BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION

SRD Research Report Number: Census/SRD/88/26

VARIANCE FORMULAE FOR THE GENERALIZED COMPOSITE ESTIMATOR UNDER A BALANCED ONE-LEVEL ROTATION PLAN

by

Patrick J. Cantwell
Statistical Research Division
Bureau of the Census
Room 3134, F.O.B. #4
Washington, D.C. 20233 U.S.A.

Recommended: Lawrence Ernst

Report completed: December 1988

Report issued: December 27, 1988

Variance Formulae for the Generalized Composite Estimator Under A Balanced One-Level Rotation Plan

ABSTRACT

In many surveys, including the Current Population Survey of the U.S. Bureau of the Census and the Labour Force Survey of Statistics Canada, participants are interviewed a number of times during the life of the survey, a practice referred to as a rotation design or repeated sampling. Often composite estimation--where data from the current and earlier periods of time are combined--is used to measure the level of a characteristic of interest. As other authors have observed, composite estimation can be used in a rotation design to decrease the variance of estimators of change in level. We derive simple expressions for the variance of a general class of composite estimators for level, average level over time, and change in level. These formulae hold under a wide range of rotation designs.

1. INTRODUCTION

The Current Population Survey of the U.S. Bureau of the Census and the Labour Force Survey of Statistics Canada are two examples of repeated sampling or rotation designs. In each case, households are interviewed a number of times before leaving the sample. In the CPS, households are interviewed for four months, then leave the sample for eight months, and finally return for four more months. In the LFS, participating households respond for six consecutive months and do not return.

A major advantage of using a rotation design is the smaller variance for estimates of change when measurements within groups are positively correlated from one time period to the next. For the CPS and the LFS, there are sample overlaps of 75% and 83%, respectively, from one month to the next. Estimates of month-to-month change or year-to-year change can be improved by selecting the proper plan and estimator. Respondent burden can be lessened by manipulating the sequence of periods when respondents are in and out of sample. For more on these topics, see Woodruff (1963), Rao and Graham (1964), or Wolter (1979).

Every ten years, during the redesign of the current surveys of the Census Bureau, many aspects of the various surveys are modified. When evaluating these changes, it may be appropriate to consider implementing a different rotation scheme. Similarly, a researcher planning a new survey may decide to use a rotation design, but must select one which accommodates his needs. Any such plan requires the variance formulae for the estimators of level and change.

Such variance derivations are not conceptually difficult, but can be quite tedious. Some of the more common estimators are "composite" in nature. In order to take advantage of repeated sampling, they combine information from the present with information from one or more previous periods. Partial estimates obtained from the same rotation group at different times are combined into a final estimator. While the variance can be decreased by selecting the combination judiciously, calculating this variance may become more complex because of the correlation patterns involved among the repeated groups.

For a general rotation plan, subject to specific restrictions, we present simple formulae for the variance of estimators of level and change. The derivations are applied to an important and quite general class of estimators called the general composite estimator (Breau and Ernst 1983). Although CPS and LFS use different estimators and rotation plans, each will be a special case of those we consider.

In Section 2, we define the generalized composite estimator and state results. An example is provided in Section 3. Proofs of the theorems are given in Section 4.

2. NOTATION AND RESULTS

Although rotation schemes can assume infinitely many forms, we restrict this discussion to one type. At each period in time, a new rotation group enters the sample, and follows the same pattern of periods in and out of sample as every preceding group. In addition, responses refer only to the current period of time, whether or not the participants were in sample in the previous period. We call this design a "balanced one-level" rotation plan. The design is "balanced" because the number of groups in sample at any time is equal to the total number of time periods any one group is included in the sample. Wolter (1979) uses the terms one-level and two-level to indicate the number of periods for which information is solicited in one interview.

The scheme used in the LFS satisfies these restrictions. Each month a new group enters, and remains in the sample for five more months. The CPS as it currently operates follows these guidelines in a 4-8-4 scheme. Before July 1953, however, CPS used a plan where five rotation groups entered, one each in consecutive months. In the sixth month, no new group entered. The process then continued in the same manner, with groups exiting after six months in sample.

One problem with the pre-1953 CPS design is the introduction of month-in-sample bias, often referred to as rotation group bias. Of greater concern here is the changing pattern of rotation group appearances. The variance of a composite estimate depends on when each participating group appeared in sample before, and the covariance structure for identical groups in different months. If the pattern of appearances changes from month to month, the variance formula of the estimator also changes. Under a balanced design with stationary covariance structure, general derivations are possible.

Throughout this paper, we will use "month" to refer to the period of time in which interviews are done, for brevity and because CPS and LFS use the month to divide the life of the survey. However, our results will apply to any period of time, provided the rotation plan is balanced and one-level.

Suppose that every rotation group is in sample for a total of m months over a period of M months, i.e., it is out of sample for $M-m$ months after first entering and before exiting. Because the rotation design is balanced, m groups are in sample during any month. Let $x_{h,i}$ denote the estimate of "monthly" level from the rotation group which is in sample for the ith time in month h. We treat only the generalized composite estimator (GCE), as defined recursively by Breau and Ernst (1983). For monthly level:

$$y_h = \sum_{i=1}^{m} a_i x_{h,i} - k \sum_{i=1}^{m} b_i x_{h-1,i} + k y_{h-1} , \tag{1}$$

where k, the a_i's and the b_i's may take any values subject to $0 \leq k < 1$, $\sum_{i=1}^{m} a_i = 1$, and $\sum_{i=1}^{m} b_i = 1$. The composite and AK composite estimators used in CPS are special cases of the GCE. For information on these, see Gurney and Daly (1965), Hanson (1978), Huang and Ernst (1981), and Kumar and Lee (1983).

The GCE is more restrictive than a general linear estimator which combines $x_{n,i}$ values from the current and many prior months. However, the GCE has been shown to perform almost as well (Breau and Ernst 1983). It has the advantage that only data from two months--the current month and the preceding one--need be stored. Although y_h incorporates earlier data, it is summarized through y_{h-1}.

To find expressions for the variance of the GCE, we assume a stationary covariance structure:

(i) $\text{Var}(x_{h,i}) = \sigma^2$ for all h and i;

(ii) $\text{Cov}(x_{h,i}, x_{h,j}) = 0$ for $i \neq j$, i.e., different rotation groups in the same month are uncorrelated; and

(iii) $\text{Cov}(x_{h,i}, x_{s,j}) = \rho_{|h-s|}\sigma^2$, if the two x's refer to the same rotation group $|h-s|$ months apart; or 0, otherwise. Take ρ_0 to be 1. (2)

As an example, the covariance structure for the 4-8-4 plan is specified in Breau and Ernst (1983).

Before stating our results, we introduce notation. Let us define the set T_0 as follows. Consider any rotation group. Let T_0 index the set of "months" when this group is not in sample, labeling as month one the month this group is first interviewed, but not going beyond month $\boldsymbol{M}$. Because the rotation plan is balanced, the composition of T_0 does not depend on which group is selected.

Next we create the $\boldsymbol{M}\times 1$ vector $\boldsymbol{a}$. Define the ith component of $\boldsymbol{a}$ to be 0 if $i \in T_0$. This step fills $\boldsymbol{M}-m$ positions in $\boldsymbol{a}$. Then the values $a_1, a_2, \ldots, a_m$ are inserted in order into the remaining m components, starting with the first. We call this a vector in "TIS (time-in-sample) form." For example, in a 4-8-4 rotation plan, $T_0 = \{5, 6, \ldots, 12\}$, and $\boldsymbol{a}^{\mathrm{T}} = (a_1, a_2, a_3, a_4, 0, 0, 0, 0, 0, 0, 0, 0, a_5, a_6, a_7, a_8)$. The $\boldsymbol{M}\times 1$ vector $\boldsymbol{b}$ is formed analogously in TIS form.

Let $\boldsymbol{L}$ be the $\boldsymbol{M}\times\boldsymbol{M}$ matrix with 1's on the subdiagonal, and 0's elsewhere. Formally, $\boldsymbol{L}_{ij} = 1$, if $i-j = 1$, and 0, otherwise. Define the $\boldsymbol{M}\times\boldsymbol{M}$ matrix $\boldsymbol{Q}$ by: $\boldsymbol{Q}_{ij} = k^{i-j}\rho_{i-j}$, if $1 \leq j < i \leq \boldsymbol{M}$, and 0, otherwise. Finally, let $\boldsymbol{I}$ be the $\boldsymbol{M}\times\boldsymbol{M}$ identity matrix.

We state several theorems, and leave the proofs to Section 4.

Theorem 1. If the GCE of level is defined as in (1), and the covariance structure of (2) holds, then

$$\mathrm{Var}(y_h) = \sigma^2 \{ \boldsymbol{a}^T\boldsymbol{a} + k^2\boldsymbol{b}^T(\boldsymbol{b}-2\boldsymbol{a}) + 2(\boldsymbol{a}-k^2\boldsymbol{b})^T\boldsymbol{Q}(\boldsymbol{a}-\boldsymbol{b})\} \;/\; (1-k^2) \tag{3}$$

Notice that when one uses an unweighted average of the estimates from the m rotation groups of the current month, $k = 0$, $\boldsymbol{Q} = 0$, and $a_i = 1/m$ for $i = 1, 2, \ldots, m$. Then $\mathrm{Var}(y_h) = \sigma^2/m$, as expected.

THEOREM 2. Let $y_h - y_{h-1}$ be the GCE estimator of "monthly" change.

(i) If $k = 0$, then $\mathrm{Var}(y_h - y_{h-1}) = 2\sigma^2\, \boldsymbol{a}^T(\boldsymbol{I}-\rho_1\boldsymbol{L})\boldsymbol{a}$;

(ii) if $0 < k < 1$, then $\mathrm{Var}(y_h - y_{h-1})$

$$= \sigma^2(\boldsymbol{a}^T\boldsymbol{a} + k^2\boldsymbol{b}^T\boldsymbol{b} - 2k\rho_1\boldsymbol{a}^T\boldsymbol{L}\boldsymbol{b})/k - (1-k)^2\mathrm{Var}(y_h)/k \tag{4}$$

The average level over a period of time, perhaps a year, is often of interest. It suffices to find the variance of the appropriate sum. Denote by $S_{h,t}$ the sum of the GCE's for the last t+1 months: $S_{h,t} = y_h + y_{h-1} + \ldots + y_{h-t}$, $t \geq 0$.

THEOREM 3. Define the series of M×1 vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$:

$$\boldsymbol{v}_i = \begin{cases} \boldsymbol{a} + [(k - k^{i+1})/(1-k)]\ (\boldsymbol{a}-\boldsymbol{b}), & i = 0, 1, \ldots, t, \\ [(k - k^{t+2})/(1-k)]\ k^{i-t-1}(\boldsymbol{a}-\boldsymbol{b}), & i = t+1, t+2, \ldots \end{cases}$$

Then

$$\mathrm{Var}\,(S_{h,t}) = \sigma^2 \left\{ \sum_{i=0}^{\infty} \boldsymbol{v}_i^T\boldsymbol{v}_i + 2\sum_{i=0}^{\infty} \boldsymbol{v}_i^T \sum_{j=1}^{M-1} \rho_j \boldsymbol{L}^j \boldsymbol{v}_{i+j} \right\} \tag{5}$$

The sums in (5) converge because $\boldsymbol{v}_i$ is proportional to $k^i(\boldsymbol{a}-\boldsymbol{b})$ for $i > t$.

Another concern, besides the variance of the estimators, is month-in-sample bias, or more generally, time-in-sample bias. Suppose that $\mathrm{E}(x_{h,i}) = Y_h + \beta_i$, for all h and i, where Y_h is the actual value of the characteristic to be estimated. This model assumes that the bias β_i depends upon how many times the respondent has been interviewed, but not which month or year it is. Stating that the unweighted monthly average of rotation group estimators is unbiased amounts to saying that $\sum_{i=1}^{m} \beta_i = 0$.

THEOREM 4. Let β be the vector of month-in-sample bias terms in TIS form. Under the model above:

(i) $E(y_h) = Y_h + \beta^T(a-kb)/(1-k)$, and

(ii) $E(y_h - y_{h-1}) = Y_h - Y_{h-1}$, i.e., the GCE for monthly change is unbiased.

COROLLARY 1. Under the foregoing assumptions, the mean squared error of the GCE for monthly level is

$$\begin{aligned} MSE(y_h) &= Var(y_h) + [E(y_h) - Y_h]^2 \\ &= \sigma^2 \{a^T a + k^2 b^T(b-2a) + 2(a-k^2 b)^T Q(a-b)\} \,/\, (1-k^2) \\ &\quad + [\beta^T(a-kb)]^2 \,/\, (1-k)^2 \,. \end{aligned}$$

3. AN EXAMPLE

Suppose the rotation pattern is 2-1-1, i.e., a participant is in sample for the first two months, leaves for one month, and then returns for a third month. Suppose further that the correlations for the characteristics to be measured are estimated to be $\rho_1 = .6$, $\rho_2 = .5$, and $\rho_3 = .4$. Good values of k, a_i, and b_i may be found through trial and error, or through optimizing routines (Wolter 1979). In this example, for simplicity let k be .5, the a_i's .3, .5, and .2, and the b_i's .6, .2, and .2, respectively. Then a^T and b^T are (.3, .5, 0, .2) and (.6, .2, 0, .2). The matrix Q is written in general (for the case when $M = 4$) and with these values of k and ρ_i:

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 \\ k\rho_1 & 0 & 0 & 0 \\ k^2\rho_2 & k\rho_1 & 0 & 0 \\ k^3\rho_3 & k^2\rho_2 & k\rho_1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ .3 & 0 & 0 & 0 \\ .125 & .3 & 0 & 0 \\ .05 & .125 & .3 & 0 \end{bmatrix}$$

Vector and matrix computations give $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a} = .38$, $\boldsymbol{b}^{\mathrm{T}}(\boldsymbol{b}-2\boldsymbol{a}) = -.20$, and $(\boldsymbol{a}-k^2\boldsymbol{b})^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{a}-\boldsymbol{b}) = -.037125$. Thus $\mathrm{Var}(y_h) = .341\sigma^2$. This might be compared to $\mathrm{Var}(\bar{x}_h) = \mathrm{Var}[(x_{h,1} + x_{h,2} + x_{h,3})/3] = .333\sigma^2$, where $\bar{x}_h$ is the unweighted average of the estimates from the three rotation groups of the current month.

For the corresponding measure of month-to-month change, $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a} = .38$, $\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b} = .44$, and $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{b} = .30$. The result is $\mathrm{Var}(y_h - y_{h-1}) = .4495\sigma^2$. The appropriate formula in Theorem 2 with $k = 0$ leads us to $\mathrm{Var}(\bar{x}_h - \bar{x}_{h-1}) = .5333\sigma^2$.

Two points should be made here. First, one observes the improvement in variance of this particular composite estimator, compared with the unweighted average of rotation group estimates. Although the GCE suffers a 2.4% increase in variance while estimating monthly level, it realizes a 15.7% decrease while estimating month-to-month change.

In fact, even the unweighted average takes advantage of repeated sampling in its month-to-month change estimate. The correlation of .6 between $x_{h,2}$ and $x_{h-1,1}$ lowers the variance from $.6667\sigma^2$ (using independent samples each month) to $.5333\sigma^2$ under the specified rotation scheme.

Secondly, the values of k, a_i, and b_i in this example were selected for simplicity. Choices which minimize the variance of the estimate of monthly level or month-to-month change, or some combination, can be determined. Such values and the resulting variances demonstrate the benefits of using composite estimation where repeated sampling is involved. However, this paper merely contributes simple formulae for the variances of GCEs under a balanced one-level rotation plan. For more on actual improvements possible with composite estimation, consult the appropriate references given above.

4. DERIVATIONS OF THE THEOREMS

The set T_0 was defined as the set of months, starting in the first month in sample, when any particular rotation group is not in sample. It is useful to introduce the $M \times M$ matrix $\boldsymbol{R}$ as: $\boldsymbol{R}_{ii} = 1$ if $i \notin T_0$, and 0 if $i \in T_0$; and $\boldsymbol{R}_{ij} = 0$ if $i \neq j$. Using terminology from Section 2, $\boldsymbol{R}$ is a diagonal matrix where diag($\boldsymbol{R}$) is a set of 1's in "TIS form." It is always true that $\boldsymbol{R}_{11}$ and $\boldsymbol{R}_{MM}$ are 1, and $\sum_{i=1}^{M} \boldsymbol{R}_{ii} = m$.

Observe that, for any $M \times p$ matrix $\boldsymbol{V}$, $\boldsymbol{RV}$ is the same as $\boldsymbol{V}$, but with 0's across each row i, if i is in T_0. In other words, premultiplication by $\boldsymbol{R}$ "removes" (turns to 0) the rows of $\boldsymbol{V}$ indexed by T_0. If the columns of $\boldsymbol{V}$ are already in TIS form, then $\boldsymbol{RV} = \boldsymbol{V}$. Similarly, for any $p \times M$ matrix $\boldsymbol{U}$, postmultiplication by $\boldsymbol{R}$ "removes" the columns of $\boldsymbol{U}$ which are indexed by T_0. If the rows of $\boldsymbol{U}$ are already in TIS form, then $\boldsymbol{UR} = \boldsymbol{U}$.

Recall that $\boldsymbol{L}$ is an $M \times M$ matrix with 1's on the subdiagonal and 0's elsewhere. For any $M \times 1$ vector written as $\boldsymbol{v}^T = (v_1, v_2, \ldots, v_M)$, the product $\boldsymbol{Lv}$ becomes $(0, v_1, v_2, \ldots, v_{M-1})^T$, and $\boldsymbol{v}^T\boldsymbol{L}$ is $(v_2, v_3, \ldots, v_M, 0)$.

Next we form vectors out of the monthly estimates from the various rotation groups. For any h, let $\boldsymbol{x}_h$ be the $M \times 1$ vector comprising $x_{h,1}$, $\ldots$, $x_{h,m}$ in TIS form. From the first two parts of (2) describing the covariance structure of the estimates, we conclude that $\mathrm{Var}(\boldsymbol{x}_h) = \sigma^2 \boldsymbol{R}$. Part three of (2) implies that $\mathrm{Cov}(\boldsymbol{x}_h, \boldsymbol{x}_{h-1}) = \sigma^2 \rho_1 \boldsymbol{RLR}$. Notice that (a) the matrix $\boldsymbol{L}$, with 1's on the subdiagonal, "represents" the one month lag between the $\boldsymbol{x}_h$ and $\boldsymbol{x}_{h-1}$ values, i.e., there is a nonzero (ρ_1) correlation between $x_{h,i}$ and $x_{h-1,j}$ if $i-j = 1$, and (b) premultiplying (postmultiplying) by $\boldsymbol{R}$ inserts 0's corresponding to 0's in $\boldsymbol{x}_h$ ($\boldsymbol{x}_{h-1}$).

It is readily seen that $L^r_{ij} = 1$ if $i-j = r \geq 0$ and $1 \leq j, i \leq M$ (take L^0 to be the identity matrix). The same development as above gives $\mathrm{Cov}(\boldsymbol{x}_h, \boldsymbol{x}_{h-2}) = \sigma^2 \rho_2 \boldsymbol{I} L^2 \boldsymbol{I}$. In general,

$$\mathrm{Cov}(\boldsymbol{x}_h, \boldsymbol{x}_{h-r}) = \sigma^2 \rho_r \boldsymbol{I} L^r \boldsymbol{I}, \text{ for } r = 0,\ 1,\ 2,\ \ldots, \text{ and all } h. \tag{6}$$

For $r \geq M$, $L^r = 0$, and $\mathrm{Cov}(\boldsymbol{x}_h, \boldsymbol{x}_{h-r}) = 0$.

The generalized composite estimator was written in (1) as

$$y_h = \sum_{i=1}^{m} a_i x_{h,i} - k \sum_{i=1}^{m} b_i x_{h-1,i} + k y_{h-1}$$

This can be put into vector form and made more manageable:

$$\begin{aligned}
y_h &= \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}_h - k \boldsymbol{b}^{\mathrm{T}} \boldsymbol{x}_{h-1} + k y_{h-1} \\
&= \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}_h - k \boldsymbol{b}^{\mathrm{T}} \boldsymbol{x}_{h-1} + k(\boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}_{h-1} - k \boldsymbol{b}^{\mathrm{T}} \boldsymbol{x}_{h-2} + k y_{h-2}) \\
&= \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}_h + k(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}} \boldsymbol{x}_{h-1} - k^2 \boldsymbol{b}^{\mathrm{T}} \boldsymbol{x}_{h-2} + k^2 y_{h-2} \\
&= \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}_h + k(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}} \boldsymbol{x}_{h-1} - k^2 \boldsymbol{b}^{\mathrm{T}} \boldsymbol{x}_{h-2} + k^2(\boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}_{h-2} - k \boldsymbol{b}^{\mathrm{T}} \boldsymbol{x}_{h-3} + k y_{h-3}) \\
&= \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}_h + k(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}} \boldsymbol{x}_{h-1} + k^2(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}} \boldsymbol{x}_{h-2} - k^3 \boldsymbol{b}^{\mathrm{T}} \boldsymbol{x}_{h-3} + k^3 y_{h-3} \\
&= \ldots \\
&= \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}_h + k(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}} \boldsymbol{x}_{h-1} + k^2(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}} \boldsymbol{x}_{h-2} + k^3(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}} \boldsymbol{x}_{h-3} + \ldots \\
&= \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x}_h + (\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}} \sum_{i=1}^{\infty} k^i \boldsymbol{x}_{h-i}
\end{aligned} \tag{7}$$

From the results in (6) and (7), we can proceed with the proofs of the theorems.

Proof of Theorem 1.

$$\begin{aligned}
\mathrm{Var}(y_h) = {} & \boldsymbol{a}^{\mathrm{T}} \mathrm{Var}(\boldsymbol{x}_h) \boldsymbol{a} + (\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}} \sum_{i=1}^{\infty} k^{2i} \mathrm{Var}(\boldsymbol{x}_{h-i})\ (\boldsymbol{a}-\boldsymbol{b}) \\
& + 2\ \boldsymbol{a}^{\mathrm{T}} \sum_{i=1}^{\infty} k^i \mathrm{Cov}(\boldsymbol{x}_h, \boldsymbol{x}_{h-i})\ (\boldsymbol{a}-\boldsymbol{b}) \\
& + 2\ (\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}} \sum_{1 \leq i < j}^{\infty} \sum^{\infty} k^{i+j} \mathrm{Cov}(\boldsymbol{x}_{h-i}, \boldsymbol{x}_{h-j})\ (\boldsymbol{a}-\boldsymbol{b})
\end{aligned}$$

$$= \mathbf{a}^{\mathrm{T}}\sigma^2\mathbf{I}\mathbf{a} \;+\; (\mathbf{a}-\mathbf{b})^{\mathrm{T}}\sum_{i=1}^{\infty} k^{2i}\sigma^2\mathbf{I}(\mathbf{a}-\mathbf{b}) \;+\; 2\mathbf{a}^{\mathrm{T}}\sum_{i=1}^{\infty} k^{i}\sigma^2\rho_i\mathbf{I}\mathbf{L}^{i}\mathbf{I}(\mathbf{a}-\mathbf{b})$$

$$+\; 2\,(\mathbf{a}-\mathbf{b})^{\mathrm{T}}\sum_{i=1}^{\infty}\;\sum_{j=i+1}^{\infty} k^{i+j}\sigma^2\rho_{j-i}\mathbf{I}\mathbf{L}^{j-i}\mathbf{I}\,(\mathbf{a}-\mathbf{b})$$

$$= \sigma^2\,\{\; \mathbf{a}^{\mathrm{T}}\mathbf{I}\mathbf{a} \;+\; (\mathbf{a}-\mathbf{b})^{\mathrm{T}}\mathbf{I}(\mathbf{a}-\mathbf{b})\; k^2/(1-k^2)$$

$$+\; 2\;\mathbf{a}^{\mathrm{T}}\mathbf{I}\;[\;\sum_{i=1}^{\infty} k^{i}\rho_i\mathbf{L}^{i}\;]\;\mathbf{I}(\mathbf{a}-\mathbf{b})$$

$$+\; 2\;(\mathbf{a}-\mathbf{b})^{\mathrm{T}}\mathbf{I}\;[\,\sum_{i=1}^{\infty} k^{2i}\sum_{j=i+1}^{\infty} k^{j-i}\rho_{j-i}\mathbf{L}^{j-i}]\;\mathbf{I}(\mathbf{a}-\mathbf{b})\;\} \tag{8}$$

Because $\mathbf{a}$ and $\mathbf{a}-\mathbf{b}$ are vectors in TIS form, $\mathbf{a}^{\mathrm{T}}\mathbf{I} = \mathbf{a}^{\mathrm{T}}$, $(\mathbf{a}-\mathbf{b})^{\mathrm{T}}\mathbf{I} = (\mathbf{a}-\mathbf{b})^{\mathrm{T}}$, and $\mathbf{I}(\mathbf{a}-\mathbf{b}) = (\mathbf{a}-\mathbf{b})$. It can be shown that the first expression in brackets in (8) is equal to the matrix $\mathbf{Q}$. Only $\mathbf{N}$-1 terms in the sum are necessary: since $\mathbf{L}$ is lower triangular, $\mathbf{L}^{i} = 0$ for $i \geq \mathbf{N}$. The ijth entry of $\sum_{i=1}^{\mathbf{N}-1} k^{i}\rho_i\mathbf{L}^{i}$ is $k^{i-j}\rho_{i-j}$, if $1 \leq j < i \leq \mathbf{N}$, and 0, otherwise. The term in the second set of brackets can be simplified. In the second sum, a change of variables to $n = j-i$ gives $\sum_{n=1}^{\infty} k^{n}\rho_n\mathbf{L}^{n}$, which was just shown to be $\mathbf{Q}$. We can rewrite (8) as:

$$\sigma^2\,\{\; \mathbf{a}^{\mathrm{T}}\mathbf{a} \;+\; (\mathbf{a}-\mathbf{b})^{\mathrm{T}}(\mathbf{a}-\mathbf{b})\; k^2/(1-k^2) \;+\; 2\mathbf{a}^{\mathrm{T}}\mathbf{Q}(\mathbf{a}-\mathbf{b})$$

$$+\; 2(\mathbf{a}-\mathbf{b})^{\mathrm{T}}\sum_{i=1}^{\infty} k^{2i}\;\mathbf{Q}\;(\mathbf{a}-\mathbf{b})\;\}$$

$$= \sigma^2\,\{\; (1-k^2)\mathbf{a}^{\mathrm{T}}\mathbf{a} \;+\; k^2(\mathbf{a}^{\mathrm{T}}\mathbf{a} - 2\mathbf{b}^{\mathrm{T}}\mathbf{a} + \mathbf{b}^{\mathrm{T}}\mathbf{b})$$

$$+\; 2\;[\;(1-k^2)\mathbf{a}^{\mathrm{T}} + k^2(\mathbf{a}-\mathbf{b})^{\mathrm{T}}\;]\;\mathbf{Q}(\mathbf{a}-\mathbf{b})\;\}\;/\;(1-k^2)$$

$$= \sigma^2\,\{\mathbf{a}^{\mathrm{T}}\mathbf{a} + k^2\mathbf{b}^{\mathrm{T}}(\mathbf{b}-2\mathbf{a}) + 2(\mathbf{a} - k^2\mathbf{b})^{\mathrm{T}}\mathbf{Q}(\mathbf{a}-\mathbf{b})\}\;/\;(1-k^2),$$

and the theorem is proved.

Proof of Theorem 2. If $k = 0$, $y_h = \mathbf{a}^{\mathrm{T}}\mathbf{x}_h$, and $y_{h-1} = \mathbf{a}^{\mathrm{T}}\mathbf{x}_{h-1}$. From the stationarity of the covariance,

$$\mathrm{Var}(y_h) = \mathrm{Var}(y_{h-1}) = \mathbf{a}^{\mathrm{T}}\sigma^2\mathbf{I}\mathbf{a} = \sigma^2\mathbf{a}^{\mathrm{T}}\mathbf{a}.$$

The estimator of difference is then $y_h - y_{h-1} = \boldsymbol{a}^T\boldsymbol{x}_h - \boldsymbol{a}^T\boldsymbol{x}_{h-1}$. Its variance is

$$\mathrm{Var}(y_h - y_{h-1}) = 2\sigma^2\boldsymbol{a}^T\boldsymbol{a} - 2\boldsymbol{a}^T\sigma^2\rho_1\boldsymbol{R}L\boldsymbol{R}\boldsymbol{a} = 2\sigma^2\boldsymbol{a}^T(I-\rho_1 L)\boldsymbol{a}.$$

If $0 < k < 1$, $y_h = \boldsymbol{a}^T\boldsymbol{x}_h - k\boldsymbol{b}^T\boldsymbol{x}_{h-1} + ky_{h-1} \equiv W + ky_{h-1}$, (9)

where W is defined as $\boldsymbol{a}^T\boldsymbol{x}_h - k\boldsymbol{b}^T\boldsymbol{x}_{h-1}$.

$$\mathrm{Var}(W) = \boldsymbol{a}^T\sigma^2\boldsymbol{R}\boldsymbol{a} + k^2\boldsymbol{b}^T\sigma^2\boldsymbol{R}\boldsymbol{b} - 2k\boldsymbol{a}^T\sigma^2\rho_1\boldsymbol{R}L\boldsymbol{R}\boldsymbol{b}$$

$$= \sigma^2 \{ \boldsymbol{a}^T\boldsymbol{a} + k^2\boldsymbol{b}^T\boldsymbol{b} - 2k\rho_1\boldsymbol{a}^T L\boldsymbol{b} \} \quad (10)$$

It follows from (9) that

$$\mathrm{Var}(y_h) = \mathrm{Var}(W) + k^2\mathrm{Var}(y_{h-1}) + 2k\mathrm{Cov}(W,y_{h-1})$$

$$= \mathrm{Var}(W) + k^2\mathrm{Var}(y_h) + 2k\mathrm{Cov}(W,y_{h-1}), \text{ and}$$

$$2\mathrm{Cov}(W,y_{h-1}) = (1/k)\ \{ (1-k^2)\mathrm{Var}(y_h) - \mathrm{Var}(W) \}$$

Now we can write $y_h - y_{h-1} = W + ky_{h-1} - y_{h-1} = W - (1-k)y_{h-1}$.

$$\mathrm{Var}(y_h - y_{h-1}) = \mathrm{Var}(W) + (1-k)^2\mathrm{Var}(y_{h-1}) - 2(1-k)\mathrm{Cov}(W,y_{h-1})$$

$$= \mathrm{Var}(W) + (1-k)^2\mathrm{Var}(y_h)$$

$$- (1-k)(1/k)\{ (1-k^2)\mathrm{Var}(y_h) - \mathrm{Var}(W) \}$$

$$= [1 + (1-k)/k]\ \mathrm{Var}(W) + [(1-k)^2 - (1/k)(1-k)(1-k^2)]\ \mathrm{Var}(y_h)$$

$$= (1/k)\ \mathrm{Var}(W) + (1/k)(1-k)^2[k - (1+k)]\ \mathrm{Var}(y_h)$$

$$= \mathrm{Var}(W)/k - (1-k)^2\mathrm{Var}(y_h)/k$$

Substituting from (10) finishes the proof:

$$= \sigma^2\{ \boldsymbol{a}^T\boldsymbol{a} + k^2\boldsymbol{b}^T\boldsymbol{b} - 2k\rho_1\boldsymbol{a}^T L\boldsymbol{b} \}/k - (1-k)^2\mathrm{Var}(y_h)/k$$

Proof of Theorem 3. To find the variance of the sum $S_{h,t} = y_h + y_{h-1} + \dots + y_{h-t}$, express $S_{h,t}$ in the form $\boldsymbol{v}_0^T\boldsymbol{x}_h + \boldsymbol{v}_1^T\boldsymbol{x}_{h-1} + \boldsymbol{v}_2^T\boldsymbol{x}_{h-2} + \dots = \sum_{i=0}^{\infty} \boldsymbol{v}_i^T\boldsymbol{x}_{h-i}$. The vectors $\boldsymbol{v}_0$, $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, ... can be determined by introducing additional y_{h-i} terms one at a time:

$$y_h + y_{h-1} = \boldsymbol{a}^T\boldsymbol{x}_h - k\boldsymbol{b}^T\boldsymbol{x}_{h-1} + ky_{h-1} + y_{h-1}$$

$$= \boldsymbol{a}^T\boldsymbol{x}_h - k\boldsymbol{b}^T\boldsymbol{x}_{h-1} + (1+k)(\boldsymbol{a}^T\boldsymbol{x}_{h-1} - k\boldsymbol{b}^T\boldsymbol{x}_{h-2} + ky_{h-2})$$

$$= \boldsymbol{a}^{\mathrm{T}}(\boldsymbol{x}_h + \boldsymbol{x}_{h-1}) + k(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{x}_{h-1} - (k+k^2)\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}_{h-2} + (k+k^2)y_{h-2}$$

Continuing,

$$\begin{aligned} y_h + y_{h-1} + y_{h-2} &= \boldsymbol{a}^{\mathrm{T}}(\boldsymbol{x}_h + \boldsymbol{x}_{h-1}) + k(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{x}_{h-1} - (k+k^2)\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}_{h-2} \\ &\quad + (1+k+k^2)(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_{h-2} - k\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}_{h-3} + ky_{h-3}) \\ &= \boldsymbol{a}^{\mathrm{T}}(\boldsymbol{x}_h + \boldsymbol{x}_{h-1} + \boldsymbol{x}_{h-2}) + k(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{x}_{h-1} + (k+k^2)(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{x}_{h-2} \\ &\quad - (k+k^2+k^3)\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}_{h-3} + (k+k^2+k^3)y_{h-3} \end{aligned}$$

Including all $t+1$ terms,

$$\begin{aligned} y_h + \dots + y_{h-t} &= \boldsymbol{a}^{\mathrm{T}}(\boldsymbol{x}_h + \boldsymbol{x}_{h-1} + \dots + \boldsymbol{x}_{h-t}) + k(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{x}_{h-1} \\ &\quad + (k+k^2)(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{x}_{h-2} + \dots + (k+k^2+\dots+k^t)(\boldsymbol{a}-\boldsymbol{b})^{T}\boldsymbol{x}_{h-t} \\ &\quad - (k+k^2+\dots+k^{t+1})\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}_{h-t-1} \\ &\quad + (k+k^2+\dots+k^{t+1})y_{h-t-1} \end{aligned} \qquad (11)$$

But according to (7),

$$y_{h-t-1} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_{h-t-1} + (\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\sum_{i=1}^{\infty} k^i \boldsymbol{x}_{h-t-1-i}$$

The third line of (11) becomes

$$\begin{aligned} &(k+k^2+\dots+k^{t+1})[\,(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{x}_{h-t-1} + (\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\sum_{i=1}^{\infty} k^i \boldsymbol{x}_{h-t-1-i}\,] \\ &= (k+k^2+\dots+k^{t+1})\ (\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\sum_{j=t+1}^{\infty} k^{j-t-1}\boldsymbol{x}_{h-j} \end{aligned}$$

Finally we can write

$$\begin{aligned} S_{h,t} = \sum_{i=0}^{\infty} \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{x}_{h-i} &= \boldsymbol{a}^{\mathrm{T}}(\boldsymbol{x}_h + \boldsymbol{x}_{h-1} + \dots + \boldsymbol{x}_{h-t}) + k(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{x}_{h-1} \\ &\quad + (k+k^2)(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{x}_{h-2} + \dots + (k+k^2+\dots+k^t)(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{x}_{h-t} \\ &\quad + (k+k^2+\dots+k^{t+1})(\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\sum_{j=t+1}^{\infty} k^{j-t-1}\boldsymbol{x}_{h-j} \end{aligned} \qquad (12)$$

From (12) it is apparent that: $\boldsymbol{v}_0 = \boldsymbol{a}$; $\boldsymbol{v}_i = \boldsymbol{a} + (k+k^2+\dots+k^i)(\boldsymbol{a}-\boldsymbol{b})$, for $i = 1, 2, \dots, t$; and $\boldsymbol{v}_i = (k+k^2+\dots+k^{t+1})k^{i-t-1}(\boldsymbol{a}-\boldsymbol{b})$, for $i = t+1, t+2, \dots$. Summing powers of k produces the series of $\boldsymbol{v}_i$ defined in Theorem 3.

All that remains is to find the variance of the derived sum. Note that the $\boldsymbol{v}_i$'s, being linear combinations of $\boldsymbol{a}$ and $\boldsymbol{a}-\boldsymbol{b}$, are vectors in TIS form: $\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{I} = \boldsymbol{v}_i^{\mathrm{T}}$ and $\boldsymbol{I}\boldsymbol{v}_i = \boldsymbol{v}_i$.

$$\begin{aligned}
\mathrm{Var}(S_{h,t}) &= \mathrm{Var}\Big(\sum_{i=0}^{\infty} \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{x}_{h-i}\Big) \\
&= \sum_{i=0}^{\infty} \boldsymbol{v}_i^{\mathrm{T}}\mathrm{Var}(\boldsymbol{x}_{h-i})\boldsymbol{v}_i + 2\sum_{0\le i<j}^{\infty}\sum^{\infty} \boldsymbol{v}_i^{\mathrm{T}}\mathrm{Cov}(\boldsymbol{x}_{h-i},\boldsymbol{x}_{h-j})\boldsymbol{v}_j \\
&= \sum_{i=0}^{\infty} \boldsymbol{v}_i^{\mathrm{T}}\sigma^2\boldsymbol{I}\boldsymbol{v}_i + 2\sum_{0\le i<j}^{\infty}\sum^{\infty} \boldsymbol{v}_i^{\mathrm{T}}\sigma^2\rho_{j-i}\boldsymbol{I}\boldsymbol{L}^{j-i}\boldsymbol{I}\boldsymbol{v}_j \\
&= \sigma^2\Big\{\sum_{i=0}^{\infty}\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_i + 2\sum_{i=0}^{\infty}\boldsymbol{v}_i^{\mathrm{T}}\sum_{j=i+1}^{\infty}\rho_{j-i}\boldsymbol{L}^{j-i}\boldsymbol{v}_j\Big\} \\
&= \sigma^2\Big\{\sum_{i=0}^{\infty}\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_i + 2\sum_{i=0}^{\infty}\boldsymbol{v}_i^{\mathrm{T}}\sum_{n=1}^{\infty}\rho_n\boldsymbol{L}^n\boldsymbol{v}_{i+n}\Big\}
\end{aligned}$$

As before, in the final sum, n need only range from 1 to $M-1$, as $\boldsymbol{L}^n = 0$ for $n \ge M$.

Proof of Theorem 4. From our definitions, recall that Y_h is the actual value to be estimated in month h, and β_i is the bias inherent in a respondent's reply in his ith time in sample. The model expressed before Theorem 4 assumed that $\mathrm{E}(x_{h,i}) = Y_h + \beta_i$. In vector form, $\mathrm{E}(\boldsymbol{x}_h) = Y_h\mathrm{diag}(\boldsymbol{I}) + \boldsymbol{\beta}$, for all h, where $\boldsymbol{\beta}$ is the set of bias terms in TIS form. The vector $\mathrm{diag}(\boldsymbol{I})$ is a set of 1's and 0's, with 1's corresponding to months in sample (i.e., for $i \notin T_0$). From (4.3),

$$y_h = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_h + (\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\sum_{i=1}^{\infty}k^i\boldsymbol{x}_{h-i}.$$ Taking expectations,

$$\begin{aligned}
\mathrm{E}(y_h) &= \boldsymbol{a}^{\mathrm{T}}(Y_h\mathrm{diag}(\boldsymbol{I}) + \boldsymbol{\beta}) + (\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\sum_{i=1}^{\infty}k^i(Y_{h-i}\mathrm{diag}(\boldsymbol{I}) + \boldsymbol{\beta}) \\
&= Y_h\,\boldsymbol{a}^{\mathrm{T}}\mathrm{diag}(\boldsymbol{I}) + \boldsymbol{a}^{\mathrm{T}}\boldsymbol{\beta} + (\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\mathrm{diag}(\boldsymbol{I})\sum_{i=1}^{\infty}k^iY_{h-i} \\
&\quad + (\boldsymbol{a}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{\beta}\sum_{i=1}^{\infty}k^i
\end{aligned}$$

$$= Y_h \sum_{i=1}^{m} a_i + \boldsymbol{a}^T\boldsymbol{\beta} + \left(\sum_{i=1}^{m} a_i - \sum_{i=1}^{m} b_i\right) \sum_{i=1}^{\infty} k^i Y_{h-i}$$

$$+ (\boldsymbol{a}-\boldsymbol{b})^T\boldsymbol{\beta}\, k/(1-k)$$

$$= Y_h + \{ (1-k)\boldsymbol{a}^T\boldsymbol{\beta} + k(\boldsymbol{a}-\boldsymbol{b})^T\boldsymbol{\beta} \}/(1-k) + 0$$

$$= Y_h + (\boldsymbol{a} - k\boldsymbol{b})^T\boldsymbol{\beta}/(1-k)$$

We have used the fact that $\sum_{i=1}^{m} a_i = \sum_{i=1}^{m} b_i = 1$.

The second part of the theorem follows because the bias in month h depends on k, $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{\beta}$, but not on h. When evaluating $E(y_h - y_{h-1})$, the bias term cancels.

REFERENCES

Breau, P. and Ernst, L. R. (1983). Alternative Estimators to the Current Composite Estimator, *Proceedings of the Section on Survey Research Methods*, *American Statistical Association*, 397-402.

Gurney, M. and Daly, J. F. (1965). A Multivariate Approach to Estimation in Periodic Sample Surveys, *Proceedings of the Social Statistics Section*, *American Statistical Association*, 242-257.

Hanson, R. H. (1978). *The Current Population Survey: Design and Methodology*, Technical Paper 40, U.S. Bureau of the Census, Washington, D.C.

Huang, E. T. and Ernst, L. R. (1981). Comparison of an Alternative Estimator to the Current Composite Estimator in CPS, *Proceedings of the Section on Survey Research Methods*, *American Statistical Association*, 303-308.

Kumar, S. and Lee, H. (1983). Evaluation of Composite Estimation for the Canadian Labour Force Survey, *Proceedings of the Section on Survey Research Methods*, *American Statistical Association*, 403-408.

Rao, J. N. K. and Graham, J. E. (1964). Rotation Designs For Sampling on Repeated Occasions, *Journal of the American Statistical Association*, 59, 492-509.

Wolter, K. M. (1979). Composite Estimation in Finite Populations, *Journal of the American Statistical Association*, 74, 604-613.

Woodruff, R. S. (1963). The Use of Rotating Samples in the Census Bureau's Monthly Surveys, *Journal of the American Statistical Association*, 58, 454-467.