VARIANCE ESTIMATION USING AUXILIARY INFORMATION UNDER A ONE UNIT PER STRATUM SAMPLE DESIGN

Elizabeth T. Huang and Cary T. Isaki, U.S. Bureau of the Census[1]

1. INTRODUCTION

Estimators of the variance of estimators of total or mean using auxiliary information under a one primary sampling unit (PSU) per stratum sample design have been proposed by several authors. Under a one PSU per stratum sample design, the most common variance estimator of the usual estimator of total is the collapsed stratum variance estimator (Cochran (1977)) which has positive bias if the true stratum totals of the collapsed pairs differ to a large extent. One way to reduce the bias of the collapsed stratum variance estimator is to form stratum pairs so that the true stratum totals of the characteristics under study are as similar as possible. Hansen, Hurwitz and Madow (1953) proposed the use of an auxiliary variate in conjunction with a collapsed stratum variance estimator and showed that the resulting estimator was positively biased. Hartley, Rao and Kiefer (1969) developed a variance estimator based on an assumed linear relationship between the true stratum means and some auxiliary variables. Isaki (1983a) developed a class of variance estimators using auxiliary information in the form of the variance of known variables to reduce the bias of the collapsed stratum estimator.

In the Spring of 1983, we designed a sampling scheme for a content evaluation survey of the 1982 Economic Censuses. The content evaluation survey was intended to measure the accuracy of the reported establishment data for employment, payroll, and receipts of the 1982 Census of Wholesale - Wholesale petroleum distributors, Standard Industrial Classifications (SIC's) 5171 (petroleum bulk stations and terminals) and 5172 (petroleum and petroleum products wholesalers, except bulk stations and terminals). For each SIC code, the universe of establishments consisted of single-unit and multiunit establishments. Multiunit establishments are establishments affiliated with firms consisting of two or more establishments. The sampling design treats "large" payroll and employment establishments as certainty; uses stratified simple random sampling without replacement for selection of multiunit establishments, and a one PSU per stratum probability proportional to size (PPS) two stage sample design for the selection of single-unit establishments.

The research that follows was motivated for several reasons. First, we have some auxiliary information from the 1977 Economic Censuses that could be used in variance estimation. Second, we were interested in comparing the different variance estimators of the estimator of total or ratio for the characteristics in the content evaluation survey. Finally, we sought to recommend a variance estimator for use in the analysis of the content evaluation survey. In the following, we compare several variance estimators for estimators of total and ratio, respectively, by a Monte Carlo study.

II. ESTIMATING THE VARIANCE OF A TOTAL

II.A. Variance Estimators for Total Under a One Unit Per Stratum Sample Design

Given a PPS one unit per stratum sample design with n strata, several potential bias-reducing variance estimators using auxiliary information were considered for the Monte Carlo study.

Let Y_{hi}, $h=1, 2, \ldots n$, $i=1, 2, \ldots, N_h$ denote the value of the characteristic of interest associated with the i-th primary sampling unit (PSU) in the h-th stratum, and N_h denote the number of PSUs in the h-th stratum. Let Z_{hi} be the corresponding auxiliary variate. Let P_{hi} denote the selection probability of the i-th PSU in the h-th stratum.

To estimate the total for characteristics y and z, under a single stage one PSU per stratum PPS sampling design, the unbiased estimator of totals Y and Z were

$$\hat{Y} = \sum_{h=1}^{n} P_{hi}^{-1} Y_{hi}, \text{ and } \hat{Z} = \sum_{h=1}^{n} P_{hi}^{-1} Z_{hi}$$

respectively, with covariance of $\hat{Y}$ and $\hat{Z}$

$$\mathrm{Cov}(\hat{Y}, \hat{Z}) = \sum_{h=1}^{n} \sum_{i=1}^{N_h} P_{hi}[P_{hi}^{-1} Y_{hi} - Y_h][P_{hi}^{-1} Z_{hi} - Z_h], \quad (2.0)$$

where Y_h and Z_h are the stratum totals.

Let the n strata be grouped into G pairs in a manner to be described later. When the number of strata n is odd, at least one group must consist of three strata.

Several proposed variance estimators for estimating $V(\hat{Y})$ under a one PSU per stratum sample design considered for the Monte Carlo study are as follows:

1. The collapsed stratum variance estimator. (See Cochran (1977)).

$$\hat{V}_{CS}(\hat{Y}) = \sum_{j=1}^{G} L_j(L_j-1)^{-1} \sum_{k=1}^{L_j} (\hat{Y}_{jk} - \hat{Y}_j/L_j)^2 \quad (2.1)$$

where $\hat{Y}_{jk}$ is the estimated stratum total, $\hat{Y}_j$ is the estimated total for group j, L_j is the number of strata in the j-th group.

2. The Hansen, Hurwitz, and Madow (1953) collapsed stratum variance estimator

$$\hat{V}_{CSA}(\hat{Y}) = \sum_{j=1}^{G} L_j(L_j-1)^{-1} \sum_{k=1}^{L_j} (\hat{Y}_{jk} - A_{jk}A_j^{-1} \hat{Y}_j)^2 \quad (2.2)$$

where A_{jk} is some measure of stratum level highly correlated with Y_{jk}, and A_j is the sum over the strata in the group j.

3. Isaki's variance estimators (1983a) are

(a) $$\hat{V}_G(\hat{Y}) = \hat{V}_{CS}(\hat{Y}) + \hat{B}_W^2 \{V(\hat{Z}) - \hat{V}_{CS}(\hat{Z})\} \quad (2.3)$$

where

$$\hat{B}_w = [\sum_{i=1}^{n} (z_i/P_i - \bar{z}_p)^2]^{-1}$$

$$\times [\sum_{i=1}^{n} (z_i/P_i-\bar{z}_p)(y_i/P_i-\bar{y}_p)] , \qquad (2.3a)$$

$$\bar{z}_p = n^{-1}\sum_{i=1}^{n} z_i/P_i , \ \bar{y}_p = n^{-1}\sum_{i=1}^{n} y_i/P_i .$$

The auxiliary variables in (2.3) are the selection probability P and Z. y_i and z_i are the sampled PSU totals from each stratum i. $V(\hat{Z})$ is assumed to be known.

(b) $\hat{V}_{GR1}(\hat{Y}) = \hat{V}_{CS}(\hat{Y})+\{\hat{B}_0^2[V(\hat{Z}_0)-\hat{V}_{CS}(\hat{Z}_0)] + \hat{B}_1^2 [V(\hat{Z}) - \hat{V}_{CS} (\hat{Z})]\}$ (2.4)

where

$$\hat{B}_1 = [\sum_{i=1}^{n} P_i^{-1} (z_i - \bar{z}_w)^2]^{-1}$$

$$\times [\sum_{i=1}^{n} P_i^{-1}(z_i-\bar{z}_w)(y_i-\bar{y}_w)] , \qquad (2.4a)$$

$$\hat{B}_0 = \bar{y}_w - \hat{B}_1 \bar{z}_w ,$$

$$\bar{y}_w = (\sum_{i=1}^{n} P_i^{-1} y_i) (\sum_{i=1}^{n} P_i^{-1})^{-1} ,$$

$\bar{z}_w$ is defined similarly as $\bar{y}_w$.

$z_{0i} = 1$, $i=1,\ldots,n$.

(c) $\hat{V}_{GR2}(\hat{Y}) = \hat{V}_{GR1}(\hat{Y}) + 2 \hat{B}_0\hat{B}_1[Cov(\hat{Z},\hat{Z}_0) - \hat{Cov}_{CS}(\hat{Z},\hat{Z}_0)]$ (2.5)

where $Cov(\hat{Z},\hat{Z}_0)$ and $V(\hat{Z})$ are assumed to be known.

The estimators $\hat{V}_{GR1}(\hat{Y})$ and $\hat{V}_{GR2}(\hat{Y})$ assume that a column of ones, z_0, is also used as auxiliary variables and they differ only in that the covariance term appearing in $\hat{V}_{GR2}$ was arbitrarily deleted from $\hat{V}_{GR1}$.

4. Hartley, Rao and Kiefer's variance estimator

$$\hat{V}_{HRK}(\hat{Y}) = \sum_{h=1}^{n} N_h^2 \hat{\sigma}_h^2 = v'\hat{\delta}^2, \quad (2.6)$$

where $\hat{\sigma}_h^2$ is the stratum variance estimator for $\hat{\bar{Y}}_h = (N_hP_{hi})^{-1} Y_{hi}$,

$v' = (N_1^2,\ldots,N_n^2)$, $\hat{\delta}^2 = C^{-1}D$, where D is an n-vector whose elements are squares of residuals resulting from fitting the sample strata means of y to an auxiliary variable z at stratum level; the auxiliary variable z is a fixed value which does not depend on the particular sample drawn. D and C are defined in Hartley, Rao and Kiefer (1969), equations (10), (14) and (15). All variance estimators listed above are biased. We are interested in comparing the bias and MSE of all of the variance estimators through a Monte Carlo study using 1977 and 1982 Economic Census data.

II.B.Monte Carlo Study of $\hat{V}(\hat{Y})$

Data from both the 1977 and 1982 Economic Censuses covering SIC 5171 and SIC 5172 were used in the study. While the sample design of 1982 Economic Census content evaluation survey covered both single and multiunit establishments, it was the single-unit segment of the universe that was subject to the one PSU per stratum design and is of primary interest here. The sampling frame for the single-unit sample design is the 1982 Economic Census mail control file for single unit in SIC 5171 and 5172. For the single-unit establishment, a two-stage PPS one PSU per stratum sample design was used. Establishments in the single-unit stratum were arranged in 128 PSU's which were formed from groups of contiguous counties. A small number of single units were separated from the PSU's and included in the sample with certainty due to their large size. The PSU's were stratified by multiple variates via Spark's algorithm, which employed a Euclidean cluster analysis. Since only 15 interviewers could be afforded for single unit field interview, two PSU's were chosen as certainty PSU's, the rest of the 126 PSU's were stratified into 13 strata. The 13 strata of PSU's were constructed by first using SIC 5171 1982 employment data as an initial stratification variable. Spark's algorithm was then used to form 13 strata using the four stratification variables; 1982 employment and 1st quarter payroll of SIC 5171 and 5172. The 13 noncertainty strata contained from 5 to 18 PSU's. One PSU was selected from each stratum with probability proportional to 1982 Economic Census first quarter payroll of SIC 5171 at the first stage of sampling, and a stratified random sample design within the selected PSU was used at the second stage of sampling. A detailed description of the sample design can be found in Isaki (1983b).

The sample design in the Monte Carlo study is a single stage PPS one PSU per stratum design. The sample frame in the Monte Carlo study is the 1977 Economic Census single-unit PSU file of SIC 5171 and 5172 (the 1982 Economic Census data were incomplete at the time of this research). There are 128 PSUs, and the same stratification and probability measurement used in the 1982 Economic Census content evaluation survey was used in the Monte Carlo study.

In the 1977 Economic Census single-unit PSU data, for each PSU, we have the PSU total of receipts, annual payroll and employment for SIC 5171 and SIC 5172, and the probability of selection for each PSU. We used this data to compare different variance estimators in the Monte Carlo study. For each SIC, for the characteristics y, receipts, annual payroll, and the number of employment, the auxiliary variables z used in the variance estimation are annual payroll, the number of employment and annual payroll respectively.

Given the auxiliary information z, one way to reduce the bias of $\hat{V}_{CS}(\hat{Y})$ is to sort the strata in increasing order of z. Isaki (1983a) found that the bias and MSE were both reduced if the strata were sorted on the basis of the auxiliary variable rather than listed in a random fashion. For each characteristic y, the 1977 Economic Census PSU file was sorted according to the stratum total of the associated auxiliary variable z. One thousand PPS one PSU per stratum samples were selected for each characteristic of interest. In calculating the six variance estimators of total $\hat{Y}$ for 1000 samples, the associated auxiliary variable z for each characteristic y was used in all variance estimators considered except collapsed stratum variance estimators. For the Hansen, Hurwitz, and Madow collapsed stratum variance estimator, the census stratum total of the associated z variable was used. For Isaki's variance estimators, the variance of the total $\hat{Z}, V(\hat{Z})$, calculated from the census, was used. For Hartley, Rao and Kiefer's variance estimators, the census stratum mean of the auxiliary variable z was used. It may be noted that the auxiliary variable, population stratum mean or population total Z, can be used for the ratio estimation of the total Y. In Hartley, Rao and Kiefer (1969), the variance estimator for the combined ratio estimator of Y, was given in equation (35) of their paper. Isaki's variance estimator can also be extended to the estimation of the variance of the combined ratio estimator of Y by a similar approach used in Isaki (1983a). (See Appendix A). The six estimators of variance were applied to each sample and their bias and mean square errors were calculated. Four of the six estimators (excluding $\hat{V}_{CS}(\hat{Y})$ and $\hat{V}_{CSA}(\hat{Y})$), can be negative. To protect against this negative variance, all estimators were arbitrarily set equal to $\hat{V}_{CSA}(\hat{Y})$ when negative. In our Monte Carlo study with 1000 samples, none of the four estimators was negative more than five percent of the time. As a matter of fact, for $\hat{V}_{HRK}$, there were no negative variance estimates. The results of the variance comparisons for receipts, annual payroll, and employment for SIC 5171 and SIC 5172 are provided ine Table I. The conclusion to be drawn was similar to that in Isaki (1983a), where estimator $\hat{V}_G(\hat{Y})$ performed better than $\hat{V}_{GR2}$ and $\hat{V}_{CSA}$ in terms of smallest MSE, and $\hat{V}_{CS}(\hat{Y})$ was never competitive with any of the estimators studied. In two of six characteristics, the V_G, V_{CSA} and V_{HRK} give the smallest relative bias.

III. ESTIMATING THE VARIANCE OF A RATIO

III.A. Variance Estimator of a Ratio Under a One PSU Per Stratum Design

The goal of the content evaluation survey of the 1982 Economic Census was to evaluate the accuracy of the census data from the wholesale petroleum distributors SIC 5171 and 5172 using content evaluation survey data.

In the content evaluation survey, the interviewer reinterviewed the sampled bulk petroleum distributors with different questionnaires from the economic census and the reinterviewed data were obtained. The ratio of the total from the reinterview (X) to the corresponding census total (Y) was used as a measure of accuracy for each characteristic. In the following section, we extended the variance estimators studied in Section II to estimators of the variance of the ratio under a one PSU per stratum sample design.

As mentioned previously, the sample frame of the content evaluation survey has two parts: the single-unit establishments, and the multiunit establishments. The single-unit establishment sample which was a two-stage PPS one PSU per stratum sample design, is of main interest.

Let $\hat{X}$ be the estimated total from the reinterview data, and $\hat{Y}$ be the estimated total from the census data. Then

$$\hat{X} = \hat{X}_s + \hat{X}_m + X_{sc} + X_{mc}$$

$$\hat{Y} = \hat{Y}_s + \hat{Y}_m + Y_{sc} + Y_{mc}$$

where $\hat{X}_s, \hat{Y}_s, \hat{X}_m, \hat{Y}_m$ are the estimated totals from the non-certainty strata of single-unit (indexed by s) and multiunit establishments (indexed by m) respectively, and $X_{sc}, Y_{sc}, X_{mc}, Y_{mc}$ are the totals from the certainty stratum (indexed by c) of single-unit and multiunit establishments respectively.

The estimate of the ratio of the reinterviewed data to the census data used is $\hat{R} = \hat{X}/\hat{Y}$. The variance of $\hat{R}$ is

$$V(\hat{R}) = [E(\hat{Y})]^{-2}\{V(\hat{X}) + R^2V(\hat{Y}) - 2R\ Cov(\hat{X},\hat{Y})\} \quad (3.1)$$

$$= [E(\hat{Y})]^{-2}\{V(m) + V(s)\} \quad ,$$

where

$$R = [E(\hat{Y})]^{-1}\ [E(\hat{X})] \ , \quad (3.2)$$

V(m) and V(s) are the variances with the same form inside the bracket of (3.1) from multi-unit and single unit respectively. Note that we may use a ratio estimator to estimate the reinterview total X by $\hat{X}_R = (\hat{X}/\hat{Y})\ Y$ but this was not of interest.

Our interest was in the estimation of the component of $V(\hat{R})$ from the single-unit establishment, V(s), where the sample design is a two-stage PPS one PSU per stratum design and where a stratified simple random sample of establishments was selected within each selected PSU.

To help in our deliberations, we compared six variance estimators of the variance of the ratio $\hat{R}$ under a single-stage PPS one PSU per stratum sample design via a Monte Carlo study using 1977 Economic Census single unit PSU data.

Let X_{hi}, Y_{hi} be the total of reinterviewed and census data from the i-th PSU and the h-th stratum from the single-unit sample, respectively.

The estimates for the single-unit total of reinterview and census data from the noncertainty strata are

$\hat{X}_s = \sum_{h=1}^{n} P_{hi}^{-1} X_{hi}$, $\hat{Y}_s = \sum_{h=1}^{n} P_{hi}^{-1} Y_{hi}$, respectively.

To estimate $\hat{V}(s)$, we needed variance estimators for $V(\hat{X}_s)$, $V(\hat{Y}_s)$, and a covariance estimator for $Cov(\hat{X}_s, \hat{Y}_s)$. All variance estimators presented in section II can be used to estimate $V(\hat{X}_s)$ and $V(\hat{Y}_s)$, and can be extended to covariance estimation of $Cov(\hat{X}_s, \hat{Y}_s)$. For example, the collapsed variance estimator defined in (2.1) can be extended to estimate the $Cov(\hat{X}_s, \hat{Y}_s)$ by replacing square term of y in (2.1) with the cross product term of y and x. Isaki's variance estimators can also be extended to the covariance estimation by a similar approach used in Isaki (1983a) (See Appendix A).

Hartley, Rao and Kiefer (1969) provided for covariance estimation in equation (34) of their paper.

III.B. Data Analysis

We considered 6 variance estimators for V(s) in (3.1). We sought the "best" variance estimator from among these 6 variance estimators for use in the content evaluation survey.

To assist in the comparison, a Monte Carlo study was carried out using data from the 1977 Economic Census single-unit PSU file. Lacking reinterview data for the entire universe, the reinterview data were simulated under a ratio model using the 1977 Economic Census PSU file and the ratio R was estimated from the 1977 Economic Census Content Evaluation Survey (see Corby (1984)). For example, for the characteristic of interest, (say, receipts) the reinterview receipts of SIC 5171 or SIC 5172 for PSU i were generated using

$$x_i = 0.8792\ y_i + e_i\ ,\ i=1,\dots,126,$$

where y_i is the 1977 economic census receipts total of the i-th PSU,

x_i is the reinterview receipts total of the i-th PSU,

0.8792 is the estimated ratio of the reinterview receipts and the 1977 economic census tabulation receipts from the wholesale trade of all single units in the U.S.,

e_i is a normal random variable with mean 0 and variance σ^2.

Several sets of reinterview receipts data at PSU level for SIC 5171 and SIC 5172 were simulated by varying σ^2, and hence the correlations of reinterview and 1977 economic census data differ accordingly. The same one thousand samples selected previously together with the selected simulated reinterview data were used. The auxiliary variables used for all 6 variance estimators except $\hat{V}_{CS}$ are the census annual payroll for SIC 5171 and SIC 5172, respectively. To compare different variance estimators of $V(\hat{R})$ under the single-unit sample design using the prescribed six variance estimators, the bias and MSE were calculated from 1000 samples, and the relative bias and MSE were tabulated in Table II. As before, when calculating $V(\hat{R})$ from the single-unit sample replicate, if any of the variance estimates of $V(\hat{X}_s)$ and $V(\hat{Y}_s)$ were negative, they were set equal to $\hat{V}_{CSA}(\hat{X}_s)$ and $\hat{V}_{CSA}(\hat{Y}_s)$ respectively.

Based on the Monte Carlo study for estimating the variance of the ratio of reinterview receipts versus census receipts, Hartley, Rao and Kiefer's variance estimator gave the smallest bias and MSE. The collapsed variance estimator is second best. Isaki's variance estimators using auxiliary variables performed poorly. This is probably because the simulated reinterview receipts were so highly correlated with the census receipts, ($\rho \geq 0.9497$), that there was little need for the auxiliary variable z to come in to reduce the residual error of $x_i - \hat{R}y_i$ fitted with z's. For the Hartley, Rao and Kiefer variance estimator of ratio, both the simulated reinterview receipts x and the census receipts y had a good linear relationship with the census payroll z at the stratum mean level.

As a result of this study, we recommended using either the Hartley, Rao and Kiefer's variance estimator of ratio or the collapsed variance estimator for the single-unit component of the variance $V(\hat{R})$ in the 1982 Economic Census Content Evaluation Survey. Since the collapsed variance estimator is easier to apply than the Hartley, Rao and Kiefer variance estimator, the collapsed variance estimators would appear to be more desirable.

IV. SUMMARY

This article compares several variance estimators for estimating the variance of the total or ratio under a one unit per stratum sample design. For estimating the variance of the total, Isaki's variance estimator, $\hat{V}_G(\hat{Y})$, performed better in terms of smallest MSE. For estimating the variance of the ratio, Hartley, Rao and Kiefer's variance estimator gave the smallest bias and MSE. The collapsed variance estimator is second best. We recommended using the collapsed variance estimator for variance estimation in the content evaluation survey.

REFERENCES

1. Cochran, W.G. (1977). Sampling Techniques (3rd edition), New York: John Wiley.
2. Corby, Carol (1984). Content Evaluation of the 1977 Economic Censuses. SRD Research Report, Statistical Research Division,

Bureau of the Census.
3. Hansen, H.H., Hurwitz, W.N., and Madow, W.G. (1953). Sampling Survey Methods and Theory (Volumes 1 and 2), New York: John Wiley.
4. Hartley, H.O., Rao, J.N.K., and Kiefer, G. (1969). "Variance Estimation with One Unit Per Stratum," Journal of the American Statistical Association, 64, 841-851.
5. Isaki, Cary T. (1983a). "Variance Estimation Using Auxiliary Information," Journal of the American Statistical Association, 78, 117-123.
6. Isaki, Cary T. (1983b). "Sample Design for Bulk Oil (1982 Economic Censuses Content Evaluation Survey)" Internal Memorandum, Bureau of the Census.

Appendix A

Isaki's Variance Estimation of a Ratio Estimator under a single stage One Unit Per Stratum PPS sample Design

Let $\hat{Y}_R = (\hat{Y}/\hat{X})\, X$ be the ratio estimator of the total Y, where $\hat{X} = \sum_{i=1}^{n} P_i^{-1} x_i$ and $\hat{Y} = \sum_{i=1}^{n} P_i^{-1} y_i$ are the unbiased estimated totals of X and Y from a one unit per stratum PPS sample design; and X is the known total. y_i and x_i are the selected unit total from i-th stratum. P_i is the unit probability selection. Let $\hat{R}' = \hat{Y}/\hat{X}$ be the estimate of the ratio $R', (Y/X)$.

Define $u_i = y_i - R' x_i$,

$$\hat{u}_i = y_i - \hat{R}' x_i \ ,$$

$$\hat{U} = \sum_{i=1}^{n} P_i^{-1} u_i .$$

We have

$$V(\hat{Y}_R) \doteq V(\hat{Y} - R'\hat{X}) = V(\hat{U}) .$$

Let auxiliary variable z_i has a linear relationship with $\hat{u}_i$ as defined in Isaki (1983). Then analygous to Isaki (1983), the three Isaki variance estimators of $V(\hat{Y}_R)$, $\hat{V}_G(\hat{Y}_R)$, $\hat{V}_{GR1}(\hat{Y}_R)$ and $\hat{V}_{GR2}(\hat{Y}_R)$, are given in (2.3), (2.4) and (2.5) in this paper with $\hat{u}_i$ replacing y_i. More explicitly, it can be shown that the regression coefficients of $\hat{u}_i$ with z_i can be expressed in terms of y_i and x_i with z_i, e.g. $\hat{B}_{w,\hat{u}z} = \hat{B}_{w,yz} - \hat{R}'\, \hat{B}_{w,xz}$. We have

$$\hat{V}_G(\hat{Y}_R) = \hat{V}_G(\hat{Y}) + \hat{R}'^2\, \hat{V}_G(\hat{X}) - 2\hat{R}'\, \hat{Cov}_G(\hat{Y},\hat{X})$$

$$\hat{V}_{GR1}(\hat{Y}_R) = \hat{V}_{GR1}(\hat{Y}) + \hat{R}'^2\, \hat{V}_{GR1}(\hat{X}) - 2\hat{R}'\, \hat{Cov}_{GR1}(\hat{Y},\hat{X})$$

$$\hat{V}_{GR2}(\hat{Y}_R) = \hat{V}_{GR2}(\hat{Y}) + \hat{R}'^2\, \hat{V}_{GR2}(\hat{X}) - 2\hat{R}'\, \hat{Cov}_{GR1}(\hat{Y},\hat{X})$$

where $\hat{V}_G$, $\hat{V}_{GR1}$, $\hat{V}_{GR2}$ are defined in (2.3), (2.4) and (2.5) respectively, and

$$\hat{Cov}_G(\hat{Y},\hat{X}) = \hat{Cov}_{CS}(\hat{Y},\hat{X}) + \hat{B}_{w,yz}\, \hat{B}_{w,xz}\, [V(\hat{Z}) - \hat{V}_{CS}(\hat{Z})]$$

$$\hat{Cov}_{GR1}(\hat{Y},\hat{X}) = \hat{Cov}_{CS}(\hat{Y},\hat{X}) + \hat{B}_{0,yz}\, \hat{B}_{0,xz}[V(\hat{Z}_0) - \hat{V}_{CS}(\hat{Z}_0)] + \hat{B}_{1,yz}\hat{B}_{1,xz}[V(\hat{Z}) - \hat{V}_{CS}(\hat{Z})]$$

$$\hat{Cov}_{GR2}(\hat{Y},\hat{X}) = \hat{Cov}_{GR1}(\hat{X},\hat{Y}) + (\hat{B}_{0,xz}\, \hat{B}_{1,yz} + \hat{B}_{0,yz}\, \hat{B}_{1,xz}) \{Cov(\hat{Z},\hat{Z}_0) - \hat{Cov}_{CS}(\hat{Z},\hat{Z}_0)\} .$$

For estimating $V(\hat{R}')$, since $V(\hat{R}') = X^{-2} V(\hat{Y}_R)$, we have $\hat{V}_G(\hat{R}') = \hat{X}^{-2}\, \hat{V}_G(\hat{Y}_R)$, and similar results for $\hat{V}_{GR1}(\hat{R}')$ and $\hat{V}_{GR2}(\hat{R}')$.

[1]This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributed to the authors and do not necessarily reflect those of the Census Bureau.

TABLE I

The Estimated Relative Bias and MSE of Six Variance Estimators of $V(\bar{Y})$

	$V(\bar{Y})$ x10^5	VCS	VCSA	VG	VGR1	VGR2	VHRK
5171 Receipts ($$10^{12}$)	9.4739						
Relative bias (in %)[a]		70.93	-7.78	-0.73	-2.04	-3.34	2.28
Relative MSE[b]		1.216	0.640	0.547	0.557	0.546	0.577
Ratio of MSE		1	0.277	0.202	0.210	0.202	0.225
5171 Payroll ($$10^{12}$)	0.0117						
Relative bias		145.17	-1.73	52.44	33.47	26.28	6.06
Relative MSE		1.802	0.513	0.786	0.691	0.605	0.758
Ratio of MSE		1	0.081	0.190	0.147	0.113	0.177
5171 # Employment	111.1184						
Relative bias		102.96	0.84	0.10	11.43	14.89	4.18
Relative MSE		1.551	0.614	0.533	0.632	0.593	0.628
Ratio of MSE		1	0.157	0.118	0.166	0.146	0.164
5172 Receipts ($$10^{12}$)	23.3953						
Relative bias		23.73	-3.85	10.73	8.13	5.77	8.35
Relative MSE		1.376	0.986	1.109	1.073	1.043	1.217
Ratio of MSE		1	0.513	0.650	0.608	0.575	0.782
5172 Payroll ($$10^{12}$)	0.0085						
Relative bias		16.77	4.89	12.17	15.22	19.65	1.50
Relative MSE		1.100	0.950	0.641	0.654	0.674	0.884
Ratio of MSE		1	0.746	0.340	0.353	0.375	0.646
5172 # Employment	66.4388						
Relative bias		3.60	4.36	0.66	-4.92	-3.69	-0.35
Relative MSE		0.748	0.635	0.429	0.421	0.396	0.581
Ratio of MSE		1	0.721	0.329	0.317	0.280	0.603

TABLE II

The Estimated Relative Bias and MSE of Six Variance Estimators of $V(\hat{R})$

	$V(\hat{R})$x10^5	VCS	VCSA	VG	VGR1	VGR2	VHRK
SIC 5171 Receipts							
1. $\rho = 0.9998$	0.0289						
Relative bias (%)[a]		13.6	14.7	958.8	3334.4	3268.9	8.1
Relative MSE[b]		0.581	0.639	86.778	214.384	220.381	0.491
2. $\rho = 0.9769$	2.8809						
Relative bias (%)		13.6	14.8	24.6	44.2	46.8	8.2
Relative MSE		0.581	0.638	1.071	2.344	2.492	0.492
3. $\rho = 0.9671$	4.1486						
Relative bias (%)		13.6	14.8	21.2	33.3	35.0	8.2
Relative MSE		0.581	0.638	0.880	1.720	1.721	0.492
4. $\rho = 0.9558$	5.6467						
Relative bias (%)		13.6	14.7	18.9	25.3	27.9	8.2
Relative MSE		0.581	0.638	0.751	1.315	1.346	0.492
5. $\rho = 0.9497$	6.4821						
Relative bias (%)		13.6	14.7	18.1	22.8	26.0	8.2
Relative MSE		0.581	0.637	0.719	1.150	1.229	0.492
SIC 5172 Receipts							
1. $\rho = 0.9877$	7.2908						
Relative bias (%)		27.7	27.5	36.2	173.7	146.9	24.4
Relative MSE		1.013	1.001	1.527	8.363	7.291	0.966
2. $\rho = 0.9718$	17.0009						
Relative bias (%)		30.1	30.9	37.8	93.1	82.7	24.9
Relative MSE		0.963	0.975	1.184	3.665	3.251	0.894
3. $\rho = 0.9599$	24.3553						
Relative bias (%)		30.4	31.4	37.0	74.2	67.5	25.0
Relative MSE		0.979	0.995	1.116	2.645	2.382	0.909
4. $\rho = 0.9503$	30.6248						
Relative bias (%)		30.5	31.5	36.5	66.2	59.8	25.1
Relative MSE		0.992	1.010	1.094	2.182	1.985	0.922

[a] Relative bias in percentages = (Bias $(\hat{V})/V(\bar{Y})$) x 100.

[b] Relative MSE = MSE $(\hat{V})^{1/2}/V(\bar{Y})$.