A Comparison of Different Ratio and Regression Type Estimators for the Total of a Finite Population

Sandra A. West
Bureau of Labor Statistics

1. Introduction

In this paper the results of a theoretical and empirical investigation of different regression and ratio type estimators are presented. The investigation began in connection with a revision program 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, benchmark employment is obtained every year or so from Unemployment Insurance administrative records. Monthly estimates of change between benchmarks are obtained from a large voluntary monthly mail survey, known as the 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 all employment, as called in the CES proogram, and that is the only one considered in this paper. In this study important complications (such as "births" of new establishments) are deliberately ignored.

Most of the estimators discussed in this paper will be considered from the point of view of 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 (1982) that the CES data do indeed follow a linear model. A number of linear models will be considered in Section 2.

The current estimator for all employment is a link relative estimator, which is essentially a ratio type estimator. This estimator, along with a number of related competitors will be discussed in Section 2. Among the regression estimators considered is one derived by applying a least squares approach to a linear model with two independent variables with data missing from each variable but not both simultaneously.

Using a real population, consisting of several months of data, an empirical investigation was undertaken to examine the different estimators for the population total at each month. In this paper ten estimators are compared. The empirical investigation is described in Section 3 and the results are presented in Section 4.

2. Estimation

2.1 Notation and definitions.

Let $Y_k(i)$ be a random variable denoting the all employment for establishment i at month k, for $k = 0, 1, \ldots$. $k = 0$ 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
$S_0 = S_1 = .. S_k = \ldots$ and $n_0 = n_1 = \ldots n_k = \ldots$.

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

$$y_k(S_k) = \sum_{i \varepsilon S_k} y_k(i) = \sum_{i=1}^{n_k} y_k(i) .$$

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, which is essentially what is 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 (P), is

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

and in general,

$$\hat{y}_k(P) = \hat{y}_{k-1}(P) \cdot \frac{y_k(S_{k-1}S_k)}{y_{k-1}(S_{k-1}S_k)} \qquad (2.2.1)$$

for k = 1,2 ... .

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

In Madow and Madow (1978) the link relative estimator is discussed from the point of view of a simple statistical model. The type of models that were found the most promising were proportional regression models specifying that the expected all employment for establishment i for the kth month, $Y_k(i)$, given the set Y_{k-1} of y values for month k-1, is proportional to $Y_{k-1}(i)$, its previous months employment.

That is, $E(Y_k(i) \mid \underline{Y}_{k-1} = \underline{y}_{k-1}) = \beta y_{k-1}(i)$.

It is further assumed that the Y's are conditionally uncorrelated

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

where $v_k(i)$ represents the conditional variance of $Y_k(i)$, which in general will depend on $Y_{k-1}(i)$. Choosing a specific simple function to represent the variance $v_k(i)$ accurately is difficult. Fortunately, knowledge of the precise form of $v_k(i)$ is not essential. An estimator of the population total which is efficient for a given class of v's does not become very inefficient when the true v's differ somewhat from those given, see Royall and Cumberland (1978). In particular, in Madow and Madow (1978) it was assumed that

$$v_k(i) = \sigma^2 y_{k-1}(i).$$

Rewriting the model as,

$$y_k(i) = \beta y_{k-1}(i) + e(i) \tag{2.2.2}$$

and using a weighted least squares approach with weight equal $1/y_{k-1}(i)$ yields the following estimator for β

$$\hat{\beta} = \frac{\sum_{i \in S_{k-1} S_k} y_k(i)}{\sum_{i \in S_{k-1} S_k} y_{k-1}(i)} = \frac{y_k(S_{k-1}S_k)}{y_{k-1}(S_{k-1}S_k)} \tag{2.2.3}$$

which is the link relative given in (2.2.1), and actually used in the computation of the 790 estimate.

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,

$$\hat{y}_k(P) = \sum_{i \in S_k} y_k(i) + \sum_{i \notin S_k} \hat{y}_k(i)$$

$$= y_k(S_k) + \hat{y}_k(R_k) \tag{2.2.4}$$

Assuming the model in (2.2.2) and $V_k(i) = \sigma^2 y_{k-1}(i)$,

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

$$k = 2, 3 \ldots \tag{2.2.5}$$

$$\hat{y}_1(P) = y_1(S_1) + \frac{y_1(S_0S_1)}{y_0(S_0S_1)} \cdot [y_0(P) - Y_0(S_1)].$$

Note that looking at the problem in the manner of (2.2.4) the resulting estimator (2.2.5) is not quite the same (unless there is no non-response) as the link relative estimator in (2.2.1). However (2.2.5) 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.

There are many modifications to this simple model. Using real data, it is shown in West (1982) that the simple linear model assumed by Madow is not an unreasonable fit, but a model with a non-zero intercept may be better. Also it seems reasonable to take $v_k(i) = \sigma^2$. Thus, consider the model

$$y_k(i) = \alpha + \beta y_{k-1}(i) + e_k(i)$$

where $E(e_k(i)) = 0$ and $V(e_k(i)) = \sigma^2$.

Using least squares one arrives at the estimator

$$\hat{y}_k(P) = y_k(S_k) + (N-n_k)\, y_k(S_{k-1}S_k) / n_{k-1,k} + \hat{\beta}\left[\hat{y}_{k-1}(P) - y_{k-1}(S_k) - (N-n_k) \cdot y_k(S_{k-1}S_k)\, n_{k-1,k}\right] \tag{2.2.6}$$

where

$$\hat{\beta} = n_{k-1,k} \sum_{i \in S_{k-1} S_k} y_{k-1}(i)\, y_k(i) - y_{k-1}(S_{k-1}S_k) \cdot y_k(S_{k-1}S_k) \Big/ \left[n_{k-1,k} \sum_{i \in S_{k-1} S_k} (y_{k-1}(i))^2 - y_{k-1}(S_{k-1}S_k)^2\right]$$

and n_k is the number of elements in S_k, and $n_{k-1,k}$ is the number of elements in $S_{k-1}S_k$, for $k = 1,2,3, \ldots$. Note for k=1, $\hat{y}_0(P) = Y_0(P)$, the benchmark value.

An unweighted regression through the origin ($\alpha = 0$) leads to the following estimator:

$$y_k(P) = y_k(S_k) + \hat{\beta}[y_{k-1}(P) - y_{k-1}(S_k)]$$

where $\hat{\beta} = \sum_{i \in S_{k-1} S_k} y_{k-1}(i)\, y_k(i) \Big/ \sum_{i \in S_{k-1} S_k} (y_{k-1}(i))^2.$ (2.2.7)

A natural extension of the previous regression models is to include data from two previous months; this is discussed in 2.3. In the rest of this subsection a natural extension of the link relative estimator is considered.

Recall that the link relative estimator only uses data from establishments that are in both months, the current and previous. It is not too difficult to write down similar estimators which are not as wasteful of data. First, write the estimator of total employment as the sum of three terms

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

That is, the estimate of total employment for month k is the sum of the estimate of the total for units in the current sample S_k, plus the estimate of the total for units in the previous sample but not in the current sample, $S_{k-1}R_k$, plus the estimate of the total for units not in the sample for either month, $R_{k-1}R_k$. The first term is just the sample sum = $y_k(S_k)$. Note that

$$\frac{y_k(S_{k-1}R_k)}{y_{k-1}(S_{k-1}R_k)} \approx \frac{y_k(S_{k-1}S_k)}{y_{k-1}(S_{k-1}S_k)} \quad \text{therefore}$$

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

Similarly one could take

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

thus

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

for $k > 2$, (2.2.8)

$$\hat{y}_1(P) = y_1(S_1) + \frac{y_1(S_0S_1)}{y_0(S_0S_1)} \cdot \hat{y}_0(S_0R_1) + y_1(S_0S_1)/y_0(S_0S_1)\ \hat{y}_0(R_0R_1).$$

An alternative for $\hat{y}_k(R_{k-1}R_k)$ is

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

Note that if the same units are always in the sample then the link relative estimator is the same as the ones in (2.2.8).

2.3 Regression model with two independent variables

A natural extension of the previous regression models is to include data from a further month. Consider the following model

$$y_k(i) = \beta_0 + \beta_1 y_{k-1}(i) + \beta_2 y_{k-2}(i) + e_k(i)$$

where $E(e_k(i)) = 0$ and $V(e_k(i)) = \sigma^2$.

Using $\hat{y}_k(P) = y_k(S_k) + \hat{y}_k(R_k)$ and

$$\hat{y}_k(R_k) = \sum_{i \notin S_k} \hat{\beta}_0 + \hat{\beta}_1 \sum_{i \notin S_k} y_{k-1}(i) + \hat{\beta}_2 \sum_{i \notin S_k} y_{k-2}(i)$$

$$= (N-n_k)\ \hat{\beta}_0 + \hat{\beta}_1\ \hat{y}_{k-1}(R_k) + \hat{\beta}_2\ \hat{y}_{k-2}(R_k)$$

it remains to obtain the regression coefficients.

The method derived for estimating the parameters will not be presented in this paper; just the results will be given. The method assumes that the independent variables are fixed numbers and that each observation contains the values of the dependent variable and at least one of the independent variables. The two main advantages of this method are: the resulting estimators are consistent and the asymptotic variances of these estimators are smaller than those of comparable estimators described in the literature. The regression coefficeints will now be described.

The only set of triplets that will be used in estimating the parameters are the ones that always have a y_k value and values of y_{k-1} and y_{k-2} are never missing simultaneously. The following notation will be used.

$$S_1 = S_{k-1}S_k$$
$$S_2 = S_{k-2}S_k$$
$$S_c = S_{k-1}S_{k-2}S_k$$
$$S_a = S_k S_{k-2} R_{k-1}$$
$$S_b = S_k S_{k-1} R_{k-2}$$
$$S_T = S_a + S_b + S_c$$

and n_j = the number of elements in s_j, for j = 1,2,a,b,c, and $n_s = n_a + n_b + n_c$. Let,

$$y_{m,k}(S_j) = \sum_{i \in S_j} (y_{k-m}(i) - \bar{y}_{k-m})\ (y_k(i) - \bar{y}_k)$$

$m = 1,2$

$$y_{1,2}(S_j) = \sum_{i \in S_j} (y_{k-1}(i) - \bar{y}_{k-1})\ (y_{k-2}(i) - \bar{y}_{k-2})$$

$$y_{1,1}(S_j) = \sum_{i \in S_j} (y_{k-1}(i) - \bar{y}_{k-1})^2$$

$$y_{2,2}(S_j) = \sum_{i \in S_j} (y_{k-2}(i) - \bar{y}_{k-2})^2$$

$$\hat{\sigma}_{12} = \frac{y_{1,2}(S_c)}{n_c}\left[1 + \frac{y_{2,2}(S_a)}{y_{2,2}(S_c)}\ \frac{n_c}{n_2} + \frac{y_{1,1}(S_b)}{y_{1,1}(S_c)}\ \frac{n_c}{n_1} - \left(\frac{n_a}{n_2} + \frac{n_b}{n_1}\right)\right]$$

$$\hat{\sigma}_{11} = \frac{y_{1,1}(S_1)}{n_1} + \frac{y_{1,2}^2(S_c)}{y_{2,2}^2(S_c)}\ \frac{1}{n_1 n_2} \cdot \left[n_c\, y_{2,2}(S_a) - n_a\, y_{2,2}(S_c)\right]$$

$$\hat{\sigma}_{22} = \left[\frac{y_{2,2}(S_2)}{n_2} + \frac{y_{1,2}^2(S_c)}{y_{1,1}^2(S_c)}\ \frac{1}{n_1 n_2}\right] \cdot \left[n_c\, y_{1,1}(S_b) - n_b\, y_{1,1}(S_c)\right]$$

$$D = (\hat{\sigma}_{11} \cdot \hat{\sigma}_{22} - \hat{\sigma}_{12}^{2})$$

$$V = 1 - \frac{(n - n_1)(n - n_2)\hat{\sigma}_{12}^{2}}{n_1 \cdot n_2 \cdot \hat{\sigma}_{11}\hat{\sigma}_{22}}$$

Now the regression coefficients can be written as:

$$\hat{\beta}_1 = \frac{1}{DVn_1n_2}\Bigg[\frac{n_2 D y_{1,k}(S_1)}{\hat{\sigma}_{11}} + \frac{n_s \hat{\sigma}_{12}^{2}}{\hat{\sigma}_{11}} \cdot y_{1,k}(S_c) - n_s \hat{\sigma}_{12} y_{2,k}(S_c) - (n_s - n_2) D' \, \hat{\sigma}_{12}\, y_{2k,}(S_a) / \hat{\sigma}_{11}\hat{\sigma}_{22}\Bigg]. \qquad (2.3.1)$$

The formula for $\hat{\beta}_2$ can be be found easily by interchanging 1 and 2 as well as a and b in (2.3.1) then,

$$\hat{\beta}_0 = \frac{y_k(S_T)}{n_s} - \hat{\beta}_1 \frac{y_{k-1}(S_1)}{n_1} - \hat{\beta}_2 \frac{y_{k-2}(S_2)}{n_2} .$$

3. Empirical Investigation

An empirical investigation was conducted on a data base of real employment data. The data, for the most part, comes 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 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 is 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 this empirical investigation is to evaluate the current sampling plan and estimation procedures used in the 790 Survey, and compare these with viable alternatives. In this report only one sampling plan is considered along with the estimators described in Section 2. The investigation can best be described in modular form; there are 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, Febraury, 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 20 samples are reported.

Estimation module

In this paper the behavior of ten estimators are reported. Seven of these estimators are described in Section 2: the link relative estimator in (2.2.1) will be denoted by LR, the estimator in (2.2.5) will be denoted by ROW, three unweighted regression estimators- (2.2.6) denoted by R1, (2.2.7) denoted by R0, (2.3.1) denoted by R2- and the two extensions of the link relative estimator suggested in (2.2.8) which will be denoted by LC1 and LC2 respectively. After looking at the population data weighted regressions with weight equal to $1/y_{k-1}$ and $1/y^2_{k-1}$ were tried. These estimators are denoted by RW1 and RW2 respectively. The tenth estimator is the Horvitz Thompson estimator, which 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 j stratum is defined as

$$HT_{kj} = \frac{N_j}{n_{jk}} y_k(S_k^j) \qquad (3.1)$$

where N_j and n_{jk} are the population size and sample size respectively, of the jth stratum at month k.

Evaluation module

Within each sample three measures of evaluation that consider the error in each stratum were used. For each stratum, j, j = 0,1,...8 the population total $Y_k(S^j)$ is known for k = 0,1,2. For all the estimators, with the exception of the Horvitz Thompson, k=0 is taken as the benchmark month (January '80).

The main interest is in the estimate of total employment for the entire industry; that is, letting

$$\hat{Y}_k(P) = \sum_{j=0}^{8} \hat{Y}_k(P^j). \text{ and } Y_k(P) = \sum_{j=0}^{8} Y_k(P^j),$$

the absolute error in the total is

$$AE(\hat{Y}_k(P)) = \left| \hat{Y}_k(P) - Y_k(P) \right| . \qquad (3.5)$$

In addition to the level of employment there is also much interest in the change of employment from one month to the next. As a measure of an estimators performance in estimating the change AE $(\hat{D})$ is used,

$$\text{where} \quad \hat{D} = \hat{Y}_k(P) - \hat{Y}_{k-1}(P) \qquad (3.6)$$

Twenty samples were randomly drawn and the indicators were computed on each sample for each of the estimators. These indicators were averaged over the twenty samples for the different estimators In addition the mean, variance, mean square error and estimated bias for the estimator of the total were computed over the twenty samples. In order to save space only two measures are reported. The behavior of the estimators on these two measures are representative of the other measures. Letting l denote the subscript for the sample number, the average absolute error is defined as

$$AAE(\hat{\theta}) = \sum_{l=1}^{20} AE_l(\hat{\theta}) / 20 \qquad (3.7)$$

and the absolute average error is defined as

$$ABAE(\hat{\theta}) = \left| \sum_{l=1}^{20} (\hat{\theta}_l / 20) - \theta \right| . \qquad (3.8)$$

The values for each estimator are reported in Tables 1 and 2, where Table 1 represents the situations of 100% response rate and Table 2 represents the 80% response rate. The mean and variance of the estimators over the 20 samples are reported in Tables 3 and 4.

4. Conclusions

From Table 1, which represents the 100% response rate, it is clear that the ranking of the six estimators is the same for the two months by average absolute error, AAE, and by absolute average error, ABAE. However the first through fourth ranked estimators according to ABAE are reversed according to AAE. The top four estimators according to ABAE are RW2, LR, R1 and RW1. The top four estimators according to AAE (as well as variance) are RW1, R1, LR and RW2.

In terms of estimating change the top four estimators are R1, RW1, HT and LR. In this situation of 100% response rate, it is clear that the regression estimators with non-zero intercept do the best, but the link relative estimator, LR, is not far behind.

From Table 2, which represents the 80% response rate one can see that the top three estimators are the same for the two months by AAE and ABAE; the top three are LC1, LC2 and LR. From Table 4, it is seen that the top three estimators according to the variance are for month 1: RW2, R1, RW1 and for month 2: RW1, R1, LC1. It is noted that the variance of ROW is smaller than the variance of LR, which one would expect, since ROW estimates the sample total by its known value. (In computing R2, if the data with missing values had been omitted, the variance would have increased by 50 percent.)

In terms of estimating change the top three estimators are LR, LC2 and R0. In this more realistic situation of 80% response rate, the slight modifications of the link relative estimators, LC1 and LC2, are good estimators. Again LR is not far behind the best two for level and it is the best estimator for estimating change in this situation. It is clear that the regression type estimators will not do as well as one gets further away from the benchmark month. Woodruff (1982) using simulated data found this to be the case. Future work will include variance estimation and alternative sampling plans.

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) Woodruff S. (1982), "A Comparison of Some Estimators in Sampling For a Time Series with Linear Trend" ASA Proceedings of the Section on Survey Research Methods.

Acknowledgement

Thanks are due to Wesley L. Schaible, Alfreda Reeves and Margo Minor for their help in preparing this paper.

TABLE 1

Absolute Average Error and Average
Absolute Error over 20 Samples
(100% Response Rate)

$\hat{Y}$	ABAE($\hat{Y}_1$)	ABAE($\hat{Y}_2$)	AAE($\hat{Y}_1$)	AAE($\hat{Y}_2$)	AAE($\hat{D}$)
HT	977(5)	1071(5)	1651(5)	1978(5)	1285(3)
LR	288(2)	478(2)	1359(3)	1460(3)	1310(4)
R0	2509(6)	4647(6)	2966(6)	4756(6)	2521(6)
R1	410(3)	774(3)	1268(2)	1400(2)	1200(1)
RW1	534(4)	819(4)	1260(1)	1382(1)	1263(2)
RW2	274(1)	96(1)	1363(4)	1612(4)	1696(5)

Note ROW = LC1 = LC2 = LR

TABLE 2

Absolute Average Error and Average
Absolute Error over 20 Samples
(80% Response Rate)

$\hat{Y}$	ABAE($\hat{Y}_1$)	ABAE($\hat{Y}_2$)	AAE($\hat{Y}_1$)	AAE($\hat{Y}_2$)	AAE($\hat{D}$)
HT	1381(4)	987(4)	2111(5)	2788(5)	1957(6)
LR	316(3)	632(3)	1508(3)	1791(3)	1162(1)
ROW	1838(9)	3747(10)	2195(8)	3747(10)	2180(8)
R0	1388(5)	1984(5)	2484(9)	2821(6)	1408(3)
R1	1710(7)	3734(9)	2158(6)	3734(9)	2090(7)
RW1	1793(8)	3563(8)	2161(7)	3562(8)	1948(5)
RW2	1633(6)	2425(6)	1791(4)	2489(4)	2196(9)
R2	-	3293(7)	-	3415(7)	-
LC1	186(1½)	127(1)	1474(1½)	1715(2)	1518(4)
LC2	186(1½)	406(2)	1474(1½)	1627(1)	1210(2)

TABLE 3

Mean and Variance of Estimators
over 20 Samples
(100% Response Rate)

$\hat{Y}$	$\bar{\hat{Y}}_1$	$\bar{\hat{Y}}_2$	$S^2(\bar{\hat{Y}}_1)$	$S^2(\bar{\hat{Y}}_2)$
HT	74,494	77,248	3,872,240	4,846,659
LR	73,805	76,655	2,862,720	4,100,312
R0	71,008	71,530	5,581,631	6,014,575
R1	73,927	76,951	2,155,152	3,313,726
RW1	74,052	76,996	2,295,155	3,124,342
RW2	73,790	76,081	2,936,199	4,818,531

TABLE 4

Mean and Variance of Estimators
over 20 Samples
(80% Response Rate)

$\hat{Y}$	$\bar{\hat{Y}}_1$	$\bar{\hat{Y}}_2$	$S^2(\bar{\hat{Y}}_1)$	$S^2(\bar{\hat{Y}}_2)$
HT	74,898	77,164	5,751,956	11,358,945
LR	73,833	76,809	4,102,258	5,862,324
ROW	75,355	79,924	3,806,600	5,438,332
R0	72,129	74,193	8,874,145	8,085,278
R1	75,227	79,911	3,473,652	4,073,983
RW1	75,310	79,740	3,481,525	3,755,546
RW2	75,150	78,602	3,332,533	5,863,097
R2	-	79,470	-	8,781,949
LC1	73,330	76,304	3,671,790	4,323,052
LC2	73,330	76,583	3,671,790	5,145,155