1.19	4.77
		2006	4.18E+7	6.29E+12	0.99	1.17	1.19	1.00	8.78
	Structures	2002	1.31E+7	3.99E+11	0.97	1.14	1.17	1.00	15.3
		2003	1.86E+7	5.78E+12	0.94	1.22	1.27	1.00	2.78
		2004	1.70E+7	8.39E+11	0.99	1.42	1.43	1.00	11.3
		2005	2.64E+7	3.84E+12	0.98	1.22	1.24	1.16	4.64
		2006	3.55E+7	5.41E+12	0.99	1.19	1.21	1.00	8.76
	Equipment	2002	3.20E+6	6.14E+10	0.98	1.15	1.17	1.00	7.26
		2003	4.18E+6	8.39E+11	0.97	1.22	1.24	1.00	1.70
		2004	6.01E+6	1.54E+12	0.97	1.13	1.16	1.00	1.39
		2005	4.33E+6	1.34E+11	0.97	1.22	1.25	1.15	6.17
		2006	6.31E+6	7.14E+11	0.99	1.12	1.13	1.00	2.68
212300	Total	2002	1.56E+6	4.14E+10	0.81	1.06	1.24	1.20	3.19
		2003	1.33E+6	1.21E+10	0.94	1.18	1.24	1.36	5.43
		2004	2.01E+6	2.86E+10	0.96	1.60	1.65	2.20	6.04
		2005	1.96E+6	1.93E+10	0.98	1.12	1.14	2.30	6.04
		2006	2.28E+6	2.19E+10	0.96	1.26	1.30	3.22	11.7
	Structures	2002	2.22E+5	4.36E+8	1.00	1.11	1.11	1.64	8.61
		2003	1.49E+5	2.27E+8	0.96	1.28	1.32	1.48	7.32
		2004	4.14E+5	1.03E+8	0.96	46.6	46.6	75.3	426
		2005	2.23E+5	9.33E+8	0.99	1.12	1.13	1.32	1.88
		2006	2.20E+5	1.88E+9	0.97	1.20	1.22	1.19	2.29
	Equipment	2002	1.33E+6	4.05E+10	0.81	1.06	1.25	1.15	2.86
		2003	1.18E+6	1.13E+10	0.94	1.20	1.26	1.32	5.07
		2004	1.60E+6	2.82E+10	0.96	1.40	1.44	1.53	3.30
		2005	1.73E+6	1.62E+10	0.97	1.16	1.19	2.33	6.69
		2006	2.06E+6	2.14E+10	0.96	1.26	1.30	2.94	10.8

Table 1 (continued)
Variance estimates for $\hat{Y}$ with ratio adjustment in ACE-1 survey

Industry	Item	Year	$\hat{Y}$	v_{J1}	$\frac{v_{L1}}{v_{J1}}$	$\frac{v_L}{v_{J1}}$	$\frac{v_{JL}}{v_{J1}}$	$\frac{v_J}{v_{J1}}$	$\frac{\tilde{v}_J}{v_{J1}}$
339900	Total	2002	1.75E+6	1.94E+10	0.99	1.27	1.29	1.10	3.71
		2003	1.58E+6	2.99E+10	0.98	1.24	1.27	1.10	1.60
		2004	1.70E+6	1.00E+10	0.99	1.40	1.40	1.69	4.61
		2005	1.77E+6	2.55E+10	0.99	1.28	1.29	1.25	3.02
		2006	1.94E+6	5.51E+10	0.99	1.23	1.25	1.12	2.15
	Structures	2002	2.99E+5	1.21E+9	0.99	1.24	1.24	1.09	3.55
		2003	1.93E+5	8.54E+8	0.99	1.27	1.28	1.09	1.75
		2004	2.10E+5	2.00E+8	0.99	1.86	1.87	2.08	5.89
		2005	2.56E+5	5.07E+8	0.99	1.80	1.81	1.97	9.61
		2006	5.97E+5	4.93E+10	0.99	1.19	1.20	1.01	1.16
	Equipment	2002	1.45E+6	1.62E+10	0.99	1.27	1.28	1.07	3.02
		2003	1.39E+6	2.71E+10	0.97	1.24	1.27	1.09	1.58
		2004	1.49E+6	9.14E+9	0.99	1.40	1.41	1.62	4.61
		2005	1.51E+6	2.45E+10	0.99	1.22	1.23	1.15	2.12
		2006	1.34E+6	5.65E+9	0.99	1.42	1.43	1.60	6.20
541300	Total	2002	3.38E+6	2.32E+10	0.99	1.47	1.48	1.67	5.02
		2003	3.09E+6	2.61E+10	0.99	1.26	1.27	1.05	1.62
		2004	3.97E+6	1.12E+11	1.00	1.23	1.23	1.03	1.37
		2005	4.94E+6	2.54E+11	1.00	1.20	1.20	1.04	1.71
		2006	4.96E+6	2.82E+10	1.00	1.40	1.40	1.75	8.36
	Structures	2002	7.41E+5	6.32E+9	1.00	1.70	1.71	1.64	7.47
		2003	4.29E+5	3.32E+9	1.00	1.29	1.29	1.01	1.33
		2004	6.96E+5	4.38E+10	1.00	1.22	1.22	1.00	1.40
		2005	7.12E+5	9.00E+9	1.00	1.25	1.25	1.08	2.08
		2006	8.73E+5	3.44E+9	1.00	1.58	1.59	1.63	9.88
	Equipment	2002	2.96E+6	1.39E+10	0.99	1.37	1.38	1.54	3.95
		2003	2.66E+6	1.94E+10	0.99	1.25	1.26	1.05	1.59
		2004	3.27E+6	5.83E+10	1.00	1.22	1.23	1.04	1.29
		2005	4.23E+6	2.40E+11	1.00	1.19	1.20	1.03	1.59
		2006	4.09E+6	2.35E+10	1.00	1.27	1.28	1.49	5.47

Table 2
Population characteristics for the simulation study

Industry	Stratum	Population size	Sampling fraction	Correlation with Payroll		
				Total	Structures	Equipment
211000	10	26	1.00	0.65	0.53	0.95
	2A	128	0.77	0.68	0.66	0.22
	2B	372	0.11	0.57	0.51	0.51
	2C	1,800	0.02	-0.07	0.00	-0.10
	2D	10,406	0.00	0.28	0.00	0.28
212300	10	30	1.00	0.96	0.95	0.94
	2A	108	0.37	0.85	0.74	0.77
	2B	414	0.07	0.03	0.76	-0.03
	2C	1,310	0.03	0.42	0.13	0.43
	2D	4,762	0.01	0.44	-0.22	0.44
339900	10	158	1.00	0.80	0.40	0.80
	2A	498	0.26	0.40	0.04	0.51
	2B	2,048	0.05	0.20	0.24	0.18
	2C	6,310	0.02	0.19	0.48	0.09
	2D	25,288	0.00	0.37	0.67	0.36
541300	10	160	1.00	0.60	0.56	0.59
	2A	959	0.38	0.20	0.39	0.06
	2B	4,531	0.06	0.28	0.13	0.27
	2C	17,913	0.01	0.08	0.06	0.08
	2D	67,440	0.00	0.13	-0.01	0.15

Table 3
Simulation results (in %) for variance estimation under covariate-dependent response mechanism

Estimate	Item	Industry		v_{J1}	v_{L1}	v_L	v_{JL}	v_J	$\tilde{v}_J$
Ratio	Total	211000	RB	-35.8	-38.1	-10.3	-8.0	39.1	113.9
			ST	49.8	50.4	47.4	48.6	252.9	182.9
			ER	19.6	19.8	12.2	11.8	10.7	1.1
		212300	RB	-20.4	-22.2	-4.48	-2.69	54.8	266.4
			ST	30.3	31.1	26.8	27.3	139.1	268.8
			ER	12.6	12.6	9.9	9.6	6.3	0.1
		339900	RB	-21.2	-22.5	0.26	1.55	-5.34	52.5
			ST	47.3	47.0	55.0	56.0	43.9	67.8
			ER	14.3	14.6	10.4	10.3	10.0	2.6
		541300	RB	-20.7	-21.0	3.83	4.08	-11.6	18.4
			ST	32.7	32.8	34.9	35.0	29.4	32.0
			ER	12.6	12.7	8.6	8.6	10.7	6.2
	Structures	211000	RB	-38.0	-40.1	-11.9	-9.59	33.9	108.1
			ST	51.3	51.9	48.5	49.6	244.4	180.8
			ER	20.9	21.1	12.9	12.6	11.1	1.1
		212300	RB	-23.2	-23.9	-12.4	-11.6	33.2	341.5
			ST	31.5	32.0	27.1	27.0	95.0	344.3
			ER	12.3	12.3	10.4	10.3	6.9	0.1
		339900	RB	-20.0	-20.4	-6.31	-5.88	-10.9	39.8
			ST	42.5	42.7	42.3	42.3	39.9	64.0
			ER	15.9	16.0	12.7	12.6	13.2	5.4
		541300	RB	-20.0	-20.1	0.09	0.33	-15.9	15.7
			ST	42.6	42.5	50.5	50.7	41.1	42.7
			ER	13.1	13.2	9.9	9.9	12.0	6.5
	Equipment	211000	RB	-15.0	-17.3	14.1	16.4	-9.37	27.9
			ST	63.9	62.6	87.7	90.0	64.1	69.6
			ER	16.2	16.7	13.3	13.0	14.7	6.7
		212300	RB	-21.4	-23.3	-4.13	-2.21	39.7	201.1
			ST	31.7	32.5	28.4	29.0	113.7	204.4
			ER	13.3	13.5	10.2	10.1	7.7	0.2
		339900	RB	-21.4	-22.8	1.18	2.57	-7.29	50.8
			ST	51.2	50.9	60.8	61.9	47.9	69.2
			ER	15.5	15.8	11.6	11.4	11.0	2.3
		541300	RB	-19.7	-19.9	6.16	6.43	-11.9	12.8
			ST	33.8	33.9	38.4	38.5	31.0	30.9
			ER	12.5	12.5	8.9	8.9	11.0	7.0
Count	Total	211000	RB	-30.1	-31.9	0.05	1.85	1.05	103.1
			ST	50.4	50.5	55.9	57.3	46.7	113.4
			ER	15.3	15.6	9.0	8.8	8.7	1.0
		212300	RB	-33.2	-34.6	-6.30	-4.96	17.6	204.5
			ST	38.7	39.6	27.7	27.8	42.8	208.6
			ER	14.1	14.7	9.1	8.7	6.9	0.4
		339900	RB	-23.9	-24.6	1.73	2.44	-14.2	46.4
			ST	47.5	47.4	55.2	55.7	43.4	62.4
			ER	13.4	13.5	9.1	9.1	10.7	2.5
		541300	RB	-22.9	-23.2	1.68	1.94	-18.8	15.4
			ST	32.9	33.0	32.0	32.2	30.2	28.9
			ER	11.5	11.6	7.2	7.1	10.6	5.2
	Structures	211000	RB	-30.3	-32.2	-0.15	1.65	-1.27	101.5
			ST	51.3	51.3	57.3	58.7	46.7	112.3
			ER	15.8	16.3	9.6	9.4	9.2	0.8
		212300	RB	-37.4	-38.0	-13.5	-12.9	3.53	250.2
			ST	41.6	42.0	28.9	28.8	32.2	254.8
			ER	15.4	15.6	9.6	9.5	8.1	0.4
		339900	RB	-20.0	-20.3	-4.33	-4.00	-14.5	38.6
			ST	42.3	42.4	42.4	42.4	40.1	62.8
			ER	14.6	14.7	11.9	11.8	13.6	5.0
		541300	RB	-20.9	-21.2	-0.54	-0.32	-18.9	14.5
			ST	41.6	41.6	47.8	48.0	40.6	40.9
			ER	12.5	12.5	9.1	9.1	12.1	6.0
	Equipment	211000	RB	-17.8	-20.0	11.2	13.3	-13.0	26.6
			ST	58.9	58.0	76.4	78.4	57.7	64.1
			ER	15.7	15.8	12.5	12.3	14.5	6.1
		212300	RB	-30.7	-32.2	-4.74	-3.27	12.1	164.3
			ST	37.6	38.6	29.1	29.3	38.7	168.9
			ER	14.1	14.5	9.6	9.5	7.9	0.6
		339900	RB	-24.1	-24.9	2.52	3.27	-15.2	45.0
			ST	51.2	51.1	61.0	61.5	47.7	64.1
			ER	14.8	15.1	9.9	9.8	11.9	2.3
		541300	RB	-21.6	-21.9	4.10	4.39	-18.1	10.1
			ST	33.6	33.7	35.2	35.3	31.5	28.2
			ER	11.1	11.1	7.2	7.1	10.3	5.9

Table 4
Simulation results (in %) for variance estimation under within-stratum uniform response mechanism

Estimate	Item	Industry		v_{J1}	v_{L1}	v_L	v_{JL}	v_J	$\tilde{v}_J$
Ratio	Total	211000	RB	-49.2	-50.4	-17.2	-16.0	89.2	138.4
			ST	55.4	56.1	43.7	43.9	310.5	258.3
			ER	26.9	27.1	13.9	13.8	9.10	1.80
		212300	RB	-5.42	-7.99	16.1	18.7	111.7	337.2
			ST	28.9	28.5	37.4	39.6	179.2	341.7
			ER	13.5	13.8	9.85	9.50	5.85	0.05
		339900	RB	-9.37	-10.5	18.0	19.2	16.5	83.8
			ST	45.8	45.4	59.0	60.1	48.4	95.6
			ER	14.5	14.7	9.55	9.55	8.60	2.65
		541300	RB	-8.83	-9.03	18.2	18.4	6.62	44.5
			ST	26.7	26.8	36.6	36.7	28.2	52.3
			ER	12.6	12.6	8.45	8.45	9.70	5.35
	Structures	211000	RB	-52.6	-53.7	-19.2	-18.0	78.4	128.0
			ST	58.0	58.7	45.3	45.4	290.8	248.5
			ER	28.7	29.0	15.1	14.9	9.80	2.25
		212300	RB	-16.2	-18.1	9.92	11.9	63.5	356.2
			ST	32.0	32.4	37.1	38.5	108.6	361.9
			ER	15.5	16.0	10.9	10.7	6.65	0.35
		339900	RB	-13.2	-13.6	13.9	14.3	1.15	54.9
			ST	47.8	47.8	59.2	59.5	46.3	82.7
			ER	17.2	17.2	12.6	12.5	13.8	6.40
		541300	RB	-8.9	-9.2	19.2	19.5	-2.22	36.0
			ST	39.4	39.3	53.7	54.0	38.6	55.9
			ER	12.9	12.9	8.85	8.85	11.3	6.85
	Equipment	211000	RB	-1.1	-2.75	27.5	29.1	12.8	60.1
			ST	64.4	63.2	88.6	90.3	71.1	89.8
			ER	15.3	15.6	12.0	12.0	12.7	5.10
		212300	RB	-6.3	-8.96	16.8	19.4	90.0	263.1
			ST	30.3	29.8	39.3	41.6	148.6	269.1
			ER	13.9	14.2	10.1	9.60	6.75	0.15
		339900	RB	-8.84	-10.1	19.5	20.7	15.8	84.6
			ST	50.8	50.3	65.7	66.9	52.9	98.8
			ER	15.1	15.3	10.3	10.3	9.50	2.45
		541300	RB	-6.89	-7.1	19.9	20.1	6.76	38.5
			ST	28.6	28.6	40.0	40.1	30.1	48.3
			ER	12.4	12.4	8.60	8.55	10.3	5.90
Count	Total	211000	RB	-27.8	-29.0	14.2	15.4	16.3	149.5
			ST	47.4	47.5	53.1	54.2	44.4	158.1
			ER	16.0	16.2	8.30	8.30	7.45	1.85
		212300	RB	-33.5	-34.9	15.3	16.7	23.9	219.9
			ST	40.0	40.9	38.5	39.5	39.4	228.5
			ER	18.8	19.3	9.80	9.65	8.55	1.90
		339900	RB	-16.5	-17.1	20.2	20.8	4.21	75.7
			ST	45.1	44.0	57.9	58.5	42.4	87.6
			ER	15.6	15.8	9.40	9.35	10.8	3.20
		541300	RB	-9.61	-9.81	18.9	19.1	-0.77	45.0
			ST	26.5	26.5	36.0	36.1	24.7	52.2
			ER	12.4	12.4	8.55	8.55	11.3	4.85
	Structures	211000	RB	-27.5	-28.7	14.5	15.7	14.6	149.0
			ST	48.1	48.1	54.5	55.6	45.1	157.6
			ER	17.1	17.5	9.05	9.00	8.50	2.05
		212300	RB	-39.4	-40.4	11.6	12.6	10.1	238.5
			ST	44.8	45.5	38.8	39.6	32.1	248.4
			ER	20.2	20.7	9.95	9.85	10.3	1.80
		339900	RB	-14.2	-14.6	13.6	14.0	-3.55	53.5
			ST	47.1	47.0	57.3	57.6	45.1	80.5
			ER	17.6	17.7	12.1	12.1	14.7	6.30
		541300	RB	-9.54	-9.76	20.0	20.2	-5.32	36.0
			ST	39.1	39.0	53.3	53.5	38.3	55.8
			ER	12.6	12.6	9.05	9.05	11.9	6.55
	Equipment	211000	RB	-8.12	-9.64	22.7	24.2	1.56	54.3
			ST	58.0	57.1	76.2	77.7	57.5	82.1
			ER	16.5	16.7	12.4	12.4	14.6	6.45
		212300	RB	-28.5	-30.0	17.1	18.6	21.5	189.7
			ST	37.6	38.4	40.9	42.0	38.2	198.9
			ER	18.1	18.5	9.95	9.80	9.25	1.75
		339900	RB	-15.8	-16.4	21.8	22.5	4.69	76.8
			ST	49.5	49.3	64.6	65.2	47.4	91.1
			ER	16.4	16.5	9.45	9.40	11.4	3.20
		541300	RB	-7.53	-7.74	20.2	20.4	0.26	38.8
			ST	28.2	28.2	39.2	39.4	27.2	48.0
			ER	12.7	12.7	8.30	8.25	11.3	5.65

Acknowledgements

This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical or methodological issues are those of the authors and not necessarily those of the U.S. Census Bureau. The authors thank two referees and an associate editor for their helpful comments and suggestions, and Carol Caldwell, Rita Petroni, and Mark Sands for their useful comments on an earlier version of this paper. Jun Shao's research was partially supported by the NSF Grant SES-0705033.

References

Fay, R.E. (1991). A design-based perspective on missing data variance. *Proceedings of the 1991 Annual Research Conference*, U.S. Bureau of the Census, 429-440.

Hidiroglou, M.A., and Särndal, C.-E. (1998). Use of auxiliary information for two-phase sampling. *Survey Methodology*, 24, 11-20.

Kott, P. (1994). A note on handling nonresponse in sample surveys. *Journal of the American Statistical Association*, 89, 693-696.

Kott, P. (2001). The delete-a-group jackknife. *Journal of Official Statistics*, 17, 521-526.

Krewski, D., and Rao, J.N.K. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *Annals of Statistics*, 9, 1010-1019.

Oh, H.L., and Scheuren, F.J. (1983). Weighting adjustment of unit nonresponse. *Incomplete Data in Sample Surveys*. New York: Academic Press, 20, 143-184.

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

Shao, J., and Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association*, 94, 254-265.

Shao, J., and Tu, D. (1995). *The Jackknife and Bootstrap*. New York: Springer-Verlag.

Thompson, J.R. (2000). *Simulation*: *A Modeler's Approach*. New York: John Wiley & Sons, Inc.

Thompson, K.J., and Yung, W. (2006). To replicate (a weight adjustment procedure) or not to replicate? An analysis of the variance estimation effects of a shortcut procedure using the stratified jackknife. *Proceedings of the Section of Survey Research Methods*, American Statistical Association, 3772-3779.

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, 3553-3558.

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