GENERALIZED COMPOSITE ESTIMATION IN REPEATED SAMPLING: DESIGNS AND VARIANCE FORMULAE

Patrick Cantwell, U.S. Bureau of the Census*

1. Introduction

The Current Population Survey and the Survey of Income and Program Participation of the U.S. Bureau of the Census are just two examples of the use of rotation plans or repeated sampling. In each case, households are interviewed a number of times before leaving the sample. One reason for using a rotation design is to decrease certain costs which arise only the first time a household appears in the sample. These costs may include listing an area, locating addresses, or explaining details of the survey and questions to participants.

Of greater interest to us is the smaller variance for estimates of change obtained when measurements within sample groups are positively correlated from one time period to the next. This improvement depends on the magnitude of the correlations and the amount of overlap in the sample design. The CPS enjoys a sample overlap of 75% from one month to the next. Estimates of month-to-month or year-to-year change can be improved by selecting the proper plan and estimator. Respondent burden, the chief drawback of repeated sampling, can be lessened by manipulating the sequence of periods when respondents are in and out of sample.

In order to compare the many rotation plans and estimators available, one might investigate the appropriate variances. Deriving such variances is not conceptually difficult, but can be quite tedious. Some 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 from the same subgroup in the sample obtained at different times are combined into a final estimate. While the variance can be decreased by selecting the combination wisely, calculating the variance may become more complex because of the correlation patterns involved among the repeated groups.

In this paper, we present simple formulae for the variances of an important and quite general class of linear estimators called the general composite estimator (Breau and Ernst 1983). The formulae are given for estimators of level and change in level, and apply to single time periods (such as months) and combinations (such as quarters or years). We classify rotation plans according to the period of reference--single or multiple time periods--and the method of replacing sample groups. Results are obtained for any survey satisfying the restrictions of either of two general classes of rotation designs.

In Sections 2 and 3 we classify various types of rotation plans, specifying those which are covered by these results. Balanced one-level rotation plans (BOLRPs) are defined and discussed in Section 2, multi-level plans (MLRPs) in Section 3. Section 4 contains a brief discussion of generalized composite estimation.

Our main results are covered in Sections 5 and 6. For each of the two types of rotation plans under consideration, we lay out the covariance structure and state several theorems giving variance formulae for generalized composite estimators. The proofs are omitted, but are found in Cantwell (1988, 1989). Finally in Section 7 we briefly mention several topics of further interest, from the usefulness of composite estimation in certain surveys, to the complexities of rotation designs not covered in this paper.

2. Balanced One-Level Rotation Plans

Although the rotational designs of government surveys differ, many share certain features which help to classify them. In the Current Population Survey (CPS), households are interviewed for four months, then leave the sample for eight months, and finally return for four more months. Participants in the Labour Force Survey (LFS) of Statistics Canada respond for six consecutive months and do not return (Kumar and Lee 1983). In either survey the period of reference for any interview is the current month only, whether or not respondents were in sample the previous month. Wolter (1979) uses the terms "<u>one-level</u>" and "two-level" to indicate the number of periods for which information is solicited in one interview. CPS and LFS are prime examples of one-level designs.

Other surveys operate under different designs. In the Survey of Income and Program Participation (SIPP), the sample is split into four groups. Each group is interviewed every fourth month for eight interviews, and respondents supply information on the previous four months. The National Crime Survey (NCS) alternately interviews one of six panels in any month, each panel reporting crimes which occurred in the prior six months. After three years in sample, households are retired.

The rotation plans used in the SIPP and NCS can be labeled four- and six-level, respectively. More generally we call these <u>multi-level</u> designs, signifying that responses are obtained for more than one period of time in a single interview. As will be seen, the covariance structure for one-level plans differs from that for multi-level plans. It follows that variance formulae will be derived separately for the two situations.

Throughout this paper, we will use "<u>month</u>" to denote the period of time (i) in which interviews are done, and (ii) about which information is obtained. This is the period used in CPS, LFS, NCS, and the SIPP. However, our theorems and results extend to any period of time. When data are compiled and/or released to the public, "months" are often combined into quarters or years.

Although rotation schemes can assume many forms, we restrict this investigation to two classes. For the one-level case, the term <u>rotation group</u> refers to all units which enter the sample in a particular "month." We observe that CPS contains "design gaps," intermediate periods which are never referenced. A respondent in sample never discloses information pertaining to the "idle" eight months between the fourth and fifth interviews. To be as general as possible, we will allow any rotation plan which satisfies the following: in each "month," a new rotation group enters the sample, and follows the same pattern of months in and out of sample as every preceding group. This design will be called "balanced" because the number of rotation groups in sample in any month (eight in CPS) is equal to the total number of months any one group is included in the sample. The first set of designs covered by the results in this paper is the class of balanced one-level rotation plans (BOLRP).

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 an unbalanced design where five rotation groups entered, one each in consecutive months. In the sixth month, no new group entered. Each group exited after six months in sample, and the process continued in the same manner.

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 during the life of the survey, the variance formula of the estimator also changes. Under a balanced design with stationary covariance structure, general derivations are possible.

3. Multi-Level Rotation Plans

In multi-level designs, we call the entire set of people who are interviewed in a given month a panel. This terminology is consistent with NCS, which employs six panels. Unfortunately, the SIPP uses the label rotation group here, and calls the collection of these groups a "panel." To avoid confusion, we proceed calling these groups panels when referring to multi-level designs.

When considering one-level designs, we allow a rotation group to assume any sequence of inclusions and exclusions from the sample, as long as the design is balanced. For a multi-level plan, however, because of recall bias, it makes little sense to allow design gaps. Consider an NCS panel which is interviewed in May and November. In November each respondent is asked about events or situations in May, June, July, August, September and October. Confusion may arise over which events occurred in April, and which in May. Yet the previous interview in May, referencing November through April, can help place these events in the proper month. NCS goes so far as to conduct a preliminary "bounding interview" for those entering the sample. The responses from this first meeting are not included in NCS estimates, but help to eliminate events which occurred before the reference period of the survey.

Suppose instead that a panel is interviewed every eight months and asked about the previous six, leaving gaps of two months after each interview. If a respondent confuses events which occurred six or seven months ago, the interviewer has no record to help determine the proper month. For this reason, and because we are not familiar with any multi-level surveys which incorporate design gaps, we will restrict our efforts to multi-level rotation plans where (i) the sample is made of p panels, (ii) each panel is interviewed every pth "month," and (iii) the period of reference is the previous p months.

At this point, the question of sample replacement must be addressed. In any SIPP sample, each of the four panels (i.e., SIPP "rotation groups") are interviewed every fourth month through eight interviews, a period of almost three years. We might call such a design longitudinal, in that initial respondents remain in sample for many interviews, and no attempt is made to balance any month's time-in-sample.

The design used in NCS, on the other hand, might be labeled "rotationally balanced." Each of the six panels is interviewed seven times, including the bounding interviewed. Within any panel there are seven rotation groups (although the group in sample for the first time is not used in the estimation process), making a total of 42 panel-rotation groups in sample at any time. After each interview, the rotation group which has just been interviewed for the seventh time leaves the sample, and a new one enters, so that data from any interview is balanced with respect to time-in-sample. The Consumer Expenditure Quarterly Survey uses a similar balanced design--each of three panels consists of five rotation groups (one is in sample only for "bounding" purposes).

Rotationally balanced multi-level designs are more involved. For any month estimates are available (eventually) from each rotation group in each panel. Realistic assumptions regarding the covariance structure and the various ways of combining these estimates grow more complex. In this paper, along with BOLRPs, we consider only "longitudinal" multi-level rotation plans (LMLRP). Effects of time-in-sample, including bias, will not be considered. We leave for further research (rotationally) balanced multi-level rotation plans (BMLRP). This is not to imply that a BMLRP will not supply longitudinal information, only that the models we consider here are simpler.

4. Generalized Composite Estimation

The interview of a rotation group (panel) in a BOLRP (LMLRP) will refer to the collective gathering of information in the proper month from all sample units in that rotation group (panel). Here we introduce some notation. Consider first any BOLRP. 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. The values m and M

are the same if there are no design gaps, as in the LFS. Because the rotation design is balanced, m groups are in sample during any month. For a particular characteristic which is to be estimated, 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, where $i = 1, 2, \ldots, m$.

Under a LMLRP, p is the number of panels in sample, and the length of the reference period for any interview. In this situation, let $x_{h,i}$ denote the estimate of monthly level for month h from the panel which is interviewed in month $h+i$, where $i = 1, 2, \ldots, p$. It is clear that i measures the recall time, i.e., the amount of time between the interview and the month of reference. In the appendix is a chart depicting the estimates $x_{h,i}$ for a four-panel four-level design. In the diagram solid horizontal lines separate estimates which are obtained in different interviews. The SIPP refers to these boundaries between the reference periods of consecutive interviews as "seams."

Using this notation, $x_{h,1}, x_{h,2}, \ldots, x_{h,p}$ represent p estimates for month h obtained from the p panels in different interviews. On the other hand, $x_{h,p}, x_{h+1,p-1}, \ldots, x_{h+p-1,1}$ denote estimates for p different months obtained from one panel in a single interview.

For the two designs, we treat only the generalized composite estimator (GCE), as defined by Breau and Ernst (1983). In the following sums, i ranges from 1 to m when we are dealing with a BOLRP, but from 1 to p under a LMLRP. For monthly level:

$$y_h = \sum_i a_i x_{h,i} - k \sum_i b_i x_{h-1,i} + k y_{h-1}, \qquad (1)$$

where k, the a_i's and the b_i's may take any values subject to $0 \leq k < 1$, $\Sigma a_i = 1$, and $\Sigma b_i = 1$.

The composite and AK composite estimators used in CPS (Huang and Ernst 1981) are special cases of the GCE. Gurney and Daley (1965) examined a general linear estimator in the case of a one-level design which combines $x_{n,i}$ values from the current and many prior months and produces significant improvement over noncomposite estimators. However, the GCE has been shown (Breau and Ernst 1983) to perform almost as well. It has the advantage that data from only 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}.

At this time, neither the SIPP nor NCS uses a composite estimation. Each uses a simple average of the estimators (with appropriate adjustments) from the several panels for any given period of time.

5. Covariance Structure and Theorems Under a BOLRP

The two classes of rotation plans are treated separately here. Although the interview design and assumptions about the covariance differ between the two, the resulting theorems appear interestingly similar. To find expressions for the variance of the GCE under a BOLRP, 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 are estimates obtained from 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 given 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, and stopping at M. Because the rotation plan is balanced, the composition of T_0 does not depend on which group is selected.

Next we create the $M \times 1$ vector $\boldsymbol{a}$. Define the ith component of $\boldsymbol{a}$ to be 0 if $i \in T_0$. This step fills $M-m$ 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 $\boldsymbol{a}^T = (a_1, a_2, a_3, a_4, 0, 0, 0, 0, 0, 0, 0, 0, a_5, a_6, a_7, a_8)$. The $M \times 1$ vectors $\boldsymbol{b}$ and $\boldsymbol{x}_h$ are formed analogously in TIS form, the latter from the estimates $x_{h,1}, x_{h,2}, \ldots, x_{h,m}$ from month h. Line (1) can now be written in vector form:

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

Let 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. Define the $M \times M$ matrix $\boldsymbol{Q}$ by: $Q_{ij} = k^{i-j}\rho_{i-j}$, if $1 \leq j < i \leq M$, and 0, otherwise. (It is easily seen that $\boldsymbol{Q}$ can be expressed equivalently as $\sum_{n=1}^{\infty} k^n \rho_n L^n$.) Finally, let I be the $M \times M$ identity matrix. We state the following theorems. Proofs can be found in Cantwell (1988).

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

$$\text{Var}(y_h) = \sigma^2 \{\boldsymbol{a}^T\boldsymbol{a} + k^2\boldsymbol{b}^T(\boldsymbol{b}-2\boldsymbol{a}) + 2(\boldsymbol{a}-k^2\boldsymbol{b})^T\boldsymbol{Q}(\boldsymbol{a}-\boldsymbol{b})\}/(1-k^2) \qquad \textbf{(3)}$$

Notice that when one uses an unweighted average of the estimates from the m rotation groups of the current month, $k = 0$, $\ell = 0$, 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 month-to-month change.

(i) If $k = 0$, then $\mathrm{Var}(y_h - y_{h-1})$

$$= 2\sigma^2 \boldsymbol{a}^T (I - \rho_1 L)\boldsymbol{a};$$

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

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

Often of interest are the average over a certain period of time, for example, a quarter or a year, the difference in these averages from one period to the next, or even 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: $S_{h,t} = y_h + y_{h-1} + \cdots + y_{h-t+1}$, $t \geq 1$. Commonly used values of t include three, four, and twelve. We will leave it to the user to divide $S_{h,t}$ by t if he desires an average rather than a sum.

THEOREM 3. (a) Let $\boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$ be any sequence of $M\times 1$ nonrandom vectors. Then

$$\mathrm{Var}\Big(\sum_{i=0}^{\infty} \boldsymbol{v}_i^T \boldsymbol{x}_{h-i}\Big) = \sigma^2 \Big\{ \sum_{i=0}^{\infty} \boldsymbol{v}_i^T \boldsymbol{v}_i + 2\sum_{i=0}^{\infty} \boldsymbol{v}_i^T \sum_{n=1}^{M-1} \rho_n L^n \boldsymbol{v}_{i+n} \Big\} \quad \textbf{(5)}$$

(b) The expressions $S_{h,t}$, $y_h - y_{h-t}$, and $S_{h,t} - S_{h-t,t}$ can be written as $\sum_{i=0}^{\infty} \boldsymbol{v}_i^T \boldsymbol{x}_{h-i}$, where:

(i) for $S_{h,t}$: $\boldsymbol{v}_i =$

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

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

(ii) for $y_h - y_{h-t}$: $\boldsymbol{v}_0 = \boldsymbol{a}$, $\boldsymbol{v}_t = k^t(\boldsymbol{a}-\boldsymbol{b}) - \boldsymbol{a}$, and $\boldsymbol{v}_i =$

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

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

(iii) for $S_{h,t} - S_{h-t,t}$: $\boldsymbol{v}_i =$

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

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

(c) $-[k^{i-2t+1}(1-k^t)^2/(1-k)](\boldsymbol{a}-\boldsymbol{b})$, for $i = 2t, 2t+1, \ldots$ **(6)**

Another concern besides the variance of the estimators is time-in-sample bias. Suppose that $E(x_{h,i}) = Y_h + \beta_i$, 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 groups is unbiased amounts to saying that $\Sigma \beta_i = 0$.

THEOREM 4. Let $\boldsymbol{\beta}$ be the $M\times 1$ vector of month-in-sample bias terms in TIS form. Under the model above:

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

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

6. Covariance Structure and Theorems Under a LMLRP

As in the case of a one-level design, the covariance structure of the monthly panel estimates in a multi-level plan is assumed to be stationary in time. But now the effect of recall time on response enters. It may be reasonable to assume that response variability changes, in fact, likely increases, with the amount of time between the interview and the point of reference. We postulate the following covariance structure:

(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 $i \neq j$, i.e., estimates for the same month from different panels are uncorrelated; 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 are estimates obtained from the same panel r months apart; or 0, otherwise. Take $\rho_{0,i}$ to be 1 for all i. **(7)**

It may well be that $d_1 \leq d_2 \leq \ldots \leq d_p$, if response variability increases with recall time. As to the correlation coefficient $\rho_{r,i}$, r counts the number of months between estimates $x_{h,i}$ and $x_{h-r,j}$. The index i indicates that the estimate for month h is recorded from an interview in month $h+i$. It may appear as if the subscript j in $x_{h-r,j}$ plays no part in determining $\mathrm{Cov}(x_{h,i}, x_{h-r,j})$. However, 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. (This value is $j = \mathrm{mod}_p(i+r-1) + 1$, where $\mathrm{mod}_p(n)$ denotes the value of the integer n, modulo p.) Otherwise, the covariance is 0.

The coefficients $\rho_{r,i}$ will likely decrease in r for fixed i, reflecting smaller correlation as the separation between points in time grows. The effect of varying i for fixed r, though, is harder to predict, and may be related to the survey conducted and the characteristic being enumerated. In some cases, it may be appropriate to replace $\rho_{r,1}, \rho_{r,2}, \ldots, \rho_{r,p}$ with a commom ρ_r. Alternatively, the values of the $\rho_{r,i}$'s for different i's could depend on how many times the relevant panel has been interviewed between months $h-r$ and h. Results will be stated with general coefficients $\rho_{r,i}$; the reader can make substitutions according to his model or experience.

Here the definitions of certain symbols are analogous though slightly different from those in Section 5. The symbols are retained to emphasize the similarities in some results.

Define the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ as $(a_1, a_2, \ldots, a_p)^T$ and $(b_1, b_2, \ldots, b_p)^T$, respectively, according to the coefficients in the GCE. I is the $p\times p$ identity matrix. Let $\boldsymbol{D}$ be the $p\times p$ diagonal matrix with $d_1, d_2, \ldots, d_p$ down the diagonal. Similarly, for any $r \geq 0$, let $\boldsymbol{R}_r$ be the $p\times p$ diagonal matrix with $\rho_{r,1}, \rho_{r,2}, \ldots, \rho_{r,p}$ on the diagonal. Define the $p\times p$ matrix J by: $J_{i,i+1} = 1$ for $i = 1, 2, \ldots, p-1$; $J_{p1} = 1$, and $J_{ij} = 0$, otherwise. Finally, let

$$\boldsymbol{Q} = \sum_{n=1}^{\infty} k^n \boldsymbol{R}_n J^n. \tag{8}$$

This matrix $\boldsymbol{Q}$ is not the same as that defined in Section 5, but plays a similar role in the results. It is not difficult to show that the sum in (8) converges.

One may notice the similarities between the theorems stated previously, 1, 2, and 3, and the following theorems, 5, 6, and 7. The proofs of the latter three mimic those of the first three. They are found in Cantwell (1989).

Theorem 5. If the GCE of level is defined as in (1), and the covariance structure of (7) holds, then

$$\mathrm{Var}(y_h) = \sigma^2 \{\boldsymbol{a}^T \boldsymbol{D}^2 \boldsymbol{a} + k^2 \boldsymbol{b}^T \boldsymbol{D}^2 (\boldsymbol{b}-2\boldsymbol{a}) + 2(\boldsymbol{a}-k^2\boldsymbol{b})\boldsymbol{DQD}(\boldsymbol{a}-\boldsymbol{b})\}/(1-k^2) \tag{9}$$

When one uses an unweighted average of the estimates for month h from the p panels, $k = 0$, $\boldsymbol{Q} = 0$, and $a_i = 1/p$ for $i = 1, 2, \ldots, p$. Then $\mathrm{Var}(y_h) = (\sigma^2/p^2)\, \Sigma\, d_i^2$.

Theorem 6. Let $y_h - y_{h-1}$ be the GCE estimator of "monthly" change.

(i) If $k = 0$, then $\mathrm{Var}(y_h - y_{h-1})$

$$= 2\sigma^2 \boldsymbol{a}^T \boldsymbol{D}(I-\boldsymbol{R}_1 J)\boldsymbol{D}\boldsymbol{a};$$

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

$$= \sigma^2 \{\boldsymbol{a}^T \boldsymbol{D}^2 \boldsymbol{a} + k^2 \boldsymbol{b}^T \boldsymbol{D}\boldsymbol{b} - 2k\boldsymbol{a}^T \boldsymbol{D}\boldsymbol{R}_1 J\boldsymbol{D}\boldsymbol{b}\}/k - (1-k)^2 \mathrm{Var}(y_h)/k \tag{10}$$

Averages and differences of averages are again important statistics, especially in multi-level designs. Define $S_{h,t}$ as in Section 5, the sum of the GCE's for the last t months.

Theorem 7. (a) Let $\boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$ be any sequence of nonrandom $p\times 1$ vectors. Then

$$\mathrm{Var}(\sum_{i=0}^{\infty} \boldsymbol{v}_i^T \boldsymbol{x}_{h-i}) = \sigma^2 \{ \sum_{i=0}^{\infty} \boldsymbol{v}_i^T \boldsymbol{D}^2 \boldsymbol{v}_i + 2\sum_{i=0}^{\infty} \boldsymbol{v}_i^T \sum_{n=1}^{\infty} \boldsymbol{D}\boldsymbol{R}_n J^n \boldsymbol{D}\, \boldsymbol{v}_{i+n}\} \tag{11}$$

(b) The expressions $S_{h,t}$, $y_h - y_{h-t}$, and $S_{h,t} - S_{h-t,t}$ can be written as $\sum_{i=0}^{\infty} \boldsymbol{v}_i^T \boldsymbol{x}_{h-i}$, where the appropriate vectors $\boldsymbol{v}_i$ are found in (6).

For our applications, the sums in (5) and (11) converge because, in each of the three expressions, $\boldsymbol{v}_i$ is proportional to $k^i(\boldsymbol{a}-\boldsymbol{b})$ for $i \geq 2t$.

7. Additional Comments

Several unrelated topics are discussed in this section. Of primary importance is how useful these results are in actual surveys. The CPS and LFS gather data on labor force characteristics, such as work force and employed status. The correlations between rotation group estimates from one month to the next tend to be moderately positive, and beneficial to the implementation of composite estimation. As we mentioned earlier, CPS already uses an AK composite estimator.

The SIPP and NCS are examples of multi-level surveys. Many of the characteristics measured in the NCS involving incidents of crime may exhibit negligible correlation from one month to the next. If so, it would appear questionable whether the NCS could profit by using composite estimation rather than simple linear estimation from the months involved.

On the other hand, the SIPP seeks information on income level, sources of income, program participation, and other items. For many of these, the correlations of interest may be large enough to make our results useful to the SIPP.

Secondly, consider the covariance structure given for multi-level designs (see (7)). We mentioned that response variability may increase with recall time. This seems reasonable in surveys where participants respond from memory.

Nevertheless it has been pointed out to us that a somewhat opposite effect may occur in some business surveys. It is sometimes the case that, to a certain extent, response variability actually decreases with time. In some situations survey data are derived from business records which may not be complete or sufficiently accurate for several months. Minimum response variance might then be obtained by interviewing several months after the target month, rather than immediately.

Such observations, however, do not invalidate our results. No assumptions are made about the constants $d_1, d_2, \ldots, d_p$ except that they are positive.

A final point to raise is the difficulty of finding easily applied general formulae for a rotationally balanced multi-level design. Such a plan is more symmetric than the longitudinal plan considered here in some aspects, including time-in-sample. For any month, estimates are eventually obtained from each panel, one panel recalling one month, a second recalling two months, etc. Each panel comprises a set of rotation groups representing the entire range of times-in-sample.

This symmetry is offset computationally by the more intricate pattern of correaltions. For any month h and any i, $1 \leq i \leq p$, consider the panel which is interviewed in month $h+i$. There is an estimate from the rotation group which is in sample for the first time (disregard any groups used only for bounding purposes). This value is correlated with estimates from the same group for the previous $p-i$ months, but with nothing else. A second group is interviewed for the second time. Its estimate for month h is correlated with those for the prior $2p-i$ months. This pattern continues.

When the contributions and relationships of all the rotation groups in this panel have been sorted, one must bring in those from the other panels. Because each panel is interviewed in a different month, the corresponding covariances may be different. The entire process, although balanced and well-structured, is more intricate. This fact is reflected in the variance formulae for the generalized composite estimators of level and change. We have obtained some initial results which we plan to document.

ACKNOWLEDGMENTS

I would like to thank Lynn Weidman and Larry Ernst for reading the draft, checking the proofs, and suggesting improvements in content and style. David D. Chapman provided helpful comments and suggestions.

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.

CANTWELL, P. (1988). Variance Formulae for the Generalized Composite Estimator Under a Balanced One-Level Rotation Plan, Statistical Research Division Research Report Series, No. 88-26, Bureau of the Census, Washington, D.C.

CANTWELL, P. (1989). Variance Formulae for the Generalized Composite Estimator Under a Longitudinal Multi-Level Rotation Plan, Statistical Research Division Research Report Series, No. 89-4, Bureau of the Census, Washington, D.C.

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.

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.

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

APPENDIX

Estimates For 14 Months From 4 Panels:

MONTH ↓	PANEL 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 panel which is interviewed in month $h+i$. Solid horizontal lines ("seams") separate estimates which are obtained in different interviews.

*This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.