A MULTIVARIATE APPROACH TO ESTIMATION IN PERIODIC SAMPLE SURVEYS

Margaret Gurney and Joseph F. Daly
U.S. Bureau of the Census

1. Introduction:

The increasing use of periodic probability-based sample surveys to study the way in which selected characteristics of a population vary with time has raised many interesting problems of sample design and estimation. The degree of interest which this topic has generated may be gauged by the (admittedly incomplete) bibliography at the end of this paper, listing articles that have come to our attention.

Although the fundamental ideas in this area are already implicit in the basic papers of Jessen [10] and Patterson [18], they are often obscured in the literature by the myriad of sampling details surrounding the problems to which they are applied. In this paper, which is essentially expository, we shall examine a specific periodic sample survey in an effort to show how the problem of constructing "good" estimators can be fruitfully regarded as a problem in standard multivariate analysis and treated by means of techniques currently in use for handling general stochastic processes.

2. The Problem:

A sample survey which continues over a period of time is capable of producing, for each time period, many estimates of each of the characteristics of the population being studied. Each individual observation can be used to make an estimate, or the individual observations can be combined in some desired manner to make one or more estimates for a particular time period.

An estimate which does not make use of the survey data for any time period except that period to which the estimate refers may be called an "elementary estimate." It should be possible to improve such an estimate by making use of correlated elementary estimates available from other time periods. The purpose of this paper is to discuss minimum variance unbiased linear combinations of elementary estimates, and to outline a method of computation which will determine the weights to be used on the various elementary estimates to obtain the best linear unbiased estimate, whether for an estimate of level, of change over time, of an average over time, or in general, for any linear combination of the elementary estimates.

3. A Special Case:

The Current Population Survey of the Bureau of the Census is a monthly household survey, which has been in operation for many years. Data are collected on labor force items, demographic characteristics, and on other characteristics of the population for which a household survey is an appropriate vehicle. At each month estimates are made, for many characteristics, of the current level, of changes since earlier months, and of averages over several months.

The survey is based on a rotating sample in which one-fourth of the households are replaced by a new selection of households each month. In each month, one-fourth of the households are new, one-fourth have been interviewed in the preceding month, one-fourth have been interviewed for three consecutive months, and one-fourth have been interviewed for four months. Each fourth of the sample is treated separately and an elementary estimate for the month is made from it.

The following additional information is available for the survey, and can be used to improve the estimates which are made from the sample for a particular characteristic: (a) the elementary estimates for past months, for each panel, (b) estimates of the average correlations over time for observations one month, two months, and three months apart, and (c) an estimate of the average variance for a single panel, averaged over time.

Consider an estimate for a particular month, say January. If the panels are labelled "A", "B", "C", etc., with "A" corresponding to the panel which has just entered the survey in January, the rotation pattern may be described by the following diagram. The numbering of the elementary estimates X_i in the following chart is chosen to facilitate the computations. This numbering makes the covariance matrix of the X's a straightforward direct product of simple covariance matrices. It is obvious how to terminate the numbering, for a finite number of months.

Month	Panel									
	A	B	C	D	E	F	G	H	I	etc.
Jan.	X_1	X_2	X_4	X_7						
Dec.		X_3	X_5	X_8	X_{11}					
Nov.			X_6	X_9	X_{12}	X_{15}				
Oct.				X_{10}	X_{13}	X_{16}	X_{19}			
Sept.					X_{14}	X_{17}	X_{20}	X_{23}		
Aug.						X_{18}	X_{21}	X_{24}	X_{27}	
etc.							etc.			

As a first approximation we shall assume the following characteristics for the survey:

A. The expected value of the estimates of the 4 panels for a given month is the same, e.g., for January $EX_1 = EX_2 = EX_4 = EX_7$.

B. The covariances between estimates from the same panel at two different months depend only on the number of months between the two estimates. For example,

$$\mathrm{var}(X_1) = \mathrm{cov}(X_1,X_1) = \mathrm{cov}(X_2,X_2) = \sigma^2$$

$$\mathrm{cov}(X_2,X_3) = \mathrm{cov}(X_4,X_5) = \mathrm{Cov}(X_5,X_6) = \rho_1\sigma^2$$

$$\mathrm{cov}(X_4,X_6) = \mathrm{cov}(X_7,X_9) = \mathrm{cov}(X_{12},X_{14}) = \rho_2\sigma^2$$

$$\mathrm{cov}(X_7,X_{10}) = \mathrm{cov}(X_{11},X_{14}) = \mathrm{cov}(X_{15},X_{18}) = \rho_3\sigma^2$$

The average values of ρ_1, ρ_2 and ρ_3 are estimated from past data from the survey. For convenience we shall put $\sigma^2 = 1$.

C. The panels are selected independently, so that (for example)

$$\mathrm{cov}(X_1,X_2) = \mathrm{cov}(X_2,X_4) = \mathrm{cov}(X_2,X_6) = 0, \text{ etc.}$$

Under these assumptions an estimate for January may be made in several ways. The simplest is to use only the observations for the month of January for panels, A, B, C, and D. Since the panels are independent, the best estimate in this case is

$$_1\hat{X}_{Jan} = (X_1 + X_2 + X_4 + X_7)/4$$

It can easily be shown that this estimate has the minimum variance of any linear combination of X_1, X_2, X_4, and X_7; its variance is .25 times the variance of an individual elementary estimate.

When such a survey has been in operation for two months, starting in December, the data available for use in making the best estimate for January are the four observations for panels A, B, C, and D for January, and the four observations for panels B, C, D, and E for December. It is possible to determine the coefficients C_i so that the estimate

$$_2\hat{X}_{Jan} = C_1X_1 + C_2X_2 + C_4X_4 + C_7X_7 + C_3X_3 + C_5X_5 + C_8X_8 + C_{11}X_{11}$$

shall have the minimum variance, knowing the relations

$$\mathrm{cov}(X_2,X_3) = \mathrm{cov}(X_4,X_5) = \mathrm{cov}(X_7,X_8) = \rho_1$$

$$EX_1 = EX_2 = EX_4 = EX_7 \,;\; EX_3 = EX_5 = EX_8 = EX_{11}$$

For an estimate of the civilian labor force, with $\rho_1 = .8$, the coefficients corresponding to the minimum variance are shown in Table 1.

Table 1.--ESTIMATE OF CIVILIAN LABOR FORCE FOR JANUARY FROM CPS SURVEY: COEFFICIENTS OF MINIMUM VARIANCE UNBIASED LINEAR ESTIMATES FOR SURVEYS STARTING IN JANUARY, DECEMBER, OCTOBER, AUGUST AND APRIL

Correlation Pattern: $\rho_1 = .8$, $\rho_2 = .7$, $\rho_3 = .65$

Duration of survey	Panel												
	A	B	C	D	E	F	G	H	I	J	K	L	M
1. One month (started in January)													
January....................	.250	.250	.250	.250									
2. Two months (started in preceding December)													
January....................	.219	.260	.260	.260									
December....................		-.052	-.052	-.052	.156								
3. Four months (started in preceding October)													
January....................	.195	.258	.269	.278									
December....................		-.079	-.036	-.032	.147								
November....................			-.063	-.025	-.005	.093							
October....................				-.061	-.021	.004	.078						
4. Six months (started in preceding August)													
January....................	.189	.258	.271	.282									
December....................		-.086	-.037	-.032	.155								
November....................			-.074	-.026	-.002	.102							
October....................				-.075	-.005	.007	.073						
September....................					-.045	-.006	.001	.050					
August....................						-.027	-.013	.000	.040				
5. Ten months (started in preceding April)													
January....................	.187	.259	.271	.283									
December....................		-.089	-.037	-.032	.158								
November....................			-.077	-.026	-.002	.105							
October....................				-.080	-.006	.007	.079						
September....................					-.053	-.006	.003	.056					
August....................						-.037	-.005	.002	.040				
July....................							-.026	-.003	.001	.028			
June....................								-.018	-.002	.001	.019		
May....................									-.012	-.001	.000	.013	
April....................										-.007	-.003	.000	.010

Table 1 shows also the coefficients for an estimate of the civilian labor force for January when the survey has been in operation four, six, and ten months, respectively. It may be seen that the coefficients for the earliest months in the sample are smaller than the coefficients for the most recent months; they approach zero rather rapidly as the number of months increases. Moreover, the coefficients for the observations from the most recent month approach constant values as the number of months increases; if the survey had been started much earlier, the coefficients for January would be approximately equal to those which are shown for ten months. In fact, after six months, the coefficients for the most recent months are close to the corresponding coefficients for ten months. Hence data from the most recent six to ten months will provide practically all of the improvement which can be achieved in the estimate. The speed of convergence depends upon the covariances between estimates for the same panel at different times. If the covariances are very high (for example $\rho_1 = .95$), the time required for convergence is longer; on the other hand, if ρ_1 is small, say .50, the convergence will be much faster, and only three or four months may be required to approximate the optimum coefficients.

If the survey has been in operation for two or more months, it is possible to make a revised estimate for a preceding month which will have a smaller variance than the one originally obtained for that month. For example, an estimate for December, made when data for January are available, will usually have a smaller variance than the original estimate for January. Coefficients for an estimate of civilian labor force for December, using the data for January, as well as all earlier data, are shown in Table 2, for a survey which has been in operation for two, four, six, and ten months, respectively.

In this manner one can make a revised estimate for November, using data through January, which will have a smaller variance than one which used data only through November (or December).

Table 2.--ESTIMATE OF CIVILIAN LABOR FORCE FOR DECEMBER FROM CPS SURVEY: COEFFICIENTS OF MINIMUM VARIANCE UNBIASED LINEAR ESTIMATES FOR SURVEY STARTING IN JANUARY, DECEMBER, OCTOBER, AUGUST AND APRIL

Correlation Pattern: $\rho_1 = .8$, $\rho_2 = .7$, $\rho_3 = .65$

Duration of survey	Panel												
	A	B	C	D	E	F	G	H	I	J	K	L	M
1. One month (started in January)													
No estimate for December													
2. Two months (started in preceding December)													
January	.156	-.052	-.052	-.052									
December		.260	.260	.260	.219								
3. Four months (started in preceding October)													
January	.128	-.052	-.041	-.035									
December		.224	.264	.267	.245								
November			-.060	-.033	-.035	.128							
October				-.042	-.047	-.006	.095						
4. Six months (started in preceding August)													
January	.120	-.051	-.039	-.030									
December		.214	.263	.267	.256								
November			-.072	-.034	-.032	.138							
October				-.060	-.023	-.003	.086						
September					-.063	-.005	.005	.063					
August						-.035	-.015	.000	.050				
5. Ten months (started in preceding April)													
January	.118	-.051	-.038	-.029									
December		.211	.263	.267	.259								
November			-.077	-.034	-.031	.142							
October				-.067	-.024	-.002	.093						
September					-.073	-.005	.007	.071					
August						-.047	-.006	.003	.050				
July							-.032	-.005	.002	.035			
June								-.022	-.003	.001	.024		
May									-.015	.002	.001	.016	
April										-.009	-.004	.000	.013

The variances corresponding to the best selection of coefficients for an estimate of the civilian labor force, for January and for earlier months, are shown in Table 3.

Table 3.--VARIANCES OF DELAYED ESTIMATES OF CIVILIAN LABOR FORCE: SURVEY ENDING IN JANUARY

Correlation Pattern: $\rho_1 = .8$, $\rho_2 = .7$, $\rho_3 = .65$

Estimates	Duration of survey (months)				
	1	2	4	6	10
A. Monthly Level:					
Jan..........	.250	.219	.195	.189	.187
Dec..........		.219	.183	.173	.170
Nov..........			.183	.167	.162
Oct..........			.195	.167	.157
Sept.........				.173	.155
Aug..........				.189	.155
July.........					.157
June.........					.162
May..........					.170
April........					.187
B. Month-to-month change:					
Jan. - Dec...		.125	.122	.122	.122
Dec. - Nov...			.121	.120	.120
Nov. - Oct...			.122	.120	.119
Oct. - Aug...				.120	.118
Sept. - Aug..				.122	.118
Aug. - July..					.118
July - June..					.119
June - May...					.120
May - April..					.122

The best coefficients for any linear combination of the estimates for January, December, etc., are obtained by taking the same linear combination of the coefficients of the best estimates of the corresponding months. For example, the coefficients for the best estimate of changes from December to January, $X'_{Jan} - X'_{Dec}$, are obtained by subtracting the coefficients in Table 2 from the corresponding coefficients in Table 1. The variance of such a difference is given by

$$\mathrm{var}(X'_{Jan} - X'_{Dec}) = \mathrm{var}(X'_{Jan}) + \mathrm{var}(X'_{Dec}) -$$

$2\mathrm{cov}(X'_{Jan}, X'_{Dec})$, where the covariance between X'_{Jan} and X'_{Dec} is a number which is obtained as part of the general solution of the system. Table 3 presents also variances for estimates of month-to-month change. when the survey has been in operation two, four six, and ten months. The variance of month-to-month change is quite stable, as can be seen from the table.

4. General Procedure:

To obtain the coefficients of the desired minimum variance linear unbiased estimate, we find it helpful to use a geometric approach suggested by the work of Parzen [16], [17] on the application of Hilbert space methods to stochastic processes.

In our problem we have a finite set of chance variables $X_1, \ldots, X_n$ whose joint probability distribution is assumed to belong to a family of distributions subject only to the conditions:

A. No one of the X's is essentially a linear function of the remaining X's. (If initially we had

$$X_n = \sum_{i=1}^{n-1} C_i X_i + \text{const.},$$

with probability one, we assume that we have eliminated X_n from the set, and so on.)

B. The covariances $K_{ij} = E(X_i - EX_i) \cdot (X_j - EX_j)$ are finite and known.

C. The expected values $\mu_i = EX_i$ are subject only to certain linear homogeneous restrictions such as

$$\mu_2 = \mu_3, \ \mu_4 = \mu_5 = \mu_6, \text{ etc.}$$

or more generally,

$$\sum_{i=1}^{n} a_{hi}\, \mu_i = 0 \qquad (h = 1, \ldots, p)$$

Our approach, in brief, is to let the variables $X_1, \ldots, X_n$ correspond to some vectors $\underline{k}_1, \ldots, \underline{k}_n$ which form a basis for an n-dimensional Hilbert space $\underline{V}$. (Halmos [6], [7]). The vector $\underline{m} = (\mu_1, \ldots, \mu_n)$ lies in this space, and is by virtue of condition C free to range over some subspace $\underline{M}$ of $\underline{V}$. To find the coefficients λ_i of the best linear estimator $\Sigma \lambda_i X_i$ of an arbitrary homogeneous linear combination of the μ_i, say $\Sigma C_i \mu_i$, we form the vector $\Sigma C_i \underline{k}_i$, find its projection $\underline{v}^*$ on the subspace $\underline{M}$, and then express $\underline{v}^*$ in terms of the $\underline{k}$'s. The coefficients in this expression will be the required λ's.

To set up the correspondence, we begin by regarding the rows of the covariance matrix

$$K = \begin{pmatrix} K_{11} & \cdots & K_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ K_{n1} & \cdots & K_{nn} \end{pmatrix}$$

as vectors $\underline{k}_i = \{K_{i1}, \ldots, K_{in}\}$ in an ordinary n-dimensional vector space $\underline{V}$. Next we show that these vectors $\underline{k}$ form a basis for $\underline{V}$, so that any vector $\underline{a}$ with arbitrary components $\alpha_1, \ldots, \alpha_n$ can be represented uniquely in the form $a = \Sigma\, \theta_i \underline{k}_i$. This follows from the fact that the $\underline{k}_i$ are linearly independent, which in turn follows from our condition A. above and from the fact that

$$\mathrm{var}(\Sigma\, \lambda_i X_i) = \Sigma\, \lambda_i K_{ij} \lambda_j.$$

Since any vector $\underline{v}$ in $\underline{V}$ is a linear combination of the basis vectors $\underline{k}$, we can define the inner product of any two vectors

$$\underline{a} = \Sigma\, \theta_i \underline{k}_i, \qquad \underline{b} = \Sigma\, \Psi_j \underline{k}_j$$

in $\underline{V}$ by

$$(\underline{a},\underline{b}) = (\Sigma\, \theta_i \underline{k}_i, \Sigma\, \Psi_j \underline{k}_j) = \Sigma\, \theta_i K_{ij} \Psi_j.$$

This inner product can be thought of as defining angles and lengths in $\underline{V}$ by interpreting the inner product of $\underline{a}$ and $\underline{b}$ as the product of the "lengths" of $\underline{a}$ and $\underline{b}$ by the cosine of the "angle" between them. In particular, we shall say that $\underline{a}$ and $\underline{b}$ are perpendicular if $(\underline{a},\underline{b}) = 0$, and shall call $(\underline{a},\underline{a})$ the square of the length of $\underline{a}$. It then follows that the variance of any linear estimator $\Sigma\, \lambda_i X_i$ is equal to the length of the corresponding vector $\Sigma\, \lambda_i \underline{k}_i$:

$$(\Sigma\, \lambda_i \underline{k}_i, \Sigma\, \lambda_j \underline{k}_j) = \Sigma\, \lambda_i K_{ij} \lambda_j = \mathrm{var}(\Sigma\, \lambda_i X_i) \tag{1}$$

We note further that

$$(\underline{a},\underline{k}_j) = (\Sigma\, \theta_i \underline{k}_i, \underline{k}_j) = \Sigma\, \theta_i K_{ij},$$

which is by definition the j-th component of the vector $\underline{a}$. Hence if $\underline{m}$ is the mean-value vector with components $\mu_1, \ldots, \mu_n$ it follows that

$$(\underline{m}, \Sigma\, \lambda_i \underline{k}_i) = \Sigma\, \lambda_i \mu_i = E(\Sigma\, \lambda_i X_i) \tag{2}$$

Thus if $\underline{a} = \Sigma\, \theta_i \underline{k}_i$ is any vector in $\underline{V}$, the expected value of the estimator $\Sigma\, \theta_i X_i$ is given by $(\underline{m},\underline{a})$ and its variance by $(\underline{a},\underline{a})$.

Let us now examine the conditions C' on the components $\mu_i = EX_i$ of the vector $\underline{m}$. Suppose, for example, we have the situation (see Section 3)

Month \ Panel	A	B	C	D	E
January........	X_1	X_2	X_4	X_6	
December.......		X_3	X_5	X_7	X_8

We would then expect the mean-value vector to be of the form

$$\underline{m} = \{\alpha, \alpha, \beta, \alpha, \beta, \alpha, \beta, \beta\}$$

so that we could write $\underline{m} = \alpha\underline{u}_1 + \beta\underline{u}_2$ where

$$\underline{u}_1 = \{1, 1, 0, 1, 0, 1, 0, 0\}$$

$$\underline{u}_2 = \{0, 0, 1, 0, 1, 0, 1, 1\}.$$

More generally, the conditions C imply that there are some linearly independent vectors $\underline{u}_1, \ldots, \underline{u}_m$ such that $\underline{m}$ satisfies C if and only if $\underline{m}$ is of the form

$$\underline{m} = \Sigma\, \xi_\alpha \underline{u}_\alpha$$

i.e. $\underline{m}$ ranges over the subspace $\underline{M}$ of $\underline{V}$ spanned by the vectors $\underline{u}_1, \ldots, \underline{u}_m$.

Let us now single out a particular basis vector $\underline{k}_h$ corresponding to the estimator X_h. As we have seen,

$$(\underline{m},\underline{k}_h) = \mu_h = EX_h$$

for any mean-value vector $\underline{m} = \{\mu_1, \ldots, \mu_h\}$. Suppose $\underline{v}$ is any other vector in $\underline{V}$ such that

$$(\underline{m},\underline{v}) = \mu_h$$

for every vector $\underline{m}$ in $\underline{M}$. Since each of the $\underline{u}_\alpha$ is then a possible choice of $\underline{m}$, we must have

$$(\underline{u}_\alpha,\underline{v}) = u_{\alpha h} \qquad (\alpha = 1, \ldots, m) \tag{3}$$

where $u_{\alpha h}$ stands for the h-th component of $\underline{u}_\alpha$.

We now show that among such vectors $\underline{v}$ satisfying (3) we can find one, say $\underline{v}^*$, which lies in $\underline{M}$. This is equivalent to the problem of finding numbers $\gamma_1, \ldots, \gamma_m$ which satisfy the equations

$$(\underline{u}_\alpha, \Sigma\, \gamma_\beta \underline{u}_\beta) = \Sigma(\underline{u}_\alpha,\underline{u}_\beta)\gamma_\beta = u_{\alpha h}.$$

Unless the matrix of elements $L_{\alpha\beta} = (\underline{u}_\alpha, \underline{u}_\beta)$ is singular, these equations have the (unique) solution

$$\gamma_\beta = \Sigma\, u_{\alpha h} L^{-1}_{\alpha\beta} \tag{4}$$

Where $L^{-1}_{\alpha\beta}$ stands for the (α,β) element of the inverse of $(L_{\alpha\beta})$. But if $(L_{\alpha\beta})$ were singular, there would be numbers ζ_α not all zero such that $\Sigma\, \zeta_\alpha L_{\alpha\beta} = 0$. This would imply $0 = \Sigma\, \zeta_\alpha L_{\alpha\beta} \zeta_\beta = (\Sigma\, \zeta_\alpha \underline{u}_\alpha, \Sigma\, \zeta_\beta \underline{u}_\beta)$, so so that the vector $\Sigma\, \zeta_\alpha \underline{u}_\alpha$ would have zero length. However, in view of our earlier discussion of the K matrix, the only vector of zero length in $\underline{V}$ is the zero vector. And the linear independence of the vectors $\underline{u}_\alpha$ makes it impossible to have $\Sigma\, \zeta_\alpha \underline{u}_\alpha = 0$ unless every ζ is zero.

We now have a vector $\underline{v}^*$ which lies in $\underline{M}$ and which satisfies the conditions

$$(\underline{u}_\alpha, \underline{v}^*) = (\underline{u}_\alpha, \underline{k}_h) = u_{\alpha h} \qquad \alpha = 1, \ldots, m.$$

As an immediate consequence of these conditions, we have

$$(\underline{u}_\alpha, \underline{k}_h - \underline{v}^*) = 0 \qquad \alpha = 1, \ldots, m.$$

Geometrically speaking, we have resolved the vector $\underline{k}_h$ into two components: a component $\underline{v}^*$ which lies in $\underline{M}$ and a component $\underline{k}_h - v^*$ which is perpendicular to every vector in $\underline{M}$. In this sense we call $\underline{v}^*$ the projection of $\underline{k}_h$ on the subspace $\underline{M}$.

The fact that $\underline{v}^*$ is the shortest vector satisfying (3) now follows readily. Let $\underline{v}$ be any other vector in $\underline{V}$ which satisfies (3). Then

$$(\underline{u}_\alpha, \underline{v} - \underline{v}^*) = 0 \qquad \alpha = 1, \ldots, m$$

so that $\underline{v} - \underline{v}^*$ is perpendicular to every vector in $\underline{M}$, and in particular, $(\underline{v}^*, \underline{v} - \underline{v}^*) = 0$. Consequently

$$(\underline{v}, \underline{v}) = (\underline{v}^* + \underline{v} - \underline{v}^*, \underline{v}^* + \underline{v} - \underline{v}^*) = (\underline{v}^*, \underline{v}^*) + (\underline{v} - \underline{v}^*, \underline{v} - \underline{v}^*)$$

since the usual cross-product term $2(\underline{v}^*, \underline{v} - \underline{v}^*)$ vanishes. Hence

$$(\underline{v}, \underline{v}) \geq (\underline{v}^*, \underline{v}^*)$$

Thus the estimator corresponding to the vector $\underline{v}^*$ will have minimum variance among all unbiased linear estimators of μ_h.

To find the coefficients of this best linear estimator, we must express the vector $\underline{v}^*$ in terms of the basis vectors $\underline{k}_i$.

From (4) we have

$$\underline{v}^* = \Sigma\, u_{\alpha h} L^{-1}_{\alpha\beta} \underline{u}_\beta .$$

But $\underline{u}_\beta = \Sigma\, \theta_i \underline{k}_i$ where the θ_i are determined from the relations

$$u_{\beta j} = \Sigma\, \theta^{(\beta)}_i K_{ij}, \text{ or } \theta^{(\beta)}_i = \Sigma\, u_{\beta j} K^{-1}_{ji}.$$

Hence

$$v^* = \Sigma\, u_{\alpha h} L^{-1}_{\alpha\beta} u_{\beta j} K^{-1}_{ji} \underline{k}_i \tag{5}$$

where $L_{\alpha\beta}$ can be expressed in terms of the $u_{\alpha i}$ and the K_{ij} by noting that

$$L_{\alpha\beta} = (\Sigma\, \theta^{(\alpha)}_i \underline{k}_i, \Sigma\, \theta^{(\beta)}_j \underline{k}_j) = \Sigma\, \theta^{(\alpha)}_i K_{ij} \theta^{(\beta)}_j = \Sigma\, u_{\alpha i} K^{-1}_{ij} u_{\beta j}.$$

From Equation (5) we then see that the coefficient of X_i in the best linear unbiased estimator of μ_h is

$$\lambda^{(h)}_i = \Sigma\, u_{\alpha h} L^{-1}_{\alpha\beta} u_{\beta j} K^{-1}_{ji} \tag{6}$$

The preceding discussion was formally restricted to finding the best linear unbiased estimator of some particular component μ_h of the mean-value vector. But from the easily demonstrated fact that the projection on $\underline{M}$ of a linear combination of vectors $\underline{v}_h$ is the same linear combination of their projections, it follows that the optimum estimator of some linear combination of the μ's, such as $\underline{\mu}_3 - \mu_1$, is the same linear combination of the optimum estimators of the individual components $\underline{\mu}_h$.

Equations (6) in the matrix form

$$(\lambda) = U^T (U K^{-1} U^T)^{-1} U K^{-1}, \text{ where } U = (u_{\alpha i}) \tag{7}$$

are readily programmed on an electronic computer to determine the λ's for given μ's and K's.[1]

Two final comments seem in order at this point. In the first place from $\underline{v}^* = \Sigma\, u_{\alpha h} L^{-1}_{\alpha\beta} u_\beta$ we readily obtain the variance of the optimum estimator of μ_h:

$$(\underline{v}^*, \underline{v}^*) = \Sigma\, u_{\alpha h} L^{-1}_{\alpha\beta} u_{\beta h} \tag{8}$$

which simplifies considerably in the problem studied in this paper because in general for a given h one of the $u_{\alpha h}$ is one, and the rest are zero. And secondly the matrix

$$(U K^{-1} U^T)^{-1} U K^{-1}$$

---

[1] The computation is illustrated in Appendix A.

has the properties of a "generalized inverse" of U^T, a notion that is being increasingly exploited in recent research on least squares estimation (Greville [5], Goldman and Zelen [4]).

5. Approximation to the Optimum Estimate:

The purpose of this section is to examine some alternative estimators which approximate the optimum estimate, and to compare them with the optimum. A desirable feature of these estimators is that they are somewhat easier to compute. Moreover, they provide estimates which have most of the gains of the optimum estimators.

Several forms of "composite estimators" will be considered. A composite estimate is a weighted average to two (or more) linear unbiased estimates of the same characteristic for a given time period; the weights are selected so as to reduce the variance, as compared with the variances of the original estimates. These composite estimators are defined recursively, and use only a limited number of elementary estimates, combined with composite estimates which have already been computed. The description below of several composite estimators will illustrate the definition.

A. Simple Composite Estimator:

To form a "simple composite estimator" for a given month, say January,

(1) make a simple average of the elementary estimates for January from panels A, B, C, and D:

$$X'_{Jan} = (X_1 + X_2 + X_4 + X_7)/4$$

(2) make a difference estimate for January by adding to the (already computed) composite estimate for December the estimate of the December-January change, based on identical panels. Let the composite estimator be designated by X^*, and the change by $\Delta_{J,D}$:

$$\Delta_{J,D} = (X_2 + X_4 + X_7 - X_3 - X_5 - X_8)/3$$

The difference estimate for January is

$$X^*_{Dec} + \Delta_{J,D}$$

(3) make a weighted average of the estimates of (1) and (2) above:

$$X^*_{Jan} = (1-K)X'_{Jan} + K(X^*_{Dec} + \Delta_{J,D}) \qquad (9)$$

B. Composite Estimator with Change from Three Previous Time Periods:

The rotation pattern of Section 3 permits estimates of change for identical panels to be made for two successive months, for times two months apart, and for times three months apart. An estimator which permits the use of this additional information is the following:

$$X^*_{Jan} = (1-K-L-M)X'_{Jan} + K(X^*_{Dec} + \Delta_{J,D}) + L(X^*_{Nov} + \Delta_{J,N}) + M(X^*_{Oct} + \Delta_{J,O}) \qquad (10)$$

Here X'_{Jan} is defined as before, and the Δ's are self-explanatory; X^* as used here is of course different from that in Equation (9).

C. A Modification of the Simple Composite Estimator:

Some improvement in the estimate of Equation (9) can be made if the linear combination of observations for January has more weight on panel A, and less on panels B, C, and D. Such a change will bring the coefficients on the observations for January more in line with coefficients of the optimum estimate. Let the term $(1-K)X'_{Jan}$ in Equation (9) be replaced by

$$\{(1-K+A)X_1 + (1-K-A/3)(X_2+X_4+X_7)\}/4 \qquad (11)$$

which is equivalent to

$$(1-K)X'_{Jan} + \{(AX_1 - \tfrac{A}{3}(X_2+X_4+X_7)\}/4$$

The expected value of the term in braces is zero. This form is called an "AK-Composite Estimator"; except for the addition of the terms in braces, the formula is the same as that in Equation (9).

D. Composite Estimator with Year-to-year Change:

The Current Population Survey is actually based on eight panels rather than four, which rotate in such a way that there is a 50 percent overlap of households from year to year, as well as the 75 percent overlap from month to month. When the year-to-year correlation is high (relative to the 12th power of the month-to-month correlation), appreciable gains may be obtained by the use of year-to-year change in the composite estimate. An appropriate formula is

$$X^*_{Jan} = (1-K-Q)X''_{Jan} + K(X^*_{Dec} + \Delta_{J,D}) + Q(X^*_{Jan-12} + \Delta_{Jan,Jan-12}) \qquad (12)$$

where X''_{Jan} is an average based on the eight January panels, and Jan-12 refers to January of the preceding year. X''_{Jan} may be a simple average, or may have unequal weights on the eight panels.

E. More General Forms of Composite Estimators:

More general composite estimators can be developed; however, an estimate which is too general will lose some of the advantages of the simpler estimates: it may require the use of too many of the elementary estimates, and may require the retention of too many earlier composite estimates.

The four forms of the composite estimator given above are versions of

$$X^*_{Jan} = Z_{Jan} + KX^*_{Dec} + LX^*_{Nov} + MX^*_{Oct} + QX^*_{Jan-12} \qquad (13)$$

where Z_{Jan} is composed of contributions from the estimates of level for January, and from several estimates of change.

F. Comparisons with the Minimum Variance Unbiased Linear Estimator:

How well a particular composite estimator will approximate the minimum variance unbiased linear estimator may be measured by a comparison of the variances which are obtained by the two estimators. One may also compare the coefficients to be used on the elementary estimates, for different estimators.

Table 4 shows the coefficients for the most recent four months (Panels A-G) which are appropriate for several composite estimates, for an estimate of the civilian labor force; it compares them with the coefficients for the minimum variance unbiased linear estimate based on data for ten months. The comparisons are made for the months of January, December, November, and October. In each case, the constants (K, L, M, Q) used for the composite estimate are the ones corresponding to the smallest value of the variance for the particular estimate considered.

Table 4 shows also the variances of estimates for January of the civilian labor force, for the various estimators.[2] The numbers are relative to the variance of the simple average for January. It is seen that the AK-composite estimate approximates the minimum variance unbiased linear estimator quite closely: its variance is only slightly larger than that of the best estimator, and the coefficients are close to those for the best estimate.

The estimate using year-to-year change (Estimate 5) is not strictly comparable to the other estimates in this table, as it uses information ($\Delta_{Jan,Jan-12}$) which is not available to the ten-month minimum variance unbiased linear estimate. It is included to point out that a high year-to-year correlation pattern can effect further improvements in the estimate.

6. The AK-Composite Estimate Used with Several Characteristics:

The AK-composite estimate is a good approximation to the optimum linear unbiased estimate for a characteristic such as the civilian labor force when A is .4 and K is .7. For an estimate such as the change in monthly level of the labor force, or for another characteristic such as unemployment (which has much lower correlations over time than the civilian labor force), these values of A and K will not be the best. Table 5 shows the variances expected for several correlation patterns, for estimates of level and of month-to-month change, for a number of values of A and K.

It may be seen from an examination of Table 5 that the values A = .2, and K = .7, while not the best for all characteristics, still provide appreciable gains over the simple average of elementary estimates, and even over the present composite estimate for the CPS (which uses the values A = 0 and K = .5) when the correlations are moderate or high.

When the correlations are low very small gains may be expected, compared with the simple average; there may be losses if the values of A and K differ appreciably from zero.

7. Variations:

The data in Tables 1 through 5 have been based on average values of variances and covariances, and on equal means for all of the observations at all time periods. In practice the variances and covariances will change over time, the means at different time periods may have a seasonal pattern, or a long term trend, and the expected values of the observations relating to the same time period may not be identical.

A. Changing Correlation Pattern:

Table 6 presents coefficients and variances for the minimum variance unbiased linear estimate for ten months when the correlations are not equal over time, but vary by as much as 20 percent from equality. The correlations used are shown at the top of the table. Even this

[2] The method of computation is illustrated in Appendix B.

rather large departure from equal correlations has almost no affect on the variance of the estimate of the January level: by comparison with Table 3 it is increased from .187 to .189. The variance of the estimate of January-December change is unchanged at .122. The fact that the variances and covariances are not known more exactly is of little importance.

Table 4.--ESTIMATE OF CIVILIAN LABOR FORCE FOR JANUARY FROM CPS SURVEY

A. Coefficients for Recent Months and Variances, for Several Linear Unbiased Estimates

Estimate and month	Panel							Variance of estimate for January (relative to the simple average)
	A	B	C	D	E	F	G	
1. Minimum variance unbiased linear estimate -- 10 months started in April								
January................	.187	.259	.271	.283				.749
December...............		-.089	-.037	-.032	.158			
November...............			-.077	-.026	-.002	.105		
October................				-.080	-.006	.007	.079	
2. Simple composite estimate: K = .6								
January................	.100	.300	.300	.300				.817
December...............		-.140	-.020	-.020	.180			
November...............			-.084	-.012	-.012	.108		
October................				-.050	-.007	-.007	.064	
3. Three-month composite estimate: K = .4, L = .1, M = .05								
January................	.112	.246	.296	.346				.796
December...............		-.088	-.035	-.015	.138			
November...............			-.074	-.039	.023	.090		
October................				-.083	-.007	.023	.067	
4. AK-composite estimate: K = .7, A = .4								
January................	.175	.275	.275	.275				.756
December...............		-.111	-.041	-.041	.193			
November...............			-.077	-.029	-.029	.135		
October................				-.054	-.020	-.020	.094	

B. Coefficients and Variance for Composite Estimate Using Year-to-year Change

Estimate and month	Panel														
	A	B	C	D	E	F	G	...	M	N	O	P	Q	R	S
5. Month-month and year-year change: K = .5, Q = .2															
Jan.	-.037	.121	.121	.121				...	.087	.171	.171	.171			
Dec.		-.065	-.023	-.023	.061					-.040	.002	.002	.086		
Nov.			-.033	-.011	-.011	.030					-.020	.001	.001	.043	
Oct.				-.016	-.006	-.006	.015					-.010	.001	.001	.021
								⋮							
Variance of estimate for January relative to the simple average:								Jan.-12	-.042	-.026	-.026	-.026			
								Dec.-12		-.031	-.016	-.016	-.001		
								Nov.-12			-.016	-.011	-.011	.006	
.709								Oct.-12				-.015	-.007	-.007	.006
								⋮							

Table 5.--COMPARISON OF VARIANCES OF AK-COMPOSITE ESTIMATES FOR SEVERAL CORRELATION PATTERNS, AND FOR SEVERAL VALUES OF A AND K

Correlation pattern	K	Variance, relative to simple average (A = 0, K = 0) of --									
		Estimate of monthly level					Estimate of month-to-month change[1]				
		A=0	A=.1	A=.2	A=.3	A=.4	A=0	A=.1	A=.2	A=.3	A=.4
1. High-- ρ_1=.95, ρ_2=.93, ρ_3=.90	.4	.791	.784	.785	.792		.524	.543	.580	.636	
	.5	.725	.717	.716	.722	.735	.442	.455	.485	.532	.596
	.6	.651	.642	.640	.644	.657	.370	.379	.403	.443	.499
	.7	.576	.564	.560	.563	.574	.309	.314	.334	.369	.419
	.8	.518	.500	.492	.491	.499	.262	.265	.282	.312	.357
2. Moderate-- ρ_1=.80, ρ_2=.70, ρ_3=.65	.4	.857	.845	.840	.843		.728	.735	.755	.787	
	.5	.829	.812	.803	.802	.808	.690	.692	.706	.732	.770
	.6	.817	.792	.777	.770	.771	.661	.660	.669	.689	.721
	.7	.848	.806	.780	.764	.756	.641	.637	.642	.658	.684
	.8	.978	.920	.874	.837	.812	.635	.627	.629	.640	.662
3. Low-- ρ_1=.50, ρ_2=.40, ρ_3=.30	.4	.971	.951	.938	.933		.928	.922	.925	.935	
	.5	1.005	.975	.953	.939	.934	.931	.922	.920	.926	.939
	.6	1.084	1.038	1.003	.978	.961	.942	.930	.924	.925	.933
	.7	1.262	1.187	1.130	1.084	1.048	.960	.945	.935	.932	.936
	.8	1.676	1.560	1.458	1.370	1.297	.985	.968	.955	.949	.948

[1] Based on difference of two estimates of level.

Table 6.--ESTIMATE OF CIVILIAN LABOR FORCE FROM CPS SURVEY: MINIMUM VARIANCE UNBIASED LINEAR ESTIMATE

Coefficients and Variances with a Changing Correlation Pattern

A. Changing Correlation Pattern

Correlation between:	ρ_1	Correlation between:	ρ_2	Correlation between:	ρ_3
Jan.-Dec.	.80	Jan.-Nov.	.70	Jan.-Oct.	.65
Dec.-Nov.	.76	Dec.-Oct.	.68	Dec.-Sept.	.61
Nov.-Oct.	.72	Nov.-Sept.	.64	Nov.-Aug.	.57
Oct.-Sept.	.68	Oct.-Aug.	.60	Oct.-July	.53
Sept.-Aug.	.64	Sept.-July	.60	Sept.-June	.57
Aug.-July	.68	Aug.-June	.64	Aug.-May	.61
July-June	.72	July-May	.68	July-April	.65
June-May	.76	June-April	.70		
May-April	.80				

B. Variances

Estimate	Variance
Jan. level	.189
Dec. level	.176
Jan.-Dec. change	.122

C: Coefficients for Panels (Estimate of January Level)

Month	Panel												
	A	B	C	D	E	F	G	H	I	J	K	L	M
Jan.	.189	.254	.272	.285									
Dec.		-.082	-.042	-.026	.150								
Nov.			-.073	-.039	.014	.098							
Oct.				-.077	-.017	.022	.072						
Sept.					-.046	-.014	.012	.048					
Aug.						-.032	-.010	.011	.031				
July							-.020	-.006	.006	.020			
June								-.016	-.003	.005	.014		
May									-.010	-.002	.002	.010	
April										-.007	-.003	.001	.009

B. Unequal Means at a Particular Time Period:

It may happen that the observations at a single time period do not all have the same expected value. For example, in the Current Population Survey, the first time a household is interviewed it appears to respond differently to the interview, with respect to some characteristics (for example, employment status), than at the second or later interviews. Following the diagram in Section 3, one might have for January:

$$EX_1 = (1+a)\mu$$

when

$$EX_2 = EX_4 = EX_7 = \mu$$

There may be the same kind of bias in the reports for December, November, etc.

When there are unknown response biases in the expected values, both the minimum variance unbiased linear estimator and the various composite estimators may produce estimates which are biased.

If the pattern of bias is constant over time the total bias will approach a limit for each of the estimators which have been discussed here. Table 7 shows the biases to be expected in several cases.

In Table 7 a characteristic is considered which is possessed by about 10,000,000 persons in the population and which has a correlation pattern similar to that of the civilian labor force. The sampling error of this estimate from a simple average of elementary estimates is 200,000 (i.e., about 2 percent). Two patterns of bias in the estimates from the four panels at a single month are considered:

(1) The bias occurs only at the first time at which a household is interviewed. The bias is of the same order at each time period. The pattern of expected values at a single time period is

$$(1+a)\mu,\ \mu,\ \mu,\ \mu.$$

(2) The "newest" and "oldest" panels have compensating biases. The pattern of expected values for each month is

$$(1+a)\mu,\ \mu,\ \mu,\ (1-a)\mu.$$

Table 7 is computed for values of "a" equal to 100,000 and 200,000. The resulting root mean square errors are compared with the standard errors of an unbiased estimate, for several estimators.

Table 7 shows that when the bias in the estimate from the "new" panel is one-half (or even equal to) the size of the standard error of the estimate, the root mean square error is hardly any larger than the standard error of the corresponding estimate. The gains which are achieved in using the minimum variance unbiased linear estimate, or the several composite estimates, persist, even with a bias of this size.

Table 7.--EFFECT OF CONSISTENT BIAS ON THE RELIABILITY OF SAMPLE ESTIMATES OF A CHARACTERISTIC HAVING A CORRELATION PATTERN OF $\rho_1 = .8$, $\rho_2 = .7$, $\rho_3 = .65$

(Size of Estimate is 10,000,000)

Estimator	Standard error of estimate	a = 100 x 10^3		a = 200 x 10^3			
		Bias in new panel only (1+a, 1, 1, 1)		Bias in new panel only (1+a, 1, 1, 1)		Compensating bias in new and oldest panels (1+a, 1, 1, 1-a)	
	(10^3)	Bias (10^3)	RMSE (10^3)	Bias (10^3)	RMSE (10^3)	Bias (10^3)	RMSE (10^3)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Simple average............	200	-25	202	-50	205	0	200
Minimum variance unbiased linear estimate-10 months.	173	-21	174	-42	178	-201	265
Simple composite:							
K = .6................	181	-25	182	-50	188	-200	270
K = .5................	182	- 8	182	-17	183	- 33	185
AK-composite:							
K=.7, A=.4................	174	-19	175	-39	178	-222	282
K=.6, A=.3................	175	- 6	176	-12	176	-149	230
K=.5, A=.3................	179	+ 7	179	+13	180	- 93	202

If biases should occur in two panels, and should be compensating, (see columns 7 and 8 of Table 7) the simple average of the elementary estimates would be unbiased. The estimators which make the most use of past data--the minimum variance estimator and the AK-composite with a high value of K--show the largest increases due to such a bias. The columns for compensating biases are included in the table to illustrate perhaps the worst situation which could occur; the likelihood of compensating biases is very small.

If the deviations in expected values are known, or can be measured quite accurately, one may consider the advisability of adjusting the sample estimates accordingly. However, the assumption that the bias will continue to be the same in the future as in the past may lead to serious errors. A better procedure is to try to eliminate the response bias, if it is significant.

C. A Different Rotation Pattern: The Census Current Business Reports Survey:

For small retail and service establishments, The Current Business Reports Survey of the Bureau of the Census is based on 12 panels, one of which is enumerated each month. At the time of enumeration, information on retail sales is obtained for the preceding month and for the next earlier month. After a year, the panel which was in the sample a year ago is enumerated again.

The rotation pattern is

Month	Jan.	Dec.	Nov.	Oct.	Sept.	Aug.	July	June	May	April	Mar.	Feb.
January.	X_1											
December	X_2	X_3										
November		X_4	X_5									
October.			X_6	X_7								
September				X_8	X_9							
August..					X_{10}	X_{11}						
July....						X_{12}	X_{13}					
June....							X_{14}	X_{15}				
May.....								X_{16}	X_{17}			
April...									X_{18}	X_{19}		
March...										X_{20}	X_{21}	
February											X_{22}	X_{23}
January.	X_{25}											X_{24}
December	X_{26}	X_{27}										
						etc.						

For this rotation pattern, the minimum variance unbiased linear estimate is very close in form to a composite estimate. If the rotation pattern is altered so that the sample for any month is independent of that for any other month (i.e., the sample for the months of one year are not repeated in subsequent years), then the minimum variance estimate can be expressed exactly in composite form.

For an estimate of January level, we write

$$X^*_{Jan} = X_1 - K_J X_2 + K_J X^*_{Dec}.$$

This estimator will have minimum variance when

$$K_J = \text{cov}(X_1,X_2)/\{\text{var}(X_2) + \text{var}(X^*_{Dec})\}.$$

Here X^*_{Dec} is defined in a similar manner, in terms of X_3, X_4, X^*_{Dec} and a constant K_D, which is defined in terms of $\text{cov}(X_3,X_4)$, $\text{var}(X_4)$, and $\text{var}(X^*_{Oct})$.

These relationships hold, whether the survey has started recently, or whether it has been in operation for a long time. When the covariances are all equal to ρ, and the variances of the individual elementary estimates (X_i) are all equal, say σ^2, the value of K approaches a limit, as time passes:

$$K = \frac{1 - \sqrt{1-\rho^2}}{\rho}$$

and the variance becomes

$$\text{var}(X^*) = \sigma^2 \sqrt{1-\rho^2}.$$

With the altered pattern, the best estimate of month-to-month change (say from December to January) will be obtained by making a revised estimate for December, using the data available from the January survey, and subtracting this from the best January estimated. The revised estimate for December can also be written in composite form

$$X^*_{Dec(rev)} = (1-K/\rho)X_2 + (K/\rho-K)X_3 + K(X^*_{Nov(rev)} + X_3 - X_4)$$

where, as before,

$$K = (1 - \sqrt{1-\rho^2})/\rho$$

The variance of the revised estimate is

$$\sigma^2 K/\rho \sqrt{1-\rho^2} \;;$$

it is smaller than the variance of the unrevised estimate by a factor of K/ρ.

The best estimate of the December-January difference is then $X^*_{Jan} - X^*_{Dec(rev)}$ as was noted by Patterson [18]. In fact, an estimate of this form, which uses all of the available data, has the minimum variance of any linear estimate of month-to-month change, even when the variances and correlations between panels are not constant over time.

APPENDIX A

Computation of Minimum Variance Unbiased Linear Estimate

To illustrate the computation of the minimum variance unbiased linear estimate, consider a set of three elementary estimates, which might be obtained at the beginning of a survey with a rotation pattern like that of the Current Business Reports Survey, as shown in the diagram in paragraph C of Section 7. The observation X_1 is an elementary estimate for the month of January; observations X_2 and X_3 are elementary estimates for the preceding month, December. It is desired to make the minimum variance unbiased linear estimate of level for January, having the following information:

1. $EX_1 = \mu_1;\; EX_2 = EX_3 = \mu_2$
2. The covariance matrix is

$$K = \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

3. The mean value vector is

$$\underline{m} = (\mu_1, \mu_2, \mu_2)$$

which can be written as

$$\underline{m} = \mu_1 \underline{u}_1 + \mu_2 \underline{u}_2$$

where

$$\underline{u}_1 = (1, 0, 0)$$

$$\underline{u}_2 = (0, 1, 1)$$

The matrix of coefficients of the minimum variance unbiased linear estimator for estimating the expected values of the three variables X_1, X_2, and X_3 is (see Equation 7)

$$C = U^T(U\,K^{-1}\,U^T)^{-1}\,UK^{-1}$$

The covariance matrix of the optimum solutions is (see Equation 8)

$$P = U^T(U\,K^{-1}\,U^T)U.$$

The solution is indicated in the following equations

$$K^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho & 0 \\ -\rho & 1 & 0 \\ 0 & 0 & 1-\rho^2 \end{pmatrix}$$

Define

$$L = U\,K^{-1}U^T = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 2-\rho^2 \end{pmatrix}$$

$$L^{-1} = \frac{1}{2}\begin{pmatrix} 2-\rho^2 & +\rho \\ +\rho & 1 \end{pmatrix}$$

Define the matrix

$$P = U^T\,L^{-1}\,U = \frac{1}{2}\begin{pmatrix} 2-\rho^2 & \rho & \rho \\ \rho & 1 & 1 \\ \rho & 1 & 1 \end{pmatrix}$$

and the matrix of coefficients

$$C = PK^{-1} = \begin{pmatrix} \underline{c}_1 \\ \underline{c}_2 \\ \underline{c}_3 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 2 & -\rho & +\rho \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

The minimum variance unbiased linear estimate for January level is obtained when the coefficients $\underline{c}_1$ are used:

$$\hat{X}_1 = X_1 - \rho(X_2 - X_3)/2.$$

The optimum estimates of December level employ the coefficients $\underline{c}_2$ and $\underline{c}_3$, respectively, and lead to identical solutions:

$$\hat{X}_2 = \hat{X}_3 = (X_2 + X_3)/2$$

The covariance matrix P, which gives the variances and covariances between

$$\hat{X}_1,\ \hat{X}_2, \text{ and } \hat{X}_3$$

leads to

$$\text{var}(\hat{X}_1) = (2 - \rho^2)/2$$

$$\text{var}(\hat{X}_2) = \text{var}(\hat{X}_3) = 1/2$$

The variance of the best estimate of change $(\hat{X}_1 - \hat{X}_2)$ is

$$\begin{aligned}\text{var}(\hat{X}_1 - \hat{X}_2) &= \text{var}(\hat{X}_1) + \text{var}(\hat{X}_2) - 2\,\text{cov}(\hat{X}_1, \hat{X}_2)\\ &= (3 - 2\rho - \rho^2)/2.\end{aligned}$$

These results may be verified by conventional methods. For example, one may construct the variance of

$$\hat{X}_1 = \sum_{j=1}^{3} C_{1j} X_j,$$

and determine the values of the coefficients which minimize the variance, subject to conditions 1 and 2.

APPENDIX B

Computation of Variances for Composite Estimates

The general composite estimator, Equation (13), may be written as

$$X^*_1 = Z_1 + KX^*_2 + LX^*_3 + MX^*_4 + QX^*_{13} \qquad (14)$$

where the subscript "1" designates the most recent month, "2" is the preceding month, etc.

To illustrate the computation of the variance of a composite estimate, consider the simple composite estimator:

$$\begin{aligned}X^*_1 &= Z_1 + KX^*_2\\ &= Z_1 + KZ_2 + K^2Z_3 + K^3Z_4 + \ldots \qquad (15)\end{aligned}$$

1. Define, at each time period t, for $t = 1, 2, 3, \ldots$ and $i = 0, 1, 2, \ldots$

$$\begin{aligned}Y_o &= \text{var}(X^*_t)\\ Y_i &= \text{cov}(X^*_t, X^*_{t+i})\\ A_i &= \text{cov}(Z_t, X^*_{t+i})\\ \rho_{iz}\,\text{var}(Z) &= \text{cov}(Z_t, Z_{t+i})\end{aligned}$$

and

$$\rho_{o,z} = 1$$

2. Take covariances between Equation (15) and X^*_1, and X^*_2:

$$\begin{aligned}Y_o &= A_o + KY_1\\ Y_1 &= A_1 + KY_o\end{aligned} \qquad (16)$$

These two simultaneous equations can be solved for Y_o and Y_1 if $|K| < 1$. The solution will give values for $\text{var}(X^*_1)$ and $\text{cov}(X^*_1, X^*_2)$; higher covariances can be obtained successively from

$$Y_i = A_i + KY_{i-1} \qquad (17)$$

for $i = 2, 3, 4, \ldots$

3. The $\{A_i\}$ satisfy a set of covariance equations obtained by taking covariances between Z_1 and $\{X^*_i = Z_i + KX^*_{i+i}\}$ for $i = 1, 2, 3, \ldots$:

$$A_i = \rho_{iz}\,\text{var}(Z) + KA_{i+1} \qquad (18)$$

In particular, for the simple composite estimate (Equation 17)

$$\begin{aligned}A_o &= \rho_{o,z}\,\text{var}(Z) + KA_1\\ A_1 &= \rho_{1z}\,\text{var}(Z) + KA_2\\ A_2 &= \rho_{2z}\,\text{var}(Z) + KA_3\end{aligned} \qquad (19)$$

etc.

4. The form of Z_t is determined by the rotation pattern and the weights assigned to the panels. For the simple composite estimate, with the rotation pattern of the diagram in Section 3, we find

$$Z_1 = (1-K)(X_1+X_2+X_4+X_7)/4 + K(X_2+X_4+X_7-X_3-X_5-X_8)/3$$

or

$$Z_1 = \frac{(1-K)}{4}X_1 + \frac{3+K}{12}(X_2+X_4+X_7) - \frac{K}{3}(X_3+X_5+X_8) \qquad (20)$$

Z_2 is defined similarly, using the elementary estimates with subscripts corresponding to months "2" and "3,", etc.

The variances and covariances

$$\text{cov}(Z_\alpha, Z_{\alpha+i}) = \rho_{iz}\,\text{var}(Z)$$

may be expressed in terms of the variances and covariances of the original (elementary) estimates X_i; that is, in terms of $\text{var}(X)$ and $\{\rho_{ix}\}$.

For the rotation pattern of the diagram, correlations more than three months apart are zero; it turns out that only var(X), ρ_{1X}; ρ_{2X} and ρ_{3X} are non-zero. Starting with $A_4 = 0$, in Equation (18), A_3, A_2, A_1, and A_o may be determined recursively, for particular values of K, and of

var(X) and ρ_{1X}, ρ_{2X}, ρ_{3X}.

5. Having found $\{A_i\}$, Equations (16) may be solved for Y_o and Y_1; in general, $\{Y_i\}$ may be found by applying Equation (17) successively.

6. The form of the $\{Y_i\}$ lends itself readily to the evaluation of a number of estimators, in addition to the estimate of level for a single month. For example:

Variance of monthly level $= \text{var}(X_1^*)$

$= Y_o$

Variance of month-to-month change in level $= \text{var}(X_1^* - X_2^*)$

$= 2(Y_o - Y_1)$

Variance of year-to-year change in level $= \text{var}(X_1^* - X_{13}^*)$

$= 2(Y_o - Y_{12})$

7. By varying the value of K, while keeping $|K| < 1$, it is possible to set up a table of variances for each estimator (of level, month-to-month change, etc.) and to determine the value of K which leads to the minimum value for each. Frequently a compromise value of K can be found which will make gains for several characteristics, and for several estimators, although it may not be the best value for any one.

8. For more complicated composite estimators the system will consist of more than two equations. For example, if year-to-year change is incorporated into the estimate, the system will consist of 13 equations. The form of $\{Z_i\}$ may also be quite complicated, depending on the conditions set on the coefficients of the estimate of level, and on the number of estimates of change which are used.

REFERENCES

[1] Bershad, M.A., and Nisselson, H., "Sampling for Weekly Time Series," Proceedings of the Social Statistics Section, American Statistical Association (1962), pp. 207-213.

[2] Cochran, W.G. Sampling Techniques. New York: John Wiley & Sons, Inc., Second Edition (1963), pp. 341-354.

[3] Eckler, A.R., "Rotation sampling," Annals of Mathematical Statistics,26(1955),pp.664-685.

[4] Goldman, A.J., and Zelen, M., "Weak generalized inverses and minimum variance linear unviased estimation," Journal of Research, National Bureau of Standards, 68B (1964), pp. 151-172.

[5] Greville, T.N.E., "Some applications of the pseudo-inverse of a matrix," Society for Industrial and Applied Mathematics Review, 2 (1960), pp. 15-22.

[6] Halmos, P.R. Introduction to Hilbert Space and the Theory of Spectral Multiplicity. New York: Chelsea Publishing Company, Second Edition 1957).

[7] Halmos, P.R. Finite Dimensional Vector Spaces. Princeton: D. van Nostrand Co., Inc., Second Edition (1958).

[8] Hansen, M.H., Hurwitz, W.N., and Madow, W.G. Sample Survey Methods and Theory. New York: John Wiley & Sons, Inc. (1953), Vol. I, pp. 490-503, Vol. II, pp. 268-279.

[9] Hansen, M.H., Hurwitz, W.N., Nisselson, H., and Steinberg, J., "The redesign of the Census Current Population Survey," Journal of the American Statistical Association, 50 (1955), pp. 701-719. Bershad, M.A. and Gurney, M., Appendix: "Winin PSU variance of the composite estimate."

[10] Jessen, R.J., "Statistical investigation of a sample survey for obtaining farm facts," Iowa Agricultural Experiment Station Research Bulletin, 304 (1942).

[11] Kruskal, W., "The coordinate-free approach to Gauss-Markov estimation, and its application to missing and extra observations," Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. I (1960), pp. 435-451.

[12] Kulldorff, G., "Some problems of optimum allocation for sampling on two occasions," Review of the International Statistical Institute, 31 (1963), pp. 24-57.

[13] Narain, R. D., "On the recurrence formula in sampling on successive occasions, "Journal of the Indian Society of Agricultural Statistics, 5 (1953), pp. 96-99.

[14] Narain, R.D., "The general theory of sampling on successive occasions," Bulletin of the International Statistical Institute, 34 (2) (1954), pp. 87-89, 201-202.

[15] Onate, B.T., "Development of multi-stage designs for statistical surveys in the Philippines," Mimeo-multilith series No. 3: Statistical Laboratory, Iowa State University, Ames, Iowa. (1960).

[16] Parzen, E., "Statistical inference on time series by Hilbert space methods, I," Technical Report No. 23: Department of Statistics, Stanford University (1959).

[17] Parzen, E., "An approach to time series analysis," Annals of Mathematical Statistics, 32 (1961), pp. 951-989.

[18] Patterson, H.D., "Sampling on successive occasions with partial replacement of units," Journal of the Royal Statistical Society, Series B, 12 (1950), pp. 241-255.

[19] Rao, J.N.K., and Graham, J.E., "Rotation designs for sampling on repeated occasions," Journal of the American Statistical Association, 59 (1964), pp. 492-509.

[20] Singh, D., and Singh, B.D., "Double sampling for stratification on successive occasions," Journal of the American Statistical Association, (In print -- 1965).

[21] Tikkiwal, B.D., "Optimum allocation in successive sampling," Journal of the Indian Society of Agricultural Statistics, 5 (1953), pp. 100-102.

[22] Tikkiwal, B.D., "A further contribution to the theory of univariate sampling on successive occasions," Journal of the Indian Society of Agricultural Statistics, 8 (1956), pp. 84-90.

[23] Tikkiwal, B.D., "On the theory of classical regression and double sampling estimation," Journal of the Royal Statistical Society, Series B, 22 (1960), pp. 131-138.

[24] U. S. Bureau of the Census. The Current Population Survey -- A Report on Methodology. Technical Paper No. 7, U.S. Government Printing Office, Washington, D. C. (1963).

[25] Waksberg, J., and Pearl, R.B., "The Current Population Survey: A Case History in Panel Operations," Proceedings of the Social Statistics Section, American Statistical Association, (1964), pp. 217-228.

[26] Woodruff, R.S., "The use of rotating samples in the Census Bureau's monthly surveys," Journal of the American Statistical Association, 58 (1959), pp. 454-467.

[27] Yates, F. Sampling Methods for Censuses and Surveys. London: Charles Griffin and Co., Third Edition (1960), pp. 175-182, 233-235, 260-262.