A COMPARISON OF ESTIMATORS FOR THE VARIANCE OF REGRESSION-TYPE ESTIMATORS IN A FINITE POPULATION

Sandra A. West, Bureau of Labor Statistics

1. INTRODUCTION

This paper is an extension of an earlier paper West (1983), which looked at ratio-type estimators for the total of a finite population. Now estimators are developed for the variance of the error which results from specific estimators of the total. The results of a theoretical and empircial investigation are presented.

The investigations which led to this paper and the earlier paper began in connection with a modernization project for the Bureau of Labor Statistics program that provides monthly estimates of employment, hours and earnings of workers on nonagricultural establishment payrolls. In this program, population employment counts are obtained once a year from Unemployment Insurance administrative records. Monthly estimates of change between the population counts are obtained from a large voluntary monthly mail survey, known as the Current Employment Survey (CES) or 790 Survey because of its schedule number. The CES data are obtained from cooperating establishments on a voluntary mail "shuttle" schedule. The main variable is employment and that is the only one considered in this paper.

The estimators that were considered for the total (West, 1983) were developed from the point of view of probability models. The variance estimators in this paper are also based on probability models. Recent theoretical and empirical studies, such as Royall and Cumberland (1981), have shown the benefits of probability models in finite population inference. These studies show the value of approaches in which models describe the relationships among variables of interest, and inferences are guided by these relationships. The sampling plan is thus relieved of the burden of generating the probability distribution on which inferences are based, and its purpose is seen to be the selection of a good sample. It was shown in West (1981) that the CES data do indeed follow a linear model.

In Section 2 the most promising models and the resulting estimators for the total are reviewed. In Section 3, five variance estimators are developed for a number of different total estimators. Using a real population, an empirical investigation was undertaken to examine the different estimators for the population total at each month and the different estimators of their variance. The empirical investigation is described in Section 4 and the results are presented in Section 5.

2. REVIEW OF ESTIMATORS FOR TOTAL

2.1 Notation and Definitions.

Let $Y_k(i)$ be a random variable denoting the all employment for establishment i at month k, for k = 0, 1, ... k = 0 denotes the benchmark month; that is, the values of $y_0(i)$ are known for all i in the population. Note that $y_k(i)$ denotes the realized value of $Y_k(i)$.

Let N denote the number of establishments in the population under investigation. In this paper it is assumed that the number of establishments in the population is fixed from month to month. Births, deaths, as well as splits and mergers are ignored.

Let P denote the set of establishments in the population; S_k denote the set of establishments in the sample and R_k the set of establishments not in the sample in month k. Let n_k denote the size of set S_k.

A sample is chosen initially and except for non-response that sample is fixed overtime. That is, if there were no non-response

$S_0 = S_1 = \dots S_k = \dots$ and $n_0 = n_1 = \dots n_k = \dots .$

$$\text{Let } S_{k-1}S_k = S_{k-1} \cap S_k$$

That is, $S_{k-1}S_k$ is the set of establishments that responded in both the (k-1) and k months, for k=1,2,... . Let $Y_k(A)$ denote the total at month k for set A; so that the sample total for month k is

$$Y_k(S_k) = \sum_{i \in S_k} Y_k(i) = \sum_{i=1}^{n_k} Y_k(i).$$

Thus $y_k(P) = \sum_{i=1}^{N} Y_k(i)$ is for k=0 the benchmark value and for k = 1,2 ... is the quantity that is being estimated.

2.2 Link Relative and Regression Estimators.

The link relative estimator used in the 790 Survey, is one which uses a benchmark obtained periodically, together with a survey estimate of change for time periods between benchmarks. The estimator for total employment for the first month, denoted by $y_1(\hat{P})$, is

$$\hat{y}_1(P) = Y_0(P)\, Y_1(S_0S_1)/Y_0(S_0S_1)$$

$$\hat{Y}_k(P) = \hat{Y}_{k-1}(P) Y_k(S_{k-1}S_k) / Y_{k-1}(S_{k-1}S_k) \quad (2.2.1)$$

for k = 1,2 ... .

In the CES program a "bias adjustment factor" is applied to the estimator in (2.2.1).

Consider the simple model that traditionally underlies the ratio estimator. That is,

$$E(Y_k(i) \mid \underline{Y}_{k-1} = \underline{y}_{k-1}) = \beta_k y_{k-1}(i). \quad (2.2.2)$$

$$\mathrm{Cov}\,(Y_k(i), Y_k(j) \mid \underline{Y}_{k-1} = \underline{y}_{k-1}) = \begin{cases} \sigma^2 Y_{k-1}(i) & i=j \\ 0 & i \neq j \end{cases}$$

Under this model, the link relative is just the weighted least squares estimator of β. A number of different models were tried on the employment data and this simple model was the most promising.

The problem of estimating the population total can be restated in the following way. The population total can be looked at as the sum of the sampled elements plus the sum of the non-sampled elements. Thus, to estimate the population total at month k, it will only be necessary to estimate the total for non-sampled elements and add that to the known total for sampled elements. That is,

$$\hat{y}_k(P) = y_k(S_k) + \hat{y}_k(R_k). \qquad (2.2.3)$$

Assuming the model in (2.2.2)

$$\hat{y}_k(P) = y_k(S_k) + \frac{y_k(S_{k-1}S_k)}{y_{k-1}(S_{k-1}S_k)} \hat{y}_{k-1}(R_k)$$

$$k = 2, 3 \ldots \qquad (2.2.4)$$

$$\hat{y}_1(P) = y_1(S_1) + \frac{y_1(S_0S_1)}{y_0(S_0S_1)} [y_0(P) - y_0(S_1)].$$

Note that looking at the problem in the manner of (2.2.3) the resulting estimator is not quite the same as the link relative estimator, unless there is no non-response. However (2.2.4) has the attractive feature that it estimates $y_k(S_k)$ by its known value. It is easily shown, see Royall (1981), that both estimators are unbiased under the stated model. However, (2.2.4) is the best linear unbiased estimator under the model.

Another estimator that was found to be quite good was an extension of the link relative estimator that used data from establishments that responded in one month, but not both months. First, write the estimator of total employment as the sum of the three terms

$$\hat{y}_k(P) = y_k(S_k) + \hat{y}_k(S_{k-1}R_k) + \hat{y}_k(R_{k-1}R_k). \qquad (2.2.5)$$

There are a number of ways to estimate the last two terms; the one found to be the most promising was:

$$\hat{y}_k(S_{k-1}R_k) = \hat{y}_{k-1}(S_{k-1}R_k)\; y_k(S_{k-1}S_k)/y_{k-1}(S_{k-1}S_k)$$

$$\hat{y}_k(R_{k-1}R_k) = \hat{y}_{k-2}(R_{k-1}R_k)\; y_k(S_{k-2}S_k)/y_{k-2}(S_{k-2}S_k).$$

Thus

$$\hat{y}_k(P) = y_k(S_k) + \frac{y_k(S_{k-1}S_k)}{y_{k-1}(S_{k-1}S_k)} \hat{y}_{k-1}(S_{k-1}R_k) + \frac{y_k(S_{k-2}S_k)}{y_{k-2}(S_{k-2}S_k)} \hat{y}_{k-2}(R_{k-1}R_k)$$

for $k \geq 2$.

$$\hat{y}_1(P) = y_1(S_1) + \frac{y_1(S_0S_1)}{y_0(S_0S_1)} \cdot$$

$$\hat{y}_0(S_0R_1) + [y_1(S_0S_1)/y_0(S_0S_1)]\, \hat{y}_0(R_0R_1).$$

This estimator is the same as the link relative estimator if there is no non-response.

3. DEVELOPMENT OF THE VARIANCE ESTIMATORS

For an estimator $\hat{y}_k(P)$ of total employment $y_k(P)$, let D_k denote the error, $\hat{y}_k(P) - y_k(P)$. Interest is in the variance of D_k. Conditioning on $y_{k-1}(P)$, the variance can be written as the sum of two terms:

$$V(D_k) = V_{k-1}\, E(D_k | y_{k-1}(P)) + E_{k-1} V(D_k | y_{k-1}(P)]. \qquad (3.1)$$

The variance of D_k will be considered for four estimators of total.

First, let $\hat{y}_k(P)$ be the weighted regression estimator defined in (2.2.4) which will be denoted by $\hat{y}_{kROW}$. It follows that

$$D_k = \beta_k \hat{y}_{k-1}(P) - \beta_k\, y_{k-1}(P) + \varepsilon_k$$

and

$$E(D_k) = \beta_k(\hat{y}_{k-1}(P) - y_{k-1}(P)).$$

Thus the first term in (3.1) becomes

$$V_{k-1}\, E(D_k | y_{k-1}(P)) = \beta_k^2\, V_{k-1}$$

and an estimator of (3.1) can be written:

$$\hat{V}_k = \hat{\beta}_k^2\, \hat{V}_{k-1} + \hat{V}\,(D_k | y_{k-1}(P)) \qquad (3.2)$$

where

$\hat{\beta}_k = y_k(S_{k-1}S_k)/y_{k-1}(S_{k-1}S_k)$ and $V_k = V(D_k)$ for $k = 1, 2, 3, \ldots$

$$\hat{V}_o = V_o = 0.$$

Before developing estimators for the conditional variance in (3.2) recall:

$$y_k(P) = \hat{y}_k(S_k) + y_k(R_k) = \hat{y}_k(S_k) + \beta_k \hat{y}_{k-1}(\hat{R}_k)$$

$$y_k(P) = y_k(S_k) + y_k(R_k)$$

thus,

$$\hat{y}_k(P) - y_k(P) = \frac{y_k(S_{k-1}S_k)}{y_{k-1}(S_{k-1}S_k)} \hat{y}_{k-1}(R_k) - y_k(R_k).$$

The conditional variance can now be written

$$V\{D_k | k-1\} = \frac{\hat{y}_{k-1}^2(R_k)}{y_{k-1}^2(S_{k-1}S_k)} V(y_k(S_{k-1}S_k) | k-1) + V(y_k(R_k) | k-1).$$

Noting that

$$V(y_k(R_k) | k-1) = \sigma^2 y_{k-1}(R_k)$$

$$V(y_k(S_{k-1}S_k) | k-1) = \sigma^2 y_{k-1}(S_{k-1}S_k)$$

then

$$V(y_k(R_k) | k-1) \doteq \frac{\hat{y}_{k-1}(R_k)}{y_{k-1}(S_{k-1}S_k)} H_k$$

where

$$H_k = V(y_k(S_{k-1}S_k) | k-1).$$

Therefore

$$V\{D_k | k-1\} \doteq C_k^*\,(C_k^* - 1) \cdot H_k$$

where

$$C_k^* - 1 = \hat{y}_{k-1}(R_k)/y_{k-1}(S_{k-1}S_k)$$

Thus (3.2) can now be written

$$\hat{V}_k = \hat{\beta}_k^2\, \hat{V}_{k-1} + C_k^*\,(C_k^* - 1) H_k. \qquad (3.3)$$

Five different estimators will be considered for H_k. First,

$$H_k = V\left(\sum_{i \varepsilon S_{k-1}S_k} y_k(i) \middle| k-1\right) = \sum_{i \varepsilon S_{k-1}S_k} V\left(y_k(i) \middle| k-1\right) = \sigma^2 Y_{k-1}(S_{k-1}S_k) \quad (3.4)$$

Using the estimator of σ^2 from standard weighted least squares theory yields for H_k:

$$\hat{H}_k^{(1)} = \frac{Y_{k-1}(S_{k-1}S_k)}{n' - 1} \sum_{i \varepsilon S_{k-1}S_k} (\mathrm{RES}(i))^2 / Y_{k-1}(i)$$

where $\mathrm{RES}(i) = y_k(i) - \hat{\beta}_k\, Y_{k-1}(i)$, $n' = \sum_{i \varepsilon S_{k-1}S_k} i$

The corresponding $\hat{V}_k$ is denoted by $\hat{V}_k^{(1)}$. This estimator is unbiased under the model, but it can be badly biased if the model fails in that the variance of y_k is not proportional to y_{k-1}. Two alternatives suggested by Royall and Cumberland (1981), are either to make the individual estimators in the sum (3.4) unbiased or make the entire sum unbiased. First, let c_i be the constant to make the i th term unbiased, then

$$\hat{H}_k^{(2)} = \sum_{i \varepsilon S_{k-1}S_k} c_i [\mathrm{RES}(i)]^2 \quad (3.5)$$

where $c_i = \{1 - Y_{k-1}(i)/Y_{k-1}(S_{k-1}S_k)\}^{-1}$.

Next, let c be the constant to make the entire sum unbiased, so that

$$\hat{H}_k^{(3)} = c \sum_{i \varepsilon S_{k-1}S_k} (\mathrm{RES}(i))^2 \quad (3.6)$$

where $c = \{1 - \sum_{i \varepsilon S_{k-1}S_k} Y_{k-1}^2(i)/(Y_{k-1}(S_{k-1}S_k))^2\}^{-1}$

Both estimators $\hat{V}_k^{(2)}$ and $\hat{V}_k^{(3)}$ have the advantage that they remain approximately unbiased under much less restrictive assumptions. The next estimator is based on the jackknife estimator for the variance of $\hat{\beta}_k$. Note that H_k can be written as

$$H_k = Y_{k-1}^2 (S_{k-1}S_k) \cdot V(\hat{\beta}_k | k-1).$$

Letting

$$\hat{\beta}_{k\ell} = \frac{Y_k(S_{k-1}S_k) - y_k(\ell)}{Y_{k-1}(S_{k-1}S_k) - Y_{k-1}(\ell)}$$

and

$$\hat{\beta}_{k.} = 1 \sum_{\ell=1}^{n'} \hat{\beta}_{k\ell}/n'$$

then the jackknife estimator of the $V(\hat{\beta}_k | k-1)$ is

$$\hat{V}\,\hat{\beta}_k = \frac{n' - 1}{n'} \sum_{\ell=1}^{n'} (\hat{\beta}_{k\ell} - \hat{\beta}_{k.})^2.$$

The estimator of H_k is

$$\hat{H}_k^{(4)} = Y_{k-1}^2(S_{k-1}S_k) \cdot \hat{V}\,\hat{\beta}_k. \quad (3.7)$$

As a fifth estimator, a bootstrap estimator was considered. As pointed out in Efron (1982), the bootstrap gives the standard estimate of σ^2 in the linear regression case, except for a constant. That is,

$$\hat{H}_k^{(5)} = (n' - 1) \quad \hat{H}_k^{(1)}/n' \quad (3.8)$$

If one starts with the link relative estimator defined in (2.2.1), then the only difference in the error variance estimator, which will be denoted $V_{K,LR}$, is the constant multiple of H_k; that is,

$$\hat{V}_{k,LR}^{(s)} = \hat{\beta}_k^2\, \hat{V}_{k-1,LR}^{(s)} + c_k(c_k - 1)\hat{H}_k^{(s)} \quad (3.9)$$

where $c_k = \hat{Y}_{k-1}(P)/Y_{k-1}(S_{k-1}S_k) = c_k^* + Y_{k-1}(S_k)/Y_{k-1}(S_{k-1}S_k)$

for s = 1, 2, 3, 4, 5.

Note that $D_{kROW} = D_{kLR} + E$ where

$E = y_k(S_k) - \hat{\beta}_k y_{k-1}(S_k)$.

Next, the variance will be considered for the modified link relative estimator defined in (2.2.5), which will be denoted by $\hat{Y}_{LC}$. From the definition,

$$D_k = \hat{\beta}_k y_{k-1}(S_{k-1}R_k) - y_k(S_{k-1}R_k)$$
$$\hat{\beta}'_k\, y_{k-2}(R_{k-1}R_k) - y_k(R_{k-1}R_k)$$

where

$$\hat{\beta}_k = \frac{y_k(S_{k-1}S_k)}{y_{k-1}(S_{k-1}S_k)}, \quad \hat{\beta}'_k = \frac{y_k(S_{k-2}S_k)}{y_{k-2}(S_{k-2}S_k)}.$$

The conditional variance, $V(D_k | <k-1)$, denoted by A_k, is

$$A_k = \frac{\hat{y}_{k-1}^2(S_{k-1}R_k)}{y_{k-1}^2(S_{k-1}S_k)} \cdot V(Y_k(S_{k-1}S_k) | k-1) + V(Y_k(R_k) | k-1)$$
$$+ \frac{\hat{y}_{k-2}^2(R_{k-1}R_k)}{y_{k-2}^2(S_{k-2}S_k)} \cdot V(Y_k(S_{k-2}S_k) | k-1)$$
$$+ 2\, \frac{\hat{y}_{k-1}(S_{k-1}R_k)\, \hat{y}_{k-2}(R_{k-1}R_k)}{y_{k-1}(S_{k-1}S_k) \cdot y_{k-2}(S_{k-2}S_k)}$$
$$\cdot \mathrm{COV}(y_k(S_{k-1}S_k), y_k(S_{k-2}S_k) | k-1)$$

This can be rewritten as

$$A_k = \frac{H_k}{y_{k-1}(S_{k-1}S_k)}$$
$$a_k[y_{k-1}(S_{k-1}R_k) + 2\, b_k y_{k-1}(S_{k-2}S_{k-1}S_k)]$$
$$+ b_k^2\, y_{k-1}(S_{k-2}S_k) + y_{k-1}(R_k)$$

where

$$a_k = \frac{y_{k-1}(S_{k-1}R_k)}{y_{k-1}(S_{k-1}S_k)}, \quad b_k = \frac{\hat{y}_{k-2}(R_{k-1}R_k)}{y_{k-2}(S_{k-2}S_k)}.$$

In order to compute the first term in (3.1) note that in this case

$$\underset{\sim}{Y}_{k-1}E_k(\hat{y}_k(P) - y_k(P)) = V\, E_k\{(\hat{\beta}_k - \beta_k)y_{k-1}(S_{k-1}R_k)$$
$$+ \hat{\beta}'_k\, \hat{y}_{k-2}(R_{k-1}R_k) - \beta'_k\, y_{k-2}(R_{k-1}R_k) + \varepsilon\}$$
$$= \underset{\sim}{Y}_{k-1}\beta'_k\, \{\hat{y}_{k-2}(R_{k-1}R_k) - y_{k-2}(R_{k-1}R_k)\}$$
$$= \beta'^2_k\, V_{k-2}\, \{\hat{y}_{k-2}(R_{k-1}R_k) - y_{k-2}(R_{k-1}R_k)\}$$

Letting $V_{k,LC}$ denote $V(D_{kLC})$ then

$$V_{K,LC} = \beta_k'^2 V_{k-2,LCI}(R_{k-1}R_k) + A_k \qquad (3.10)$$

for $k \geq 3$.

Note the following relationship between $\hat{y}_{k,LC}$ and $\hat{y}_{k,ROW}$:

$$D_{kLC} = D_{kROW} + \hat{\beta}_k' \hat{y}_{k-2} (R_{k-1}R_k) - \hat{\beta}_k \hat{y}_{k-1}(R_{k-1}R_k)$$

for $k \geq 2$.

In the comparison of estimators of the total, the Horvitz-Thompson estimator was included in order to give a comparison between the usual probability estimator and the model-based estimators, described in Section 2. The Horvitz-Thompson estimator, HT, for the total at month k of the i th stratum is defined as

$$HT_{ki} = N_i \cdot y_k(S_{ki})/n_{ki} \qquad (3.11)$$

where N_i and n_{ki} are the population size and sample size respectively, of the i th stratum at month k. An estimator of the variance of HT, as given in Cochran (1977), is

$$\hat{V}(HT_{ki}) = \frac{N_i^2(N_i - n_{ki})}{n_{ki}^2 (n_{ki}-1)} \sum_{\ell=1}^{n_{ki}} \sum_{m>\ell}^{n_{ki}} (y_{ki\ell} - y_{kim})^2. \qquad (3.12)$$

This estimator is included in the empirical investigation.

4. EMPIRICAL INVESTIGATION

An empirical investigation was conducted on a data base of real employment data. The data, for the most part, come from the Unemployment Insurance (U.I.) accounting file. The information used to maintain the U.I. file is obtained from quarterly reports which each covered employer is required to submit. These quarterly reports contain, among other things, information on employment for each month of the first quarter. Each U.I. account also carries an industry code. The industry codes are taken from the Standard Industrial Classification (SIC) Manual, 1972 edition, as amended by the 1977 supplement. Also available for the same time period are the data from the 790 Survey (790 SAMP). In principle all the establishments on the 790 data base should also be on the U.I. file.

The purpose of the empirical investigation was to evaluate the current sampling plan and estimation procedures used in the 790 Survey, and compare these with viable alternatives. Estimators for total employment and change of employment were considered as well as estimators for the variance of the error. In this report only one sampling plan is considered and only four estimators for total are reported. For each estimator of total five estimators for the variance of the error are considered. The investigation can best be described in terms of four modules: population, sample selection, estimation, and evaluation.

Population Module.

For a given SIC (in this paper only SIC 177, concrete work, is considered) the '79 U.I. file was matched with the '80 U.I. file and this was matched with the 790 SAMP file. The population is made up of three parts: the establishments that are on both the U.I. and SAMP; those on the U.I. but not SAMP (most), and the establishments on the SAMP but not on the U.I. For those establishments that are on both the SAMP and U.I., the all employment values for March '79, January, February, and March '80 were compared. If they differed the SAMP file values were used if all four values were there, otherwise the U.I. values were used. (Note '79 U.I. file, as held by B.L.S., only has one month of data.) Initially it is assumed that there is no non-response in the population; thus the population contains only those establishments that responded in all three months in '80. In the case of SIC 177, the population size is 8419.

Sample Selection Module.

There are five variables: Sample size, strata bounds, allocation of sample, type of random selection and response rate. Initially, the sample sizes that appear on the actual 790 sample are used, however, the samples are selected randomly using nine strata, 0-3, 4-9, 10-19, 20-49, 50-99, 100-249, 250-499, 500-999, 1000 or more employees. The establishments are classified into strata by their March '79 all employment values. In SIC 177 there is no 1000+ strata and an additional strata was added to take care of establishments that had no '79 values (essentially births). The sample sizes by strata are 31, 36, 57, 75, 96, 50, 14, 5, 2. Up to this point only two response rates have been considered, 100 percent and 80 percent. For the 80 percent response rate, the 20 percent non response was simulated by a uniform random number generator. In this paper, the results from 200 samples are reported.

Estimation Module.

There are two sections to this module; one section computes the estimators of the total and the other section computes the variances. Although ten estimators of the total were evaluated only four are reported here. These are the three estimators reviewed in Section 2 - the link relative in (2.2.1) denoted by LR, the estimator in (2.2.4) denoted by ROW, and the extension of the link relative estimator in (2.2.5), denoted by LC. The fourth estimator is the Horvitz-Thompson estimator described in (3.11). For each of the three estimators in Section 2, the five estimators for the variance of the error described in Section 3, are computed. For the Horvitz-Thompson estimator, the variance estimator in (3.12) is computed.

Evaluation Module.

A number of different evaluation measures were considered for the estimators of the total and change and for the variance estimators. Measures that considered the error in each stratum, as well as the overall error, were used.

Letting $\hat{\theta}(P)$ be an estimator of $\theta(P)$, the absolute error in $\hat{\theta}(P)$, denoted by $AE(\hat{\theta})$, is defined as:

$$AE(\hat{\theta}) = \left|\hat{\theta}(P) - \theta(P)\right|.$$

$\hat{\theta}=\hat{y}_k(P)$, $\hat{y}_k(P) - \hat{y}_{k-1}(P)$, and $\hat{V}\left(\hat{y}_k(P) - y_k(P)\right)$.

Two hundred samples were randomly drawn and the indicators were computed on each sample for each of the estimators. These indicators were averaged over the two hundred samples for the different estimators. In addition the mean, variance, and mean square error were computed over the two hundred samples. In order to save space, only the mean, variance and mean square error are reported for the error variances. For the estimators of total, the average absolute error and the absolute average error, which are defined below, are also reported. Letting ℓ denote the subscript for the sample number, the average absolute error is defined as:

$$AAE(\hat{\theta}) = \sum_{\ell=1}^{200} AE_{\ell}(\hat{\theta})/200$$

and the absolute average error is defined as:

$$ABAE(\hat{\theta}) = \left|\sum_{\ell=1}^{200}(\hat{\theta}_{\ell}/200) - \theta\right|.$$

When $\hat{\theta}$ is the variance estimator θ is taken as

$$\frac{1}{r}\sum_{\ell=1}^{r} \left(\hat{y}_{k\ell}(P) - y_k(P)\right)^2.$$

5. CONCLUSIONS

First consider the estimators for total. For the 100% response rate, there are only two distinct estimators and it is clear that ROW is better than HT. For the 80% response rate, there are three distinct estimators in the first month and four in the second month. From Table 2, it is clear that in terms of level LC is the best estimator followed by ROW, LR and then HT. However, in terms of change, ROW is the best estimator followed by LR, LC and then HT. Overall, it seems that ROW would be the best estimator to use. However, it is noted that LR is not far behind the winner.

Next consider the error variance estimators. Since the ranking of the five estimators were the same for the three estimators of total, only the error variane estimators for ROW are given. Also, only the 100% response rate is shown since there was not much difference in the rankings between the two response rate. The sixth estimator in Table 3 is the jackknife method with the jackknife estimator of β instead of the weighted least squares estimator of β. As can be seen from the table, method 6 is almost identical to method 4. From table 3 it is clear that for the first month the top two estimators are Royall's robust estimators $\hat{V}^{(2)}$ and $\hat{V}^{(3)}$, with $\hat{V}^{(3)}$ at the top. For the second month it is not quite as clear. $\hat{V}^{(5)}$ and $\hat{V}^{(2)}$ have the smallest means, however, $\hat{V}^{(5)}$ and $\hat{V}^{(1)}$ have the smallest variances and $\hat{V}^{(5)}$ and $\hat{V}^{(3)}$ have the smallest MSE. Overall, the robust estimator, $\hat{V}^{(3)}$, seems like the best estimator to use with the employment data. In passing, it is noted that all the variance estimators did well in the investigation of the coverage properties of the related confidence intervals. For overall total, the short (or long) fall was less than .03.
The Horvitz-Thompson Estimator and its variance estimator finished far behind the other estimators.

REFERENCES

1) Cochran, William (1963), "Sampling Techniques," Third Edition, John Wiley and Sons, New York.

2) Madow, Lillian H., and Madow, William G. (1978), "On Link Relative Estimators", ASA Proceedings of the Section on Survey Research Methods, 534-9.

3) Royall, R.M. and Cumberland, W.G. (1978), "Variance Estimation in Finite Population Sampling", Journal of the American Statistical Association, 73, 351-8.

4) Royall, Richard M. and Cumberland, William G. (1981), "An Empirical Study of the Ratio Estimator and Estimators of its Variance", Journal of the American Statistical Association, 76, 66-77.

5) Royall, Richard M. (1981), "Study of Role of Probability Models in 790 Survey Design and Estimation," Bureau of Labor Statistics contract Report 80-98.

6) West, Sandra A. (1981), "Linear Models for Monthly All Employment Data," Bureau of Labor Statistics report.

7) West, Sandra A. (1983), "A Comparison of Different Ratio and Regression Type Estimators for the Total of a Finite Population," ASA Proceedings of the Section on Survey Research Methods.

ACKNOWLEDGEMENTS

Thanks are due to Troy Bishop for programming the system that computes the empirical investigation and to Linda Weaver for typing this paper.

TABLE 1. ESTIMATORS FOR TOTAL
Mean, Variance, ABAE, MSE, and AAE over 200 Samples
(100% Response Rate)

Estimator $\hat{Y}$	Mean Y	Variance S_Y	Abs.Avg.Error ABAE($\hat{Y}$)	Mean Sq.Error MSE($\hat{Y}$)	Avg.Abs.Error AAE($\hat{Y}$)
	(10^4)	(10^6)		(10^6)	(10^3)
			Month 1		
ROW*	7.35	2.28	4.19	2.27	1.21
HT	7.37	13.81	203.92	13.78	2.88
			Month 2		
ROW	7.62	3.86	59.86	3.84	1.58
HT	7.64	16.13	196.59	16.09	2.95

*ROW = LR = LC

TABLE 2. ESTIMATORS FOR TOTAL
Mean, Variance, ABAE, MSE, and AAE over 200 Samples
(80% Response Rate)

Estimator $\hat{Y}$	Mean Y	Variance S_Y	Abs.Avg.Error ABAE($\hat{Y}$)	Mean Sq.Error MSE($\hat{Y}$)	Avg.Abs.Error AAE($\hat{Y}$)
	(10^4)	(10^6)	(10^2)	(10^6)	(10^3)
			Month 1		
ROW	7.366	3.714	1.430	3.716	1.527
LR	7.367	3.839	1.539	3.844	1.552
LC	7.366	3.714	1.430	3.716	1.527
HT	7.380	18.299	2.822	18.287	3.264
			Month 2		
ROW	7.657	6.035	3.921	6.158	1.998
LR	7.659	6.370	4.147	6.510	2.064
LC	7.643	5.563	2.571	5.601	1.870
HT	7.604	16.554	1.315	16.489	3.032

	AAE $(\hat{Y}_2 - \hat{Y}_1)$ (10^3)	
	100% Response	80% Response
ROW	1.04	1.260
LR	-	1.287
LC	-	1.378
HT	1.08	2.095

TABLE 3. VARIANCE ESTIMATORS FOR ROW ERROR
Mean, Variance and Mean Square Error over 200 Samples
(100% Response Rate)

Estimator $\hat{V}^{(s)}$	Mean $\bar{V}^{(s)}$	Variance $SV^{(s)}$	Mean Sq.Error MSE($\hat{V}^{(s)}$)
	(10^6)	(10^{11})	(10^{11})
		Month 1	
1	2.390	9.070	9.166
2	1.953	5.979	6.964
3	1.927	5.172	6.328
4	2.026	7.607	8.171
5	2.333	8.617	8.612
6	2.026	7.607	8.171
		Month 2	
1	4.474	13.420	17.323
2	3.847	18.291	18.200
3	3.798	16.469	16.408
4	4.002	22.636	22.775
5	4.375	12.841	15.602
6	4.002	22.638	22.777