ON LINK RELATIVE ESTIMATORS II
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, \ldots need not
be independently selected samples. For example, a
sample, $s$, may be selected and at time $a$, $a=1, \ldots, g$, $s$
is the subset of $s$ consisting of elements that report
both at times $a$ and $a-1$. Thus, the design is longitudi-
unal.

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 depend-
ing 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 character-
istics for which there are no benchmarks, even though
there is a benchmark for a related variable. Link
relative type estimators are discussed for such charac-
teristics 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-consecu-
tive reporting are discussed in Section 3.

2. Characteristics Not Having a Benchmark

In addition to the variables for which bench-
marks are available, there may be variables for which
no benchmark is available. In a single survey, estima-
tors 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

$$Y'_{gX} = Y'_{gg} = \frac{Y_{gg}^g}{X_{gg}^g} \frac{X'_{gg}}{X_{gg}} R'_{Xg}$$

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 $Y$. If it is assumed that the pairs $(X_{g1} Y_{g1}), \ldots,
(X_{gN} Y_{gN})$ are independent for fixed $g$.

and

$$E(Y_{g1} | X_{g1}) = \beta_{g1} Y_{g1}$$

$$\sigma^2 \left( \frac{Y_{g1} - \beta_{g1} X_{g1}}{g} \right)^2 = \sigma^2 \left( \frac{X_{g1} - \beta_{g1}}{g} \right)^2$$

then

$$E(Y'_{gX} | (g-1), (X_{g})) = \beta_{g} X_{g}$$

and

$$E(Y'_{gX}) = E(Y_{g})$$

since

$$E(Y_{g}) = \beta_{g} C X_{g}$$

Also,

$$E(Y'_{gX} - Y_{g}) = \beta_{g} C X_{g}$$

An unbiased estimator of $\beta_{g}^2$ is

$$b_{g}^2 = \left( \frac{Y'_{gg}}{X_{gg}^g} \right)^2 - \frac{1}{X'_{gg}^g} v^2 Y_{gX} | X$$

where

$$v^2 = \frac{\sum_{i \in g} (Y_{gi} - Y_{gg}^g X_{gi}^g)^2}{(n - i) \frac{X'_{gg}^g}{X_{gg}^g} (1 - CV_{Xg}^2)}$$

$$CV_{Xg}^2 = \frac{\sum_{i \in g} (X_{gi} - \bar{X}_{gi}^g)^2}{n \frac{X'_{gg}^g}{X_{gg}^g} (n - 1) \bar{X}_{gi}^g}$$
Further, $\sigma^2_{g,Y \mid X}$ is an unbiased estimator of $\sigma^2_{g,Y \mid X}$.

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

Thus, if we define

$$\sigma^2_{g,X'g - Xg} = \beta^2_g \sigma^2_{g,Y \mid X} + \nu^2_{g,Y \mid X} X'g (X' - Xg)$$

it follows that $\sigma^2_{g,X'g - Xg}$ is an unbiased estimator of $E(Y'gX - Yg)$.

A second estimator of $Yg$ is $Y''g$ where

$$Y''g = \sum_{g=0}^{g-1} \frac{Y'00 Y'11 \cdots Y'gg}{Y'00 - Y_0}$$

Since $\sum_{g=0}^{g-1} \frac{Y'00}{X'00} = Y'0$, it follows that $E(Y''gX - Yg) = 0$.

Also, $E(Y''gX - Yg)^2 = \sigma^2_{YgX - Yg}$ 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

$$R'_{g-1} \frac{Y'_{g-1}}{Y'_{g}}$$

where

$$R'_{g} = \frac{Y'_{h}}{Y'_{h-1}}$$

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

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

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 $g$.

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

A composite estimator of $\zeta$ is defined by

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

where $\zeta$ is defined to be

$$\zeta = \frac{Y'_{g-2}}{Y'_{g}}$$

Suppose that $R'_{g-1}$, $R'_{g}$, and $R'_{g-2}$ are unbiased estimators of $\zeta$ . Then $Z$ is an unbiased estimator of $\zeta$ . 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 $0, 1, \ldots, h$.

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

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

337
Minimization of \( \sigma_Z^2 \) with respect to \( \lambda \) results in

\[
\hat{\lambda} = \frac{\sigma^2_{R'g-2g} - \zeta | (g-2)}{\sigma^2_{R'g-1g} - \zeta | (g-2)} + \frac{\sigma^2_{R'g-2g} - \zeta | (g-2)}{\sigma^2_{R'g-2g} - \zeta | (g-2)}
\]

and

\[
\min \sigma_Z^2 | (g-2) = \frac{\sigma^2_{R'g-1g} - \zeta | (g-2)}{\sigma^2_{R'g-1g} - \zeta | (g-2)} + \frac{\sigma^2_{R'g-2g} - \zeta | (g-2)}{\sigma^2_{R'g-2g} - \zeta | (g-2)}
\]

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

It can be shown that

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

An almost unbiased estimator is obtained by replacing \( Y_{g-2} \) by \( Y'_{g-2} \) and estimating the parameter \( \beta \) by

\[
\beta = \frac{1}{\sigma^2_{g-2g}} \sum_{i \in S_{g-2g}} (Y_{g-i} - R'_{g-2g} Y_{g-2i})^2
\]

\[
v_{g-2g}^2 = \frac{1}{n_{g-2g} - 1} Y_{g-2} (1 - CV^2_{Y'})_{g-2}
\]

where \( n_{g-2g} \) is the number of elements in \( S_{g-2g} \).

\[
Y'_{g-2} = \frac{1}{n_{g-2g}} \sum_{i \in S_{g-2g}} Y_{g-2i}
\]

and

\[
CV_{Y'}^2 = \frac{\sum_{i \in S_{g-2g}} (Y_{g-2i} - Y'_{g-2})^2}{n_{g-2g} (n_{g-2g} - 1) Y_{g-2}^2}
\]

Also

\[
\sigma^2_{R'g-1g} - \zeta | (g-2) = \frac{\sigma^2_{g-1g}}{Y'_{g-2g} Y_{g-2}} \left( \frac{1}{\sigma_{g-2g}^2} - \frac{1}{\sigma_{g-2g}^2} \right)
\]

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

Second, 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

\[
Y''_{g} = (1 - \lambda) Y'_{g} + \lambda Y'_{g-2} R'_{g-2g}
\]

\[
= Y'_{g-2} Z
\]

An alternative estimator would be

\[
Y'''_{g} = Y''_{g} Z
\]

A third estimator would be

\[
Y''''_{g} = (1 - \lambda) Y'_{g} + \lambda Y''''_{g-2} R'_{g-2g}
\]

All estimates of level that are mentioned above are unbiased estimates of \( Y'_{g} \).

Then

\[
\mathbb{E} (Y''_{g} - Y'_{g})^2 = \sigma_{Y''_{g} - Y'_{g}}^2
\]

\[
= (1 - \lambda)^2 \sigma_{Y'_{g} - Y'_{g}}^2 + 2 \lambda (1 - \lambda) \sigma_{Y'_{g} - Y'_{g}} \sigma_{Y''_{g-2} R'_{g-2g}} + \lambda^2 \sigma_{Y''_{g-2} R'_{g-2g}}^2
\]

\[
+ \lambda^2 \sigma_{Y''_{g-2} R'_{g-2g}}^2 + \lambda^2 \sigma_{Y''_{g-2} R'_{g-2g}}^2
\]
An unbiased estimator of $\frac{\sigma^2}{Y^g - Y^g}$ 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

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

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 $R_{g,g-1}'$ and $R_{g,g-2}'$.

$$\sigma^2 R_{g,g-1}' R_{g,g-2}' (g-2)$$

$$= \sigma^2 \frac{Y^*,g-2}{g-1} \frac{Y^*,g-2,g-1}{g-2} \frac{Y^*,g-1}{g-2}$$

$$+ \sum \frac{Y^*,g-1,g-1}{g-2,g-1} \frac{Y^*,g-1,g}{g-2,g}$$

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 $R_{g,g-1}'$ 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.,

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

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 $\lambda$ is

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

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

REFERENCES