Survey Methodology, June 1990
Vol. 16, No. 1, pp. 153–163
Statistics Canada

Variance Formulae for Composite Estimators in Rotation Designs

PATRICK J. CANTWELL[1]

ABSTRACT

In many government surveys, respondents are interviewed a set 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. In this paper, simple expressions are derived for the variance of a general class of composite estimators for level, change in level, and average level over time. Considered first are "one-level" rotation designs, where only the current month is referenced in the interview. Results are developed for any sampling pattern of m interviews over a period of M months. Subsequently, "multi-level" plans are addressed. In each month one of p different groups is interviewed. Respondents then answer questions referring to the previous p months. Results from the several sections apply to a wide range of government surveys.

KEY WORDS: Repeated sampling in surveys; Balanced designs; Month-to-month change; Yearly average.

1. INTRODUCTION

Rotation designs of various types are used in many major household surveys. The Current Population Survey (CPS) is conducted by the U.S. Bureau of the Census for the U.S. Bureau of Labor Statistics. Statistics Canada operates the Labour Force Survey (LFS). Both surveys yield estimates of labor force characteristics, including unemployment. In each survey, households are interviewed a number of times before leaving the sample. In the CPS, each household is "rotated in" for interviews in four consecutive months, rotated out of the sample for eight months, and finally back in for four more months. In the LFS, a participating household responds for six consecutive months and does not return.

A survey with a rotation design lies somewhere between a fixed panel survey, where participants remain in sample indefinitely, and a survey using independent samples, where respondents are interviewed once and retired from sample. The total overlap of a fixed panel from one time period to the next can minimize the variance of estimators of change when measurements are positively correlated across periods. Also, certain costs are incurred only the first time a unit is placed in sample. However, response burden on the members of a fixed panel can be excessive. Using a rotation design is an attempt to realize variance or cost reductions without overly burdening sample participants. In the CPS and the LFS, there are sample overlaps of 75% and 83%, respectively, from one month to the next. For more on these topics, see Woodruff (1963), Rao and Graham (1964), or Wolter (1979).

Some estimators used with rotation designs are composite in nature. In order to take advantage of repeated sampling, they combine rotation group estimates obtained for the current month with those from prior months into a final estimator.

[1] Patrick J. Cantwell, Mathematical Statistician, Statistical Research Division, Bureau of the Census, Washington, DC 20233, USA. This paper reports research undertaken by a member of the Census Bureau's staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.

While the variance of composite estimators can be decreased by selecting the combination wisely, calculating this variance may become more complex because of the correlation patterns involved among the repeated groups. For general rotation plans, subject to specific restrictions, simple formulae are presented in this paper for the variance of estimators of level and change. The derivations are applied to an important and quite general class of estimators called the generalized composite estimator (Breau and Ernst 1983).

These formulae can be of use if the correlations between estimates from the same rotation group one or more time periods apart can be estimated and are suffiently large to render composite estimation worthwhile. In continuing government surveys, past sample data will typically enable the estimation of these correlations. Characteristics involving household income and labor force usually exhibit moderately high correlations. For others, such as the incidence of crime, however, correlations across time periods may not be large enough to realize the benefits of composite estimation. Of the surveys mentioned in this paper, only CPS currently uses a composite estimator.

In the developments which follow, two types of surveys are treated separately. In surveys such as CPS and LFS, participants supply information only for the current month. Such surveys are called "one-level" surveys. On the other hand, the U.S. Census Bureau conducts the Survey of Income and Program Participation (SIPP) to acquire data on income level, sources of income, program participation, and other items. During each interview, respondents in the SIPP refer back to the previous four months. A different group is then interviewed the following month. The SIPP design is consequently called "multi-level." The level of a survey was used by Wolter (1979) to indicate the number of periods for which information is solicited in one interview.

Another distinction is made between these two types of surveys. Let the term "design gap" indicate a period of time between interviews which is never referenced in any interview. While the LFS contains no design gaps, CPS includes one of eight months. For the sections pertaining to one-level designs, the results and derivations apply *regardless of the pattern of interviews and design gaps*. Therefore, the formulae are relevant not only to the current design of CPS and LFS, but also to other designs under consideration.

For reasons discussed later, designs gaps are generally not a feature of multi-level rotation plans in practice. The SIPP is no exception. Accordingly, the multi-level plans addressed in this paper do not include design gaps.

One-level designs are treated in Sections 2 and 3. In Section 2, the generalized composite estimator is defined. Notation, definitions and covariance assumptions are introduced. The main results – Theorems 1 through 3 – are given in Section 3. Variances of estimators of level and change in level are stated. The formulae are determined for single time periods (such as months) and combinations (such as quarters or years). They apply to one-level designs with any pattern of interviews and design gaps. When seeking the optimal rotation plan and composite estimator, the user must determine how best to combine variance reductions/increases for the resulting estimators of level, "month-to-month" change, and average over many periods.

In Section 4, these results are extended from one-level to certain multi-level designs, which include the SIPP. Subject to minor restrictions – in particular, the exclusion of design gaps in the sampling scheme – theorems similar to those in Section 3 are stated. Because the derivations are analogous to those for one-level plans, the results are not proved.

2. ONE-LEVEL DESIGNS: NOTATION AND DEFINITIONS

Although rotation schemes can assume infinitely many forms, the discussion in Sections 2 and 3 is restricted to one type. At each period of time, a new rotation group enters the sample, and follows the same pattern of interviews and design gaps 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. This design is called 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.

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 plan. Before July 1953, however, CPS used an unbalanced design 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 CPS design before 1953 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, the word "month" refers to the period of time in which interviews are done, partly for brevity, but also because most government surveys use the month to divide the life of the survey. However, the results in this section and the next apply to any period of time, provided the rotation plan is balanced and one-level.

Some notation and vector definitions are now introduced. Suppose that every rotation group is in sample for a total of m interviews over a period of M months. That is, it is out of sample for $M - m$ months after first entering and before exiting. The balanced design ensures that m groups are in sample during any month.

The set T_0 is defined 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, and stopping at month M. Because the design is balanced, the composition of T_0 does not depend on which group is selected. Note that, if respondents are interviewed in m consecutive months, *i.e.*, there are no design gaps, then m and M are the same, and T_0 is empty.

Next, given a set of m values $w_1, \ldots, w_m$, it is possible to define the $M \times 1$ vector w as follows. Define the ith component of w to be 0 if $i \in T_0$. This step fills $M - m$ positions in w. Then the values $w_1, \ldots, w_m$ are inserted in order into the remaining m components, starting with the first. The resulting w is called a vector "in design form." For example, in a 4-8-4 rotation plan, $T_0 = \{5, 6, \ldots, 12\}$, and $w^T = (w_1, w_2, w_3, w_4, 0, 0, 0, 0, 0, 0, 0, 0, w_5, w_6, w_7, w_8)$.

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 $i \neq j$. It is clear that $\boldsymbol{R}$ is a diagonal matrix where diag($\boldsymbol{R}$) is a set of 1's "in design form," $\boldsymbol{R}_{11}$ and $\boldsymbol{R}_{MM}$ are 1, and $\sum_{i=1}^{M} \boldsymbol{R}_{ii} = m$.

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

Let $\boldsymbol{L}$ be the $M \times M$ matrix with 1's on the subdiagonal, and 0's elsewhere. Formally, $L_{ij} = 1$, if $i - j = 1$, and 0, otherwise. For any $M \times 1$ vector written as $w^T = (w_1, \ldots, w_M)$, the product $\boldsymbol{L}w$ becomes $(0, w_1, w_2, \ldots, w_{M-1})^T$, and $w^T\boldsymbol{L}$ is $(w_2, w_3, \ldots, w_M, 0)$.

Turning to the data, let $x_{h,i}$ denote the estimate of "monthly" level for some characteristic to be measured from the rotation group which is in sample for the ith time in month h, where $i = 1, \ldots, m$. Breau and Ernst (1983) defined the generalized composite estimator (GCE) of level recursively as follows. For monthly level, let:

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

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

The GCE is more restrictive than a general linear estimator which combines $x_{h,i}$ values from the current period with those from many prior months (see Gurney and Daly 1965). 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 to be stored. Although y_h incorporates earlier data, it is summarized through y_{h-1}.

To facilitate variance computations, (1) is expressed in vector form. Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be $M \times 1$ vectors in design form comprising, respectively, the sets of constants $a_1, \ldots, a_m$ and $b_1, \ldots, b_m$. Similarly, for any h, the observations $x_{h,1}, \ldots, x_{h,m}$ make up $\boldsymbol{x}_h$, also an $M \times 1$ vector in design form. Then

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

The data are assumed to exhibit a stationary covariance structure:

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

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

(iii) $\mathrm{Cov}(x_{h,i}, x_{s,j}) = \rho_{|h-s|}\sigma^2$, 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)

From the first two parts of (2), it is clear that $\mathrm{Var}(\boldsymbol{x}_h) = \sigma^2 \boldsymbol{R}$, for all h. Part three implies that $\mathrm{Cov}(\boldsymbol{x}_h, \boldsymbol{x}_{h-1}) = \sigma^2 \rho_1 \boldsymbol{RLR}$. This follows because (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, and (b) premultiplying (postmultiplying) by $\boldsymbol{R}$ inserts 0's corresponding to 0's in $\boldsymbol{x}_h$ ($\boldsymbol{x}_{h-1}$) (months not in sample).

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

$$\mathrm{Cov}(\boldsymbol{x}_h, \boldsymbol{x}_{h-r}) = \sigma^2 \rho_r \boldsymbol{RL}^r\boldsymbol{R}, \text{ for } r = 0, 1, 2, \ldots, \text{ and all } h. \tag{3}$$

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

For the theorems which follow, 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.

3. ONE-LEVEL DESIGNS: THEOREMS AND PROOFS

Three theorems are now stated and proved.

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

$$\mathrm{Var}(y_h) = \sigma^2\{a^Ta + k^2b^T(b - 2a) + 2(a - k^2b)^TQ(a - b)\}/(1 - k^2). \tag{4}$$

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

Proof of theorem 1. Substitution into (1a) recursively leads to

$$y_h = a^Tx_h + (a - b)^T\sum_{i=1}^{\infty} k^ix_{h-i}. \tag{5}$$

From (3), the variance of this sum is

$$\begin{aligned}\mathrm{Var}(y_h) &= a^T\sigma^2Ra + (a - b)^T\sum_{i=1}^{\infty} k^{2i}\sigma^2R(a - b)\\ &+ 2a^T\sum_{i=1}^{\infty} k^i\sigma^2\rho_iRL^iR(a - b)\\ &+ 2(a - b)^T\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} k^{i+j}\sigma^2\rho_{j-i}RL^{j-i}R(a - b)\\ &= \sigma^2\Bigg\{a^TRa + (a - b)^TR(a - b)k^2/(1 - k^2)\\ &+ 2a^TR\left(\sum_{i=1}^{\infty} k^i\rho_iL^i\right)R(a - b)\\ &+ 2(a - b)^TR\left(\sum_{i=1}^{\infty} k^{2i}\left[\sum_{j=i+1}^{\infty} k^{j-i}\rho_{j-i}L^{j-i}\right]\right)R(a - b)\Bigg\}.\end{aligned} \tag{6}$$

Because a and $a - b$ are vectors in design form, $a^TR = a^T$, $(a - b)^TR = (a - b)^T$, and $R(a - b) = (a - b)$. The sum $\sum_{i=1}^{\infty} k^i\rho_iL^i$ is seen to be the matrix Q: its ijth entry is $k^{i-j}\rho_{i-j}$, if $1 \le j < i \le M$, and 0, otherwise. A change of variables will show that the sum in brackets is also Q. Expression (6) can be rewritten as:

$$\begin{aligned}&\sigma^2\{a^Ta + (a - b)^T(a - b)k^2/(1 - k^2) + 2a^TQ(a - b)\\ &+ 2(a - b)^TQ(a - b)k^2(1 - k^2)\}.\end{aligned}$$

Simple rearrangement of these terms produces the result in (4).

Theorem 2. Let $y_h - y_{h-1}$ be the GCE estimator of "month-to-month" change. Then $\mathrm{Var}(y_h - y_{h-1})$ is

(i) $2\sigma^2 a^T(I - \rho_1 L)a$, if $k = 0$, and

(ii) $\sigma^2(a^Ta + k^2b^Tb - 2k\rho_1 a^TLb)/k - (1-k)^2\mathrm{Var}(y_h)/k$, if $0 < k < 1$.

Proof of theorem 2:

(i) If $k = 0$, $y_h = a^Tx_h$. From (3), the variance of $a^Tx_h - a^Tx_{h-1}$ is

$$2a^T\sigma^2 Ra - 2a^T\sigma^2\rho_1 RLRa = 2\sigma^2 a^T(I - \rho_1 L)a.$$

(ii) If $0 < k < 1$, define W_h as $a^Tx_h - kb^Tx_{h-1}$. From prior results, it is quickly seen that

$$\mathrm{Var}(W_h) = \sigma^2\{a^Ta + k^2b^Tb - 2k\rho_1 a^TLb\}. \tag{7}$$

From (1a), $y_h = W_h + ky_{h-1}$. Then

$$\mathrm{Var}(y_h) = \mathrm{Var}(W_h) + k^2\mathrm{Var}(y_{h-1}) + 2k\mathrm{Cov}(W_h, y_{h-1}); \tag{8}$$

the covariance term can be isolated for later use. Finally, $y_h - y_{h-1} = W_h - (1-k)y_{h-1}$. When computing the variance of this difference, substitution from (8) and (7) produces the desired result.

Often of primary importance are the average level over a certain length of time (*e.g.*, a quarter or a year), the difference in these averages from one "year" to the next, or the difference in "monthly" level for two months a year apart. Denote by $S_{h,t}$ the sum of the GCE's for the last t months:

$$S_{h,t} = y_h + y_{h-1} + \ldots + y_{h-t+1}, \quad t \geq 1. \tag{9}$$

Commonly used values of t include three, four and twelve. It is left to the reader to divide $S_{h,t}$ by t if an average desired rather than a sum.

Theorem 3:

(a) The expressions $S_{h,t}$, $S_{h,t} - S_{h-t,t}$, and $y_h - y_{h-t}$ can be written as $\sum_{i=0}^{\infty} v_i^T x_{h-1}$, where

(i) for $S_{h,t}$, $v_i =$:

$a + [(k - k^{i+1})/(1-k)](a - b)$, for $i = 0, 1, \ldots, t-1$,

$[k^{i-t}(k - k^{t+1})/(1-k)](a - b)$, for $i = t, t+1, t+2, \ldots$;

(ii) for $S_{h,t} - S_{h-t,t}$, $v_i =$:

$a + [(k - k^{i+1})/(1-k)](a - b)$, for $i = 0, 1, \ldots, t-1$,

$[(2k^{i-t+1} - k - k^{i+1})/(1-k)](a - b) - a$, for $i = t, t+1, \ldots, 2t-1$,

$-[k^{i-2t+1}(1-k^t)^2/(1-k)](a - b)$, for $i = 2t, 2t+1, \ldots$;

(iii) for $y_h - y_{h-t}$, $v_0 = a$, $v_t = k^t(a - b) - a$, and $v_i =$:

$k^i(a - b)$, for $i = 1, 2, \ldots, t-1$,

$-k^{i-t}(1-k^t)(a - b)$, for $i = t+1, t+2, \ldots$; (10)

(b) For the sets of vectors v_0, v_1, v_2, ... defined in (a),

$$\mathrm{Var}\left(\sum_{i=0}^{\infty} v_i^T x_{h-i}\right) = \sigma^2\left\{\sum_{i=0}^{\infty} v_i^T v_i + 2\sum_{i=0}^{\infty} v_i^T \sum_{n=1}^{M-1} \rho_n L^n v_{i+n}\right\}; \tag{11}$$

the sums in (11) converge.

Proof of theorem 3. For (a), successive inclusion of terms y_h through y_{h-t+1}, and the application of (5) to y_{h-t} yield

$$\begin{aligned} S_{h,t} = {} & a^T(x_h + x_{h-1} + \ldots + x_{h-t+1}) + k(a - b)^T x_{h-1} \\ & + (k + k^2)(a - b)^T x_{h-2} + \ldots \\ & + (k + k^2 + \ldots + k^{t-1})(a - b)^T x_{h-t+1} \\ & + (k + k^2 + \ldots + k^t)(a - b)^T \sum_{j=t}^{\infty} k^{j-t} x_{h-j}. \end{aligned} \tag{12}$$

The three sets of v_i's are then determined from (12) and (5).

The proof of (b) is similar to that of Theorem 1, once it is seen that the v_i's defined in (a), being linear combinationsi of a and $a - b$, are in design form. To prove convergence, note that, for all three sets of v_i's in (a), v_i is proportional to $k^i(a - b)$ for i sufficiently large. There exists a constant $\lambda > 0$ such that, for $i \geq 2t$ and each component j, $| v_{ij} | \leq k^i\lambda$. Recalling that $| \rho_i | \leq 1$, and that each row of L^n has at most one nonzero element (equal to 1), the finite sum in (11) is seen to be an $M \times 1$ vector, each of whose components is bounded above in absolute value by $k^i(M - 1)\lambda$. Convergence of the double summation then follows geometrically in k^{2i}.

4. EXTENSION TO MULTI-LEVEL DESIGNS

Although the results developed in Sections 2 and 3 apply to all balanced one-level rotation plans, it was observed that many surveys operate under multi-level designs. For example, in the Survey of Income and Program Participation (SIPP), one of four rotation groups is interviewed each month, and respondents supply information about the previous four months. Although the design is always subject to change, the first rotation group is interviewed in February, June, October, February, *etc.*, for a total of eight interviews. A second group is interviewed in March, July, *etc.* The remaining two groups follow the same sampling pattern, beginning in April and May. A SIPP panel is the set of four concurrent rotation groups covering about two and one-half years. Each year, a new panel is introduced. For example, the 1986 panel ran from 1986 through 1988, while the 1987 panel spanned 1987-89. Data from different panels are not combined, even though they may cover a common year or two. For further details on the SIPP design, see Nelson, McMillen and Kasprzyk (1984).

When one-level designs were addressed, a rotation group was allowed to assume any pattern of interviews and design gaps – intermediate months which are never referenced – provided the design was balanced. In a multi-level plan, however, design gaps can create problems with recall. Looking back several months, a respondent may find it difficult to assign an event to the correct period of time. Design gaps can only add to the confusion. For this reason, and because multi-level surveys which incorporate design gaps are rare in practice, this section considers only designs where (i) the sample comprises p rotation groups, (ii) groups are interviewed every pth "month" in an alternating sequence, and (iii) the period of reference is the previous p months.

Many multi-level surveys, for example, the National Crime Survey, sponsored by the U.S. Bureau of Justice Statistics, have a more intricate rotational pattern than that covered here. As expected, variance formulae applied to composite estimators would tend to be more complex.

The interview of a rotation group will refer to the collective gathering of information in the assigned month from all sample units in that group. For a particular characteristic which is to be estimated, let $x_{h,i}$ denote the estimate of "monthly" level for month h from the group which is interviewed in month $h + i$, where $i = 1, \ldots, p$. The index i measures recall time – the amount of time between the month of reference and the interview. Table 1 depicts the estimates $x_{h,i}$ for a four-group four-level design. In the diagram solid lines separate estimates which are obtained in different interviews. These boundaries between the reference periods of consecutive interviews are called "seams" in the SIPP.

Table 1
Layout of Estimates in a Longitudinal 4-Level Design

MONTH ↓	ROTATION GROUPS → 1	2	3	4
1	$x_{1,4}$			
2	$x_{2,3}$	$x_{2,4}$		
3	$x_{3,2}$	$x_{3,3}$	$x_{3,4}$	
4	$x_{4,1}$	$x_{4,2}$	$x_{4,3}$	$x_{4,4}$
5	$x_{5,4}$	$x_{5,1}$	$x_{5,2}$	$x_{5,3}$
6	$x_{6,3}$	$x_{6,4}$	$x_{6,1}$	$x_{6,2}$
7	$x_{7,2}$	$x_{7,3}$	$x_{7,4}$	$x_{7,1}$
8	$x_{8,1}$	$x_{8,2}$	$x_{8,3}$	$x_{8,4}$
9	$x_{9,4}$	$x_{9,1}$	$x_{9,2}$	$x_{9,3}$
10	$x_{10,3}$	$x_{10,4}$	$x_{10,1}$	$x_{10,2}$
11	$x_{11,2}$	$x_{11,3}$	$x_{11,4}$	$x_{11,1}$
12	$x_{12,1}$	$x_{12,2}$	$x_{12,3}$	$x_{12,4}$
13	$x_{13,4}$	$x_{13,1}$	$x_{13,2}$	$x_{13,3}$
14	$x_{14,3}$	$x_{14,4}$	$x_{14,1}$	$x_{14,2}$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

Note: $x_{h,i}$ denotes the estimate of "monthly" level for month h from the group which is interviewed in month $h + i$. Interviewing begins in month 5. Solid horizontal lines (seams) separate estimates which are obtained in different interviews.

Let the vector x_h, defined as $(x_{h,1}, x_{h,2}, \ldots, x_{h,p})^T$, comprise the p estimates for month h obtained from the p groups *in different interviews*. Note that $x_{h,p}, x_{h+1,p-1}, \ldots, x_{h+p-1,1}$ are estimates for p different months obtained *from one group in a single interview* (in month $h + p$).

As in Sections 2 and 3, the generalized composite estimator for monthly level is defined as

$$y_h = \sum a_i x_{h,i} - k \sum b_i x_{h-1,i} + k y_{h-1}, \tag{13}$$

where the summations now range from 1 to p. Defining $\boldsymbol{a}$ and $\boldsymbol{b}$ as $(a_1, \ldots, a_p)^T$ and $(b_1, \ldots, b_p)^T$, respectively, the GCE can again be written as

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

The covariance structure of the monthly rotation group estimates is assumed to be stationary in time. Under this multi-level design, however, the length of time between the target month h and the corresponding interview in month $h + i$ may affect the variability of the response, $x_{h,i}$. For $i = 1, \ldots, p$, let d_i^2 represent the response variability as a function of the amount of time between the reference month and the interview. The following covariance structure is postulated:

(i) $\mathrm{Var}(x_{h,i}) = d_i^2 \sigma^2$ for all h and i, where $d_i > 0$;

(ii) $\mathrm{Cov}(x_{h,i}, x_{h,j}) = 0$ for $i \neq j$; and

(iii) For $r \geq 0$: $\mathrm{Cov}(x_{h,i}, x_{h-r,j}) = \rho_{r,i} d_i d_j \sigma^2$, if the two x's refer to the same group r months apart; or 0, otherwise. Take $\rho_{0,i}$ to be 1 for all i. (14)

It may well be that $d_1 \leq d_2 \leq \ldots \leq d_p$, if response variability increases with recall time. The subscript r in the correlation coefficient $\rho_{r,i}$ is the amount of time between the months referenced by estimates $x_{h,i}$ and $x_{h-r,j}$. The subscript i indicates that the estimate for month h is obtained from an interview i months later. For specified values of h, r and i, there is only one value j, $1 \leq j \leq p$, for which the estimates $x_{h,i}$ and $x_{h-r,j}$ refer to the same panel and $\mathrm{Cov}(x_{h,i}, x_{h-r,j})$ is nonzero. (This value is $j = \mathrm{mod}_p(i + r - 1) + 1$, where $\mathrm{mod}_p(n)$ is the value of the integer n, modulo p.) Otherwise, the covariance is 0. In some cases, it may be appropriate to replace $\rho_{r,1}, \ldots, \rho_{r,p}$ with a commom ρ_r.

No assumptions are made about bias. In addition to the effect of recall on variances of group estimates as postulated in (14), a bias related to recall time might also be incurred. Another source - time-in-sample bias - can result according to the number of times a respondent has been interviewed (Bailar 1975). Although these biases need not be measured to derive the variance formulae given in this section, they might constitute a nontrivial component of mean squared error.

Define the $p \times p$ matrices $\boldsymbol{D}$, $\boldsymbol{P}_r$ and $\boldsymbol{J}$ as follows. Let $\boldsymbol{D}$ and $\boldsymbol{P}_r$, for $r \geq 0$, be diagonal matrices with $d_1, \ldots, d_p$ and $\rho_{r,1}, \ldots, \rho_{r,p}$, respectively, along the diagonal. Define $\boldsymbol{J}$ as: $J_{i,i+1} = 1$ for $i = 1, 2, \ldots, p - 1$; $J_{p1} = 1$; and $J_{ij} = 0$, otherwise. The powers of $\boldsymbol{J}$ form a cycle with $\boldsymbol{J}^p = \boldsymbol{I}$, where $\boldsymbol{I}$ is the $p \times p$ identity matrix. An argument similar to that in Section 2 leads to $\mathrm{Var}(\boldsymbol{x}_h) = \sigma^2 \boldsymbol{D}^2$ for all h, and, in general, $\mathrm{Cov}(\boldsymbol{x}_h, \boldsymbol{x}_{h-r}) = \sigma^2 \boldsymbol{D} \boldsymbol{P}_r \boldsymbol{J}^r \boldsymbol{D}$, for $r = 0, 1, 2, \ldots$, and all h.

Finally, define the matrix Z as $\sum_{n=1}^{\infty} k^n P_n J^n$. For general p, i, and j, it can be shown that the ijth cell Z_{ij} is an infinite sum of terms:

$$Z_{ij} = \sum_{m=0}^{\infty} k^u \rho_{u,i}, \quad \text{where} \quad u = pm + 1 + \text{mod}_p(p - i + j - 1).$$

Because the ρ values represent correlation coefficients, it follows easily that Z is finite.

Analogous to theorems 1, 2, and 3 proven earlier are theorems 4, 5, and 6 presented below. The former three allow any pattern of design gaps, but apply only to one-level designs. Theorems 4, 5, and 6 do not permit designs gaps.

The proofs of the theorems are similar to those in Section 3 and are not repeated. All results apply to the limiting case where rotation groups have been in sample long enough to eliminate the effect of phasing in the sample. If the $\rho_{r,i}$'s decrease rapidly with r, or if k is relatively small, the "steady-state" arrives within a couple of interviews.

Theorem 4. If the GCE of level is defined as in (13), and the covariance structure of (14) holds, then

$$\begin{aligned}\text{Var}(y_h) = \sigma^2\{&a^T D^2 a + k^2 b^T D^2 (b - 2a) \\ &+ 2(a - k^2 b)^T DZD(a - b)\}/(1 - k^2).\end{aligned}$$

Theorem 5. Let $y_h - y_{h-1}$ be the GCE estimator of "month-to-month" change. Then $\text{Var}(y_h - y_{h-1})$ is

(i) $2\sigma^2 a^T D(I - P_1 J)Da$, if $k = 0$, and

(ii) $\sigma^2(a^T D^2 a + k^2 b^T D^2 b - 2ka^T DP_1 JDb)/k - (1 - k)^2 \text{Var}(y_h)/k$, if $0 < k < 1$.

Theorem 6. Define $S_{h,t}$ as in (9), the sum of the GCE's for the last t periods. Then $S_{h,t}$, $S_{h,t} - S_{h-t,t}$, and $y_h - y_{h-t}$ can again be written as $\sum_{i=0}^{\infty} v_i^T x_{h-i}$, where the vectors v_0, v_1, v_2, ... are found in (10). For these sets of vectors,

$$\text{Var}\left(\sum_{i=0}^{\infty} v_i^T x_{h-i}\right) = \sigma^2\left\{\sum_{i=0}^{\infty} v_i^T D^2 v_i + 2\sum_{i=0}^{\infty} v_i^T \sum_{n=1}^{\infty} DP_n J^n D\, v_{i+n}\right\}; \tag{16}$$

the sums in (16) converge.

ACKNOWLEDGMENTS

The author wishes to thank Lynn Weidman and Larry Ernst for checking the text and proofs. Lynn graciously read through several drafts, and offered many helpful suggestions to improve the presentation of the paper. The comments and suggestions of an associate editor and three referees are also noted and greatly appreciated.

REFERENCES

BAILAR, B. A. (1975). The effects of rotation group bias on estimates from panel surveys. *Journal of the American Statistical Association*, 70, 23-30.

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.

NELSON, D., McMILLEN, D., and KASPRZYK, D. (1984). An Overview of the Survey of Income and Program Participation. SIPP Working Paper Series, No. 8401, U.S. Bureau of the Census, Washington, D.C.

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.