ON LINK RELATIVE ESTIMATORS II

Lillian H. Madow, U.S. Bureau of Labor Statistics
and William G. Madow, Washington, D.C.

1. Introduction

This paper is a continuation of a paper, Madow and Madow (1978), referred to as M below, which contains the background and notation for the following discussion. It should be noted that the sets of elements $s_1, s_2, \ldots$ reporting at times 1, 2,... need not be independently selected samples. For example, a sample, s, may be selected and at time a, $a=1,\ldots,g$, s_a is the subset of s consisting of elements that report both at times a and a-1. Thus, the design is longitudinal.

This paper is limited to a discussion of two aspects of link relative estimators. No alternative estimators or general designs are discussed. Clearly, the results apply to stratified link relative estimators and would be useful for composite estimators depending on link relative estimators. The latter are not discussed because to do so, it would be necessary to state a general design. However, composites of link relative estimators are discussed below.

a. Estimates are often needed for characteristics for which there are no benchmarks, even though there is a benchmark for a related variable. Link relative type estimators are discussed for such characteristics in Section 2.

b. In a voluntary survey, reporting units may skip one or more time periods, but then report again. Estimators dealing with some aspects of non-consecutive reporting are discussed in Section 3.

2. Characteristics Not Having a Benchmark

In addition to the variables for which benchmarks are available, there may be variables for which no benchmark is available. In a single survey, estimators of totals for such variables are often defined to be single or multiple ratio or regression estimators based on the benchmarked variables.

In this section we consider two link relative estimators for variables ,Y, that have no benchmarks. One estimator is

$$Y'_{gX} = X_0 R'_{X1} \cdots R'_{Xg} \frac{Y'_{gg}}{X'_{gg}} = X'_g \frac{Y'_{gg}}{X'_{gg}}$$

where R'_{Xg} is a link relative such as defined in M, and Y'_{gg} is the sum of the values of Y for the same elements as those reporting X in period g.

Then Y'_{gX} is a link relative estimator of the total Y_g. If it is assumed that the pairs $(X_{g1},Y_{g1}),\ldots,(X_{gN},Y_{gN})$ are independent for fixed g and

$$\mathcal{E}(Y_{gi} \mid X_{gi}) = \beta^*_g X_{gi}$$

$$\sigma^2_{Y_{gi} \mid (X_g)} = \sigma^2_{g,Y \mid X} X_{gi} = \sigma^{*2}_g X_{gi}$$

then

$$\mathcal{E}\,[Y'_{gX} \mid (g-1),(X_g)\,] = \beta^*_g X'_g$$

and

$$\mathcal{E} Y'_{gX} = \mathcal{E} Y_g$$

since

$$\mathcal{E} Y_g = \beta^*_g \mathcal{E} X_g$$

Also,

$$\mathcal{E}(Y'_{gX} - Y_g)^2 = \beta^{*2}_g \sigma^2_{X'_g - X_g} + \mathcal{E}\sigma^{*2}_{g,Y \mid X} \left(\frac{X'_g(X'_g - X'_{gg})}{X'_{gg}} \right)$$

An unbiased estimator of β^{*2}_g is

$$b^{*2}_g = \left(\frac{Y'_{gg}}{X'_{gg}} \right)^2 - \frac{1}{X'_{gg}} v^2_{g,Y \mid X}$$

where

$$v^2_{g,Y \mid X} = \frac{\sum_{i \varepsilon s_g} \left(Y_{gi} - \frac{Y'_{gg}}{X'_{gg}} X_{gi}\right)^2}{(n_g - 1)\, \bar{X}'^2_g \,(1 - CV^2_{\bar{X}'_g})},$$

$$CV^2_{\bar{X}'_g} = \frac{\sum_{i \varepsilon s_g} (X_{gi} - \bar{X}'_g)^2}{n_g(n_g - 1)\, \bar{X}'^2_g}$$

Further, $v^2_{g,Y|X}$ is an unbiased estimator of $\sigma^2_{g,Y|X}$.

Finally, an unbiased estimator, $v^2_{X'_g - X_g}$, of $\sigma^2_{X'g - Xg}$ was obtained in M.

Thus, if we define

$$v^2_{Y'_{gX} - Y_g} = b^{*2}_g v^2_{X'g - Xg} + v^2_{g,Y|X} \frac{X'_g (X'_g - X'_{gg})}{X'_{gg}}$$

it follows that $v^2_{Y'_{gX} - Y_g}$ is an unbiased estimator of $\mathcal{E}(Y'_{gX} - Y_g)^2$.

A second estimator of Y_g is Y''_g where

$$Y''_{gX} = X_0 \frac{Y'_{00}\ Y'_{11} \cdots Y'_{gg}}{X'_{00}\ Y'_{10} \cdots Y'_{g-1,g}} = \frac{X_0 \frac{Y'_{00}}{X'_{00}}}{Y_0} Y'_g$$

Since $\mathcal{E} X_0 \frac{Y'_{00}}{X'_{00}} = Y_0$, it follows that $\mathcal{E} Y''_{gX} = Y_g$.

Also, $\mathcal{E}(Y''_{gX} - Y_g)^2 = \sigma^2_{Y''_{gX} - Y_g}$ has an unbiased estimator obtained as above.

Whether Y'_{gX} or Y''_{gX} will be used will depend on whether better estimates are available for the link relatives in X or in Y.

3. Reporting for Non-Consecutive Periods

In voluntary longitudinal surveys, elements skip reports and then report again. The skipping cannot be assumed to be random although, for many of the elements, it may be. The skipping is voluntary and thus the skipping pattern differs from rotation designs.

In this section we first consider a simple case in which the elements that skip and the elements that do not skip are both assumed to yield unbiased estimators of change over the two periods. Thus a composite estimator can be used to obtain the combined two-period change. This estimation procedure is used to obtain estimates of level for each time period. Estimates of change for single time periods are computed from the estimates of level.

First, consider two estimates of change over an interval of two reporting periods.

The usual link relative of change over two periods is

$$R'_{g-1}\ R'_g$$

where

$$R'_h = \frac{Y'_{hh}}{Y'_{h-1,h}}$$

and the numerator and denominator are sums of values of Y at times h and h-1 respectively, for units reporting both at times h and h-1.

An alternative estimator is

$$R'_{g-2,g} = \frac{Y'_{gg(2)}}{Y'_{g-2,g(2)}}$$

where the numerator and denominator are totals of Y for times g-2 and g, for units reporting both at times g-2 and g, but not reporting at time g-1.

Every element with data either in R'_g or R'_{g-1} reports for time g-1. Thus, while R'_g and R'_{g-1} have elements in common, neither has elements in common with $R'_{g-2,g}$.

A composite estimator of ζ is defined by

$$Z = (1 - \lambda)\ R'_{g-1}\ R'_g + \lambda R'_{g-2},\ 0 \le \lambda \le 1$$

where ζ is defined to be

$$\zeta = \frac{Y_g}{Y_{g-2}}$$

Suppose that $R'_{g-1}\ R'_g$ and $R'_{g-2,g}$ are unbiased estimators of ζ. Then Z is an unbiased estimator of ζ. We are concerned with the period since g-2 and hence we condition on g-2 and earlier periods, using (h) to indicate conditioning in periods 0,1,..., h.

$$(Z - \zeta)^2 \mid (g-2) = \sigma^2_{Z \mid (g-2)}$$

$$= (1 - \lambda)^2 \sigma^2_{R'_{g-1}R'_g - \zeta \mid (g-2)} + \lambda^2 \sigma^2_{R'_{g-2,g} - \zeta \mid (g-2)}$$

Minimization of σ^2_Z with respect to λ results in

$$\hat{\lambda} = \frac{\sigma^2_{R'_{g-1}R'_g - \zeta | (g-2)}}{\sigma^2_{R'_{g-1}R'_g - \zeta | (g-2)} + \sigma^2_{R_{g-2,g} - \zeta | (g-2)}}$$

and

$$\min \sigma^2_Z | (g-2) = \frac{\sigma^2_{R'_{g-1}R'_g - \zeta | (g-2)} \; \sigma^2_{R'_{g-2,g} - \zeta | (g-2)}}{\sigma^2_{R'_{g-1}R'_g - \zeta | (g-2)} + \sigma^2_{R'_{g-2,g} - \zeta | (g-2)}}$$

Thus, if one of the variances is much larger than the other, little would be gained by using the composite estimator; min $\sigma^2_Z \mid (g-2)$ would be approximately equal to the smaller variance.

It can be shown that

$$\sigma^2_{R'_{g-2,g} - \zeta | (g-2)} = \left(\beta^2_g \sigma^2_{g-1} + \beta_{g-1} \sigma^2_g \right) \left(\frac{1}{Y'_{g-2,g}} - \frac{1}{Y_{g-2}} \right)$$

An almost unbiased estimator is obtained by replacing Y_{g-2} by Y'_{g-2} and estimating the parameter $\beta^2_g \sigma^2_{g-1} + \beta_{g-1} \sigma^2_g$ by

$$v^2_{g-2,g} = \frac{\sum_{i \in s_{g-2,g}} (Y_{gi} - R'_{g-2,g} Y_{g-2,i})^2}{(n_{g-2,g} - 1) \; Y'_{g-2} (1 - CV^2_{Y'_{g-2}})}$$

where $n_{g-2,g}$ is the number of elements in $s_{g-2,g}$.

$$\overline{Y}'_{g-2} = \frac{1}{n_{g-2,g}} \sum_{i \in s_{g-2,g}} Y_{g-2,i}$$

and

$$CV^2_{\overline{Y}'_{g-2}} = \frac{\sum_i (Y_{g-2,i} - \overline{Y}'_{g-2})^2}{n_{g-2,g} (n_{g-2,g} - 1) \; \overline{Y}'^2_{g-2}}$$

Also

$$\sigma^2_{R'_g R'_{g-1} - \zeta | (g-2)} = \beta^2_g \sigma^2_{g-1} \left(\frac{1}{Y'_{g-2,g-1}} - \frac{1}{Y_{g-2}} \right) + \sigma^2_g \mathcal{E} \left[R'_{g-1} \left(\frac{R'_{g-1}}{Y'_{g-1,g}} - \frac{1}{Y_{g-2}} \right) \right]$$

and an almost unbiased estimator of $\sigma^2_{R'_g R'_{g-1} - \zeta}$ is obtained by replacing Y_{g-2} by its estimator Y'_{g-2}, and the parameters $\beta^2_g \sigma^2_{g-1}$ and σ^2_g by their unbiased estimators conditional on (g-2) as obtained in M.

<u>Second</u>, consider estimates of level. Various estimators of level can be defined. The one that we consider is suggested by the desire to have the current estimator of level, Y'_g, be one term of the composite estimator of level.

Thus, we define

$$Y''_g = (1 - \lambda) Y'_g + \lambda \; Y'_{g-2} R'_{g-2,g}$$

$$= Y'_{g-2} Z$$

An alternative estimator would be

$$Y'''_g = Y''_{g-2} \; Z$$

A third estimator would be

$$Y^{iv}_g = (1 - \lambda) Y'_g + \lambda \; Y'''_{g-2} R'_{g-2,g}$$

All estimates of level that are mentioned above are unbiased estimates of Y_g.

Then

$$\mathcal{E} (Y''_g - Y_g)^2 = \sigma^2_{Y''_g - Y_g}$$

$$= (1-\lambda)^2 \sigma^2_{Y'_g - Y_g} + 2\lambda(1-\lambda) \sigma_{(Y'_g - Y_g),(Y'_{g-2} R'_{g-2,g} - Y_g)}$$

$$+ \lambda^2 \sigma^2_{(Y'_{g-2} R'_{g-2,g} - Y_g)}$$

An unbiased estimator of $\sigma^2_{Y'_g - Y_g}$ has already been obtained in M. Unbiased estimators of the other terms may be obtained by using the procedures discussed above.

An alternative approach takes advantage of the fact that many units report in all three time periods g, g-1 and g-2.

Define $R''_{g,g-2}$ by

$$R''_{g,g-2} = \frac{Y''_{g(g,g-2)}}{Y''_{g-2(g,g-2)}}$$

where $s_{g,g-2}$ is the set of all units that report both in period g and in period g-2. Then $R''_{g,g-2}$ will usually be correlated with $R'_g R'_{g-1}$ and

$$\sigma_{R'_g R'_{g-1},\ R''_{g,g-2}} \ (g-2)$$

$$= \beta^2_g \ \sigma^2_{g-1} \ \frac{Y^*_{g-2}}{Y'_{g-2,g-1} \ Y'_{g-2,g}}$$

$$+ \ \mathcal{E} \ \frac{Y'_{g-1,\,g-1}}{Y_{g-2,g-1} \ Y_{g-1,g} \ Y_{g-2,g}} \ \sigma^2_g \ Y^*_{g-1}$$

where Y^*_{g-1} and Y^*_{g-2} are sums of $Y_{g-1,i}$ and $Y_{g-2,i}$ over the units reporting in all three periods, g, g-1, g-2.

An unbiased estimator of the covariance of $R_g R_{g-1}$ and $R''_{g,g-2}$ can be obtained by the same procedures as those used above.

Using the covariance, it follows that the variance of a composite estimate of change in the general case can be obtained. An unbiased estimator of the variances is also obtained as above.

Composite unbiased estimators of level may also be stated, e.g.,

$$Y''_g = (1 - \lambda) Y'_g + \lambda Y'_{g-2} R''_{g,g-2}$$

The variance of Y''_g and an unbiased estimate of that variance are also easily obtained by the procedures stated earlier.

The optimizing value of λ is

$$\hat{\lambda} = \frac{\sigma^2_{R'_g R'_{g-1}} \mid (g-1) - \sigma_{R'_g R'_{g-1},\, R'_{g,g-2}} \mid (g-2)}{\sigma^2_{R_g R_{g-1}} \mid (g-1) - 2\sigma_{R'_g R'_{g-1},\, R''_{g,g-2}} \mid (g-2) + \sigma^2_{R''_{g,g-2}} \mid (g-2)}$$

which can be approximated from the estimates needed for the variance of the estimator.

REFERENCES

Madow, L.H. and Madow, W.G. (1978), "On Link Relative Estimators," ASA Proceedings of the Section on Survey Research Methods, 1978, pp. 534-539.