GRAHAM, John Elwood, 1933–
ROTATION DESIGNS FOR SAMPLING ON SUCCESSIVE OCCASIONS.

Iowa State University of Science and Technology
Ph.D., 1963
Mathematics

ROTATION DESIGNS FOR SAMPLING ON SUCCESSIVE OCCASIONS

by

John Elwood Graham

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of

The Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1963

TABLE OF CONTENTS

Page

I. INTRODUCTION

It is now almost universally recognized that a properly conducted sample survey can often be a good alternative to a complete census when information on the characteristics of some population is desired. Perhaps the two chief advantages of a sample survey are (a) the saving in cost over a complete enumeration can be appreciable, (b) it might be virtually an impossible task to economically compile detailed information on an entire population within a limited period of time so that larger non-sampling errors will be encountered with a census compared with a sample survey. The main two disadvantages of a sample survey would appear to be that (a) it does not permit the myriad of breakdowns, cross-classifications, "small area" statistics, etc., that users of the data might demand, (b) any figures derived from a sample are subject to sampling error because of the failure to enumerate the entire population. It is apparent that in a complex economy both the sample survey and the complete census have important roles to play.

A census or a sample survey taken at one point of time can obviously furnish data which are relevant to that point of time only. If the population characteristics are relatively stable then the statistics so obtained will be adequate for some time to come. However in a dynamic population characterized by significant changes in characteristics within a short period of time a census or a sample survey conducted infrequently is of limited use. It is thus of some importance that the data be collected in as brief a time interval as possible. Zarkovich (1961) has vividly illustrated both static and dynamic populations using a historical series on the number of native males per hundred females in the United States as an illustration of the former and the number of employed persons in the agricultural sector of the United States economy during the period 1955-56 as an illustration of the latter.

Since the rate of change itself can be highly variable, "non-repetitive change surveys" which are taken only two or three times for the purpose of providing information on changes can meet only short-term needs. Any such information is, of course, still preferable to a "one shot survey." There is obviously a need for sample surveys which are conducted at more or less regular time intervals; these have been referred to as "current change surveys." Such surveys are usually expensive to operate since a permanent field organization may be required to conduct them. It is therefore desirable to develop efficient designs for their specific purpose. Considerable research has been conducted towards achieving this end. It is hoped that this dissertation will prove to be a useful contribution to this body of knowledge.

Yates (1949) has listed five alternative vehicles for the collection of up-to-date information in a dynamic population. These are:

(1) A complete census may be repeated in its original form at intervals.

(2) A new sample on each occasion may be conducted without regard to previous samples, i.e., independent samples.

(3) A survey may be repeated on the same sample, i.e., a fixed sample or a fixed panel.

(4) Part of the sample may be replaced on each occasion, the

remainder being retained, i. e., partial replacement.

(5) A re-survey of a subsample of the original sample may be made, i. e., a subsample.

We shall be concerned here with the fourth of these alternatives. This sampling scheme has been referred to in various parts of the literature as "sampling on successive occasions with partial replacement of units," "rotation sampling," and "sampling for a time series." These terms all refer to the process of dropping some of the old units out of the sample and adding new units to the sample on each occasion. The rotation plan specifies the number of occasions for which any given unit provides information in the survey. It thus determines the number of units that will be matched between occasions. Our attention will be focused upon rotation designs with a fixed plan of partial replacement. It is possible to ascertain from the prespecified rotation plan on what occasions any given unit enters and leaves the sample. The sample size n and the population size N are both fixed over time and the replacement fraction from occasion to occasion is constant. By a one cycle design is meant a rotation pattern wherein a unit enters the sample for r occasions and then drops out and does not return. Similarly in a two cycle design a unit comes into the sample for r occasions, drops out for m occasions, returns for another r occasion and then drops out once again but does not return. One and two cycle designs demand that the population size be effectively infinite if the sampling conditions are to be met. With an infinite cycle design a unit is permitted to enter the sample for r occasions, to leave for m occasions and to continue doing this without limit; it is thus required that $n(m+r)/r = N$. In practice one would select a random sample of size N' from N and carry out the rotation on the N' units only. Since N' is arbitrary the above restriction does not rob the design of its generality. Infinite cycle designs are of interest when the sampling fraction n/N is not small, e.g., as when rotating within primary or secondary sampling units in a multi-stage design.

In one-level rotation sampling only sample values that have been drawn from the population on the current occasion may be added to the pattern of sample values previously available. Such a design is illustrated in the following pattern (1), in which each row represents a rotation group of units and an X in the column footed by t_{-i} denotes that the group was sampled at the i-th previous survey.

					X
				X	X
			X	X	X
		X	X	X	
	X	X	X		
X	X	X			
X	X				
X					
t_{-5}	t_{-4}	t_{-3}	t_{-2}	t_{-1}	t_0

(1)

At each t_α, $\alpha = 0, -1, -2, \ldots$, there are n sample values in the sample pattern. Of the n units in the sample at time $t_{\alpha-1}$, $(1-\mu)n$ of these are retained in the sample for observation at time t_α and the remaining μn are replaced by μn new ones. In the above example the replacement fraction is $\mu = 1/3$. Those values to the right of the vertical line are added at t_0 to the previously observed values to the left of this line. The term "one-level" refers to the fact that one column of observations is added to sample pattern at each occasion. Thus the sample overlap, or the extent to which the same sampling unit is surveyed over time, agrees with the information overlap.

In two-level rotation sampling values referring to the previous occasion as well as to the current occasion are added to the sample pattern on the current occasion. Such a design is illustrated in pattern (2).

				X	X
				X	X
				X	X
			X	X	
			X	X	
			X	X	
		X	X		
		X	X		
		X	X		
	X	X			
	X	X			
	X	X			
X	X				
X	X				
X	X				
t_{-5}	t_{-4}	t_{-3}	t_{-2}	t_{-1}	t_0

(2)

At time t_0 not only the sample values to the right of the vertical line but also those above the upper horizontal line augment the pattern of values. A sample of n new units is selected on each occasion, the replacement fraction μ therefore being unity. The previous value as well as the current value of the sampled units are recorded. It would obviously be folly to employ a partial replacement scheme under such conditions. "Two-level" indicates that sample values from two columns are added to the pattern at each occasion. The sample overlap and the information overlap do not agree.

The extension to three-level and multi-level patterns is immediate. For example, in a three-level design values referring to occasions t_{-2} and t_{-1} as well as those collected for t_0 are adjoined to the pattern on occasion t_0.

In a truncated rotation pattern only the data from the $\ell \geq 2$ most recent occasions are included in the estimator, the remainder being ignored. Truncation is performed for two possible reasons: (a) the bulk of the variance reduction may result from the use of the most recent data only, (b) the assumed correlation model may be only an accurate description of the true correlation structure locally. Neither two-level rotation sampling nor truncated rotation patterns will be dealt with here.

It is through the choice of an estimator that the statistician derives benefit from the sample overlap when sampling on successive occasions with partial replacement of units. Estimators of the current occasion mean or total can be constructed in two different ways in such situations. The data from the current occasion only can supply an unbiased estimator of the current occasion level; Yates (1949) calls these "overall estimates." A recurrence-type estimator may be formed by adding to the estimate for the previous occasion an estimate of change from the matched units only. The composite estimator of level is then a weighted average of these two types of estimates. A composite estimate of the change between any two occasions is provided by the difference of the composite estimates of level for the individual occasions. A more precise estimator of change is formed by revising the estimate of the earlier occasion level in light of the matched sample data that has been gathered since that time. The estimate of change so derived will not be consistent with the two individual estimates of level and so, for practical reasons, is not usually favoured.

There is considerable latitude in designing a rotation sample. The choice of an optimum design is governed by such factors as the variability in the different populations, the relative importance of change and current occasion statistics, various cost factors, the frequency with which estimates are to be released, the nature and variety of data to be collected, the correlation structure between matched units over time, and so on. With such factors in mind the statistician determines a compromise design requiring decisions as to the extent of the sample overlap, the possibility of information overlap, the spacing of interviews, the size of a rotation group and the number of cycles to be made, the form of the composite estimator and the choice of weights for its component parts, among other things. Many of these aspects of rotation sampling will be dealt with in the succeeding chapters. The criterion to be adopted here for selecting a certain design and/or estimator will usually be that of minimum sampling variance.

What are the advantages and disadvantages of a partial replacement sample design coupled with composite estimation relative to the use of completely independent samples or matched panels? Some of the more important considerations are listed below.

(1) The composite estimation technique makes use of past as well as current information. If there is a strong positive correlation between measurements on the same unit on successive occasions then moderate efficiency gains in the estimate of level relative to an overall estimate may be anticipated. Very significant efficiency gains are achieved when estimating the change between occasions.

(2) A rotation design possesses the benefits of both independent samples and a matched panel but exploits neither to their fullest. A matched panel is obviously best for the estimation of change whereas a complete turnover on each occasion is optimum for estimating the overall mean over several occasions.

(3) When sampling human populations the field costs are likely to be lower if the same unit is enumerated on several occasions.

(4) Once initial contact with a sampling element has been made and his confidence gained subsequent interviews may well bring improved cooperation from him.

(5) Increased response resistance is often encountered after several interviews with the same respondent. Gray and Cortlett (1950) reported a forty per cent defection from a fixed panel during the course of five consecutive monthly interviews.

Consequent serious biases in estimates were therefore anticipated since the sample could no longer be judged as being at all representative of the universe. A partial replacement design relieves the respondent of an unduly heavy reporting burden and therefore assists in maintaining the response rate. A matched sample may not be feasible at all if the character being measured is of such a nature that the respondent may be unwilling to furnish the desired information more than once.

(6) Continued reporting by a respondent may condition his response. For example, in a series of farm management surveys the respondent may actually improve his practices because of an increased awareness of the value of certain procedures and through advice solicited from the enumerator.

(7) The conditioning of response may be investigated with a partial replacement design since entirely new units as well as matched units are available for analysis on each occasion.

(8) A partial replacement design furnishes valuable information regarding variances, correlations and costs. These permit the implementation of near-optimum procedures because of the flexibility of such designs. Mahalanobis (1952) used the terms "historical schemes" or "sequential designs" to designate surveys whose design is altered over time in view of the information compiled in the course of repeated surveys. Conversely non-historical schemes or non-sequential designs do not make use of the information which becomes progressively

available to improve the design of subsequent surveys.

(9) Defects in the survey procedure are often more quickly observed in a matched survey design.

(10) The quality of current information can be improved through the possibility of comparing responses at different points of time and rectifying descrepancies either by a second visit or by editing rules. Bounded interview techniques wherein the respondent is supplied with a record of his previous interview responses serve to jog his memory and to fix the time period of reference more clearly. The caliber of response in some types of surveys may be so improved and the often quoted "telescoping effect" minimized.

(11) By revising past estimates in the light of more recent data improved historical series may be made available.

(12) Rotation designs are uniquely set up to handle the unexpected occurrence of large units in the sample which can so greatly increase the variance. The reader is referred to Bershad and Nisselson (1962) for further details.

(13) Two-level and multi-level designs exploit the advantages of both a complete matching and independent samples to the fullest. The possibility of serious memory biases when recall over a longer time period is necessary may make such designs highly undersirable.

In this dissertation we shall develop the theory of successive sampling with a fixed rotation pattern, i.e., equal sample sizes and a constant replacement fraction, using the Hansen _et al._ (1955) estimator and extensions thereof. Expressions for the estimator of the current occasion mean and change between current and previous occasions and their variances in a finite population are developed for an infinite cycle design. The resulting variance functions under the assumed exponential correlation model are of a complex nature. Consequently numerical investigations are carried out to estimate the optimum values of the various design and estimator variables. The theory is extended to two-stage sample designs where either primaries or secondaries are rotated. A discussion of estimators involving the ratio of two composite estimators is given and their application with respect to the estimation of the sample mean in two-stage designs is illustrated. Rotation designs exhibiting a finite number of cycles are of some practical interest. Attention is primarily devoted to a particular two cycle rotation design employed by the United States Bureau of the Census. An improvement to the Hansen _et al._ (1955) estimator is suggested in a specific design situation where the correlation between successive occasions is not monotonely decreasing as the interval between observations increases. In the special case of a one cycle three visit design a "multi component estimator" is developed. The derivation of the variance function is of some special interest in itself in that the solution of a second order difference equation is involved.

Eckler (1955, p. 668) fairly well summarized the situation with respect to rotation sampling when he remarked that "it seems quite likely that rotation sampling will be of most value when (a) the correlation is high, and (b) it is so difficult to draw a sample that the sample size must be kept as small as possible. If it costs no more to carry out rotation sampling than independent random sampling, then even a modest reduction of five to ten per cent in variance will be worthwhile."

II. REVIEW OF THE LITERATURE

Because the concept of sampling on successive occasions is related to the principles of double sampling it would therefore seem natural to begin by briefly summarizing the basic essentials involved therein. A complete account of the theory is available in most of the standard textbooks which dwell on sample survey methodology, e.g., Sukhatme (1954).

Neyman (1938) developed the theory for the following double sampling procedure. A large sample of size n' is selected and a character x which is correlated with the character y of interest is recorded. The cost of securing the x information is assumed to be considerably less than that of gathering the y information. This preliminary sample is then sub-divided into strata within which the character x varies little. If the correlation between x and y is large this should prove to be an effective stratification for the y variate as well. A stratified random sample of size $n < n'$ is now selected from the n' units and the y characteristic is also recorded for these units. The nomenclature "double sampling" was coined because of the two sampling investigations involved. The estimator of the y mean used by Neyman is a simple stratified mean with unknown stratum weights which are estimated from the first sample. Expressions for the sample sizes n and n' which minimize the variance of the x mean subject to a linear cost function are derived. This theory is also presented in some detail by Cochran (1953). Extensions to two-stage sampling with an example are to be found in Robson (1952) and Robson and King (1953).

Another type of double sampling scheme for the purpose of estimating a main character y is that in which a sample of size n' is observed for the x character. A random subsample of these of size $n < n'$ is observed for the y character as well. This subsample serves to determine the regression of the main character y on the auxiliary character x. The double sampling estimator of $\overline{Y}$, the y character mean, is

$$y_{ds} = \overline{y} + b(\overline{x}' - \overline{x})$$

where $\overline{x}'$ is the arithmetic mean of all n' observations on character x, $\overline{x}$ and $\overline{y}$ are the arithmetic means of the x and y characters from the subsample only, and b is the sample regression coefficient calculated from the subsample. This procedure is particularly useful in situations where the enumeration of the main character is costly but the correlated auxiliary variable can be readily measured. Double sampling can in fact be regarded mathematically as rotation sampling where the current occasion sample is a subsample of an earlier sample. Cochran (1939) mentions several examples of the use of double sampling to increase the precision of an estimator. Bose (1943) first considered the situation where the second sample of size n' is drawn independently of the first. Seal (1951) obtains estimates and their variances for sampling with or without replacement at either of the two phases for both a single auxiliary variate and many auxiliary variates.

A theory of double sampling in finite populations is given in Tikkiwal (1960). The finite population of size N is regarded as a random sample from an (infinite) bi-variate normal population (x_i, y_i). It is shown that the double sampling estimator Y_{ds} is an unbiased estimator of $\overline{Y}_N$, the population mean of the N y_i variates in an "extended sense," i.e.,

$$E(Y_{ds} | x_1, \ldots, x_N) = E(\overline{Y}_N | x_1, \ldots, x_N)$$

where $E(\ |\ )$ denotes conditional expectation. Further the variance $\hat{V}$ of Y_{ds} in the extended sense is defined by

$$\hat{V}(Y_{ds}) = E(Y_{ds} - \overline{Y}_N)^2 | x_1, \ldots, x_N) .$$

Expressions for $E(\hat{V}(Y_{ds}))$ and an unbiased estimator are presented. Extensions of this approach are described in Ajgaonkar and Tikkiwal (1961).

The problem of sampling on two consecutive occasions with a partial replacement of sampling units was first considered by Jessen (1942) in his analysis of a survey which collected farm data. The survey was designed so that of the n = 900 sampling units employed in the 1938 phase of the survey, 450 were retained for further observation in 1939. An additional 450 different units were selected to bring the 1939 sample up to strength. The 1939 sample was thus half independent of and half matched with the 1938 sample. He determined the efficiency of this incomplete matching relative to a completely independent selection of units on the second occasion as follows. Assuming a linear relationship to hold between observations on the same units in 1938 and 1939 for a given character, the adjusted mean

$$\bar{y}'_m = \bar{y}_m + b(\bar{x} - \bar{x}_m)$$

is an estimator of the population mean, $\bar{Y}$, per sample unit of the characteristic in 1939. Here $\bar{x}$ is the mean of all 900 units estimating the 1938 mean $\bar{X}$, $\bar{x}_m$ and $\bar{y}_m$ are, respectively, the 1938 and 1939 means of the 450 matched units, and b is the sample regression coefficient of y on x computed from the matched units. Combining $\bar{y}'_m$ with $\bar{y}_u$, the mean of the 450 unmatched 1939 sampling units, by weighting inversely as their variances ($\bar{y}'_m$ and $\bar{y}_u$ being independent) gives the weighted mean

$$\bar{y}_w = (\bar{y}'_m \sigma^2_{\bar{y}_u} + \bar{y}_u \sigma^2_{\bar{y}'_m}) \Big/ (\sigma^2_{\bar{y}_u} + \sigma^2_{\bar{y}'_m}) .$$

The variance of $\bar{y}_w$ is derived under the assumption of an infinite population and normality of the x variate. The normality requirement is not however needed provided that terms of higher order in $1/n$ may be neglected; this is equivalent to ignoring the variation in b which is of the order $1/n^2$. The estimator $\bar{y}_w$ was compared in efficiency with the unweighted mean $\bar{y}$ of all 900 units. For fourteen items in the questionnaire efficiency gains of from 22 to 45 per cent were achieved. Snedecor and King (1942) commented that Jessen's results might perhaps have been somewhat overoptimistic since one-third of the sampling units bore no farmsteads so that the x and y values would both be zero. This would lead to higher correlations than one might normally encounter. They also reported that partial matching techniques provided little additional information in the estimation of crop acreages with mail questionnaire returns compared with ratio-to-land estimation techniques. Jessen next raised the question of what the optimum match fraction would have been given that the 1938 sample had already been taken and that, for a given expenditure, the best possible estimate of the 1939 mean was required. Assuming the cost of matched and unmatched units to be the same, the optimum match fraction was shown to be approximately

$$m/u = 1/(1-\rho^2),$$

where m and u are the number of matched and unmatched units respectively and ρ is the coefficient of correlation between the 1938 and 1939 values of a character. He went on to consider the problem of allocating N sampling units to the first occasion and $u+m$ to the second so that (a) the variance of the sample mean was the same on each occasion and (b) that $N+u+m$ was a minimum for given sampling variances. His solution was unfortunately invalidated by an algebraic error.

The extension of Jessen's results to the situation where the population mean of the character is to be estimated on each of $h > 2$ occasions was considered by Yates (1949). He specified that (a) a fixed fraction $0 < \lambda < 1$ of the units was to be replaced on each occasion, (b) the population variance on each occasion and the correlation ρ between the same sampling unit on successive occasions were stationary, (c) an exponential correlation pattern of the type $\rho, \rho^2, \rho^3, \ldots,$ held between the same sampling units separated in time by one, two, three,..., occasions, (d) the correlation coefficient ρ was assumed to be known. The composite estimator considered was

$$\bar{y}'_h = (1 - Q_h)(\bar{y}_{h,h-1} + \rho(\bar{y}'_{h-1} - \bar{y}_{h-1,h})) + Q_n \bar{\bar{y}}_h ,$$

where $\bar{y}_{h,h-1}$ and $\bar{y}_{h-1,h}$ are the arithmetic means of the y character observations on occasions h and h-1 respectively computed from the units matched between occasions h and h-1 only, and $\bar{\bar{y}}_h$ is the sample mean on the h-th occasion of the unmatched units only. The optimum value of the weight coefficient Q_h which minimized the variance of $\bar{y}'_h$ was given as a function of ρ, λ, and the total number of occasions on which sampling had taken place. With increasing h, Q_h was observed to rapidly approach a limiting value which depends on ρ and λ only. Yates also discussed the estimation of change and the possibility of improving the composite estimator of the mean on the (h-1)-th occasion by using data provided on the h-th occasion as auxiliary information. All of the above results were given by Yates without proof. Both Cochran (1953) and Sukhatme (1954) supplied the missing details together with further extensions of the theory. In particular, the problem of selecting an optimum set of replacement fractions on the h-th occasion $(1, \lambda_1, \lambda_2, \ldots, \lambda_{h-1}, \lambda_h)$, when λ_i is not restricted to a constant value in time was first solved by Cochran (1953). As a working rule he recommended that to retain from one-quarter to one-third of the first sample on the second occasions and to thereafter employ a match fraction of one-half would serve as a good approximation to the optimum procedure. A recurrence relationship permitting the systematic evaluation of Q_h was derived by Narain (1953).

One of the classical papers on the theory of sampling on successive occasions is due to Patterson (1950). By limiting himself to the consideration of best linear unbiased estimators only he was able to exploit a property of such estimators and so quickly arrive at variances of a compact form. He first developed a set of necessary and sufficient conditions for a linear unbiased estimator to be a minimum-variance estimator as well. This theorem, which is not given here because it is not employed in this dissertation, is also quoted in its entirety by Eckler (1955); a set of covariance conditions is essentially involved. The results are applied to the problem of determining suitable estimates when sampling on successive occasions with partial replacement of units. An exponential-type correlation pattern is assumed to hold over time and population variances are taken to be equal on each occasion in the infinite populations. In the situation where the number of units and the replacement fraction are the same on all occasions he derives the efficient unbiased estimator of (a) the current occasion mean, (b) the change between the current and previous occasions, (c) the change between the current mean and the mean $k > 1$ occasions ago, and their respective variances. Modifications of the results when dealing with unequal numbers, population variances, or replacement fractions are indicated. The problem of arriving at an optimum replacement fraction for each occasion is considered. In order to minimize the sample size on the current occasion so that variances of the efficient estimators on the current and previous occasions are equal, a replacement fraction of one-half after the second occasion is found to be optimum. When ρ is not known or the correlation law is not of an exponential type the loss of efficiency in the estimation of means and the bias in the estimation of variances is investigated and found to be generally small.

Eckler (1955) simplified Patterson's approach to one-level rotation sampling to some extent by reducing the number of covariance conditions to be checked in Patterson's theorem. For two and three-level rotation designs iterative solutions for the minimum-variance estimates of the current mean are developed. The problem of improving past estimates by incorporating more recent data is discussed for the two-level case. A simple procedure yielding constant variances in an estimate over time with a two-level design is presented. He derives a criterion for deciding upon what level of rotation design to use by setting up a cost function involving the additional cost of securing past information as well as the current information on one visit.

Considerable attention has been devoted to the rotation sampling problem by Tikkiwal. In (1951) he develops the univariate theory for sampling on successive occasions by relating it to a lemma concerning the multiple regression of a p-th variate on $p - 1$ other variates, the joint distribution of all p being multivariate normal and the dispersion matrix of a somewhat more general form than that of Patterson. The possibility of improving the current occasion estimator by the inclusion of ancillary information provided by other characteristics correlated with the character under study is considered. Under a particular correlation structure he shows that the best linear unbiased estimator does not utilize such information. In a stratified sample design the optimum allocation of matched and unmatched units to strata which minimizes the estimate of the overall mean subject to a linear cost function and fixed total sample size on each occasion is derived. In the special case of constant population and sample sizes in each stratum the optimum match fraction is shown to be identically equal to one-half. If the correlation coefficient is also the same in each stratum the optimum allocation is in fact the familiar Neyman (1934) allocation. Somewhat more general results giving the optimum allocation as the Neyman allocation plus a correction term are presented in Tikkiwal (1953), the sample and population sizes being permitted to vary over time.

In his Ph.D. thesis Tikkiwal (1955a) examined the problem of establishing a suitable sampling scheme for the estimation of $k \geq 1$ characters on each of $h \geq 2$ occasions. The design was such that the i-th $(i = 2, 3, \ldots, k)$ character on any occasion is always studied on a portion of the sample on which the $(i-1)$-th character is measured. If the correlation between the i-th character on the r-th occasion and the j-th character on the s-th $(s \geq r)$ occasion is denoted by ρ_{rs}^{ij}, then the following correlation structure is assumed to hold:

(a) if $i = j,\ r \neq s,\ \rho_{rs}^{ii} = \prod_{t=r}^{s-1} \rho_{t,\,t+1} = \rho_{rs}$

where $\rho_{t,\,t+1}$ is the correlation between measurements on the same character on the t-th and $(t+1)$-th occasions;

(b) if $i \neq j$, $r = s$, $\rho_{rr}^{ij} = \prod_{t'=1}^{j-1} \rho'_{t', t'+1} = \rho'_{ij}$

where $\rho'_{t', t'+1}$ is the correlation between the t'-th and $(t'+1)$-th characters on the same occasion;

(c) If $i \neq j$, $r \neq s$, $\rho_{rs}^{ij} = \rho'_{ij} \cdot \rho_{rs}$.

The development of estimators and their variances in an infinite population hinges upon the assumption of a joint multivariate normal distribution for the k characters under the above correlation pattern. The theory is extended to cover finite populations of size N by employing a vague super-population concept. The normality assumption is not required since the estimators are shown to satisfy Patterson's (1950) conditions for a best estimator. The consequent variance and covariance relationships are utilized to derive variances of the estimators where correlation and regression coefficients are assumed known. The finite population results are also reported upon in Tikkiwal (1954) and (1955b). In a finite population it is possible to exhaust all of the sample units so that a subsample of new units cannot be selected. Tikkiwal is forced to assume that the units thus sampled are uncorrelated with the values they assumed when sampled some occasions earlier; the fact that they are treated as new units just the same contradicts the finite population assumption. It is the author's opinion that Tikkiwal's finite population theory is inadequate because (a) the super-population concept is unrealistic, (b) the sampling procedure is not carefully specified when the population is finite.

The variance of the minimum-variance estimator of the population mean is derived by Tikkiwal (1956b) when the various regression and correlation coefficients are estimated from the sample. In (1958a) he shows that whenever there is matching the efficiency of the best linear unbiased estimator relative to the simple mean increases with increasing time. The limiting value is approached more slowly for higher absolute values of the correlation coefficient but the convergence is still more rapid than that indicated by Yates (1949). In Tikkiwal (1958b) results for two-stage sampling are reported; it is assumed that the primary sample units of fixed size M are rotated and that a sample of size m is selected from each sampled primary.

Narain (1954) supplied a composite multiple regression type estimator of the mean on the h-th occasion which involved the arithmetic means on each prior occasion of those units still present on the h-th occasion. This estimator was claimed to be best but no proofs of any statements were provided.

The theory of sampling on successive occasions is applied by Tikkiwal (1956a) to data collected from a series of quarterly farm surveys concerning livestock production and marketing in the State of Iowa (Maki and Strand 1961). With some simplifying assumptions relating to the actual survey design efficiency gains ranging from 1% to 89% over the simple arithmetic mean were calculated for the various occasions and characters investigated.

The sampling and statistical aspects of the use of remeasured permanent plots and partial replacement of the initial sample for forest inventory are treated by Ware and Cunia (1962). The cost situation differs from that encountered in the sampling of human populations where the field costs would be expected to be lower if units are retained for a number of occasions. In forest inventory remeasured permanent plots will usually cost more than temporary ones because of the expense in marking, surveying and mapping permanent plots to facilitate identification at a later date. Optimum sampling plans under such a cost structure are derived. They recognized that although both growth and current volume are of equal interest in forest inventory, the optimum sampling plans for each do not coincide. A non-linear programming solution to the dilemma is presented. With respect to this last problem Mahalanobis (1952) had earlier suggested that a decision theory approach involving a risk function might perhaps be adopted. Cochran (1963) recommends the retention of 3/4 of the sample from occasion to occasion as a good practical solution to satisfying both requirements simultaneously.

Hansen et al. (1953) presented a simplified composite estimator to be used in sampling for a time series. They considered a two-level design with twelve independently selected random samples. One of the twelve is enumerated during January of each year, the second during February, and so on. If x_h and y_{h-1} represent the simple unbiased estimators of the population totals for the h-th and (h - 1)-th months from the h-th enumeration, then the composite estimator x'_h of the population total on the h-th occasion is given by

$$x'_h = Q(x'_{h-1} + x_h - y_{h-1}) + (1 - Q)x_h$$

where $0 \leq \Omega \leq 1$. Under some simplifying assumptions the variance of the monthly total x'_h, of the month-to-month change $x'_h - x'_{h-1}$, of the twelve month total $\sum_{i=1}^{12} x'_i$, and of the year-to-year change $x'_h - x'_{h-12}$ are derived.

The approach to rotation sampling in a finite population followed in this dissertation was inspired by the preliminary work of Oñate (1960c). He proposed a rotation plan for introduction into the Philippine Statistical Survey of Households for the prime purpose of minimizing the response resistance of panel households. The secondary sampling units (barrios) were to be split into segments. A rotation group within a barrio consisted of three segments; two segments were to be common from visit to visit and one segment common from year to year. A finite population theory was developed for the special case of five segments in a barrio by considering the $5!$ possible orderings of the design and the Hansen et al. (1955) composite estimator. Accounts of the theory may be found in Oñate (1960a) and Oñate (1960b) as well.

Rao (1962b) further developed the theory of rotation sampling from a finite population of arbitrary size N for a one cycle pattern. Since the theory presented in this dissertation is an extension of this preliminary research and contains the one cycle design as a special case, no details are given here on Rao's work. Schach (1962) carried out a numerical investigation of Rao's formula in order to obtain approximations to the optimum values of the weight coefficients in the composite estimator and to the optimum number of visits by a unit. Specific reference to preliminary results obtained in this dissertation is given in Rao and Graham (1962).

In February of 1954 the U.S. Bureau of the Census instituted design revisions into the Current Population Survey (C.P.S.). This monthly survey compiles information on employment and unemployment and related data; information on other national and regional characteristics such as income distribution, marital status, migration and education are compiled at less frequent intervals. Hansen et al. (1953) contains a full account of the sample design as it existed before the new features were implemented. At that time a rotation of sampling units was employed for the sole purpose of reducing the nonresponse rate. There were administrative advantages in introducing new units on a staggered basis since substantial costs are involved in introducing a household into the sample for the first time. Hansen et al. (1955) give a comprehensive summary of the new design features of the C.P.S. Essentially, a rotation group remains in the sample for four consecutive months, drops out of the sample for the next eight months, and returns for another four months. It then drops out of the sample completely and does not return again. The composite estimator developed was of the form

$$X''_h = Q(X''_{h-1} + X'_{h,h-1} - X'_{h-1,h}) + (1 - Q)X'_h$$

where $0 \leq Q \leq 1$,

X''_h is the composite estimator for month h,

X'_h is the regular ratio estimator based on the entire sample for month h,

$X'_{h,h-1}$ is the regular ratio estimator for month h made from segments that are in the sample for both months h and h-1,

$X'_{h-1,h}$ is the regular ratio estimate for the previous month (h-1) made from segments that are in the sample for both months h and h-1.

The developments in Chapter III of this dissertation will be based upon a composite estimator of the same general structure as that of X''_h.

A description of the sample for the U.S. Monthly Retail Trade Report which includes a discussion of the composite estimator and its variance is given by Kailin (1955). Neter and Waksberg (1961) report on a four visit rotation design which permitted the study of the effects of different interviewing techniques on the same households over the course of time. The design was also advantageous in that the estimation of differences between occasions was also of major concern.

The Monthly Retail Trade Survey, conducted monthly throughout the United States to report on various characteristics of retail stores, is an example of two-level rotation sampling (Woodruff (1959)). It employs a stratified two-stage design with a rotation feature supplemented by a fixed list sample of the largest retail outlets. One primary sampling unit, a county or county-group, is selected from each of 230 strata covering the entire United States. The secondary sampling units are area segments containing an everage of four retail stores and are selected at an over-all rate of six per cent. The sample is subsequently divided into twelve panels, each panel being a one-half per cent sample from the population. The panel is completely rotated each month so that the same panel is interviewed on the same month each year. At the time of interviewing for the i-th month information concerning the (i - 1)-th month is also gathered. Such a design is obviously more efficient than that employed by the Current Population Survey since an entirely new panel is available each month in addition to a complete panel match with the previous month. In the C. P. S. sampling situation, however, two-level sampling would be inadviseable due to the possibility of rather severe memory biases occurring. The Monthly Retail Trade Survey features a preliminary composite estimator published for the h-th month based on data available up to that time. A subsequent revised composite estimate is issued a month later which utilizes information furnished on the (h + 1)-th month pertaining to the h-th month as well. The gain in efficiency in so doing is believed to outweigh the possible inconvenience caused by two sets of estimates.

Bershad and Nisselson (1962) explored the feasibility of using a pattern of weekly surveys rather than a single monthly survey based on a systematic sample of weeks. They assumed that a respondent could provide adequate information covering the previous two time periods (weeks) but that any longer recall period would not yield satisfactory data. It was also assumed that monthly statistics were of prime importance and these would be published at the end of each month. Four rotation sampling plans were considered: (a) "50 - 75 - 50 Plan" characterized by a 50 per cent overlap in information from week to week, a 75 per cent sample overlap from month to month and a 50 per cent sample overlap from year to year. This is accomplished by collecting two weeks data at each interview and interviewing a given unit for one week of each month for four months, dropping it out of the sample for the next eight months, and then repeating the procedure for the next four months; (b) "50 - 0 - 100 Plan" characterized by a 50 per cent overlap in information from week to week, no monthly sample overlap and a 100 per cent yearly overlap. Only one interview per year is taken; (c) "50 - 50 - 100 Plan" characterized by a 50 per cent weekly information overlap with 50 per cent monthly and 100 per cent yearly sample overlaps. After two interviews one month apart a unit drops out of the sample for the next ten months and then returns for another two months; (d) "X - 75 - 50 Plan" with no overlap in information. This is the C.P.S. sample design described earlier. Numerical analyses showed that the first three plans were superior to the fourth in supplying monthly averages but that each of the four plans displayed certain advantages when estimating changes over various time intervals.

In view of the extensive use of rotation sampling as substantiated by the above references, it would appear that any further research conducted in this area would certainly not be superfluous.

III. ROTATION SAMPLING WITH INFINITELY MANY CYCLES

A. The General Rotation Pattern

Consider a population P whose size N remains fixed over time, with no units immigrating into or migrating from P. Samples of size n are selected from P on occasions 0, -1, -2, ..., where 0 denotes the most recent or current occasion, according to the following rotation pattern. A rotation group consisting of $n_2\ (\geq 1)$ units remains in the sample for $r \geq 2$ consecutive occasions so that $n = rn_2$. It then leaves the sample for m occasions, returns for another r occasions, drops out for m more occasions, and so on. Any rotation group is said to make infinitely many cycles in order to distinguish the pattern from other designs to be considered later. The maximum value of m is obviously $r(N/n - 1)$. If m is less than its maximum value then the rotation is, in fact, taking place within a subpopulation of size $N' = (m + r)n_2$ from P. It will be assumed initially that the rotation is imposed on all N units of P and hence that $N = (m + r)n_2$. This restriction will be removed in D. The case $m \geq r$ will be considered in detail; the case $m < r$ is exceedingly more complex and less useful in practice and a discussion has been relegated to the Appendices. An example of the foregoing rotation pattern is presented in Figure 1.

The general rotation pattern is established in the following manner. The units in P are randomly assigned integers from 1 to N. The first n_2 units comprise the first rotation group, the second n_2 units the second rotation group, etc. There are a total of N! possible rotation patterns that can be so formed by taking all permutations of the

N units. Any given unit will appear in a given rotation group in $(N-1)!$ of these randomizations since it necessarily appears in every group an equal number of times. The particular rotation pattern observed is therefore one random pattern selected from a finite population of $N!$ rotation patterns.

Unit number	Occasion -6	-5	-4	-3	-2	-1	0
1	X				X	X	
2	X	X				X	X
3		X	X				X
4			X	X			
5				X	X		

Figure 1. General rotation pattern with $N = 5$, $n = r = 2$, and $m = 3$

B. The Simple Composite Estimator of the Current Occasion Mean

Let $x_{\alpha, k}$ denote the value of the characteristic being measured for the k-th unit on the α-th occasion $(\alpha = 0, -1, \ldots, -u$, and $k = 1, 2, \ldots, N)$, where $-u + 1$ is the occasion on which a composite estimator of the current occasion mean is first employed. It will be assumed that sampling has taken place for several occasions in the past so that u is large. This will introduce certain simplifications in deriving variances in subsequent sections.

The simple composite estimator of the current occasion population mean, $\overline{X}_0$, is

$$\bar{x}_0' = Q(\bar{x}'_{-1} + \bar{x}_{0,-1} - \bar{x}_{-1,0}) + (1 - Q)\bar{x}_0 , \tag{3.1}$$

where $$0 \le Q \le 1 ,$$

and

$$\bar{x}_{a,a-1} = \sum_{k=1}^{n_1} x_{a,k}/n_1 , \quad \bar{x}_{a-1,a} = \sum_{k=1}^{n_1} x_{a-1,k}/n_1 , \quad \bar{x}_a = \sum_{k=1}^{n} x_{a,k}/n ,$$

and $\bar{x}'_{-1}$ is the simple composite estimator for occasion -1. Here $n_1 = (r-1)n_2$ is the number of units common to occasions $a-1$ and a, and n_2 is the number of units entering the sample for a first visit of some cycle on occasion a.

Now $\bar{x}_0'$ can be written as

$$\bar{x}_0' = \sum_{a=0}^{-u} Q^{-a} W_a = \sum_{a=0}^{-u} \sum_{k=1}^{N} w_{a,k} x_{a,k} \tag{3.2}$$

where

$$W_a = Q(\bar{x}_{a,a-1} - \bar{x}_{a-1,a}) + (1 - Q)\bar{x}_a \tag{3.3}$$

for $a = 0, -1, \ldots, -u+1$, and

$$W_{-u} = \bar{x}_{-u} . \tag{3.4}$$

The weights $w_{a,k}$ are functions of Q, r and n_2.

From (3.3) and (3.4) it is seen that the weights $w_{a,k}$ are as follows:

(1) For $a = 0$ (current occasion),

(a) $w_{0,k} = (1-Q)/n$ for the n_2 units on the 1-st visit of a cycle,

(b) $w_{0,k} = (1-Q)/n + Q/n_1 = (1 + n_2 Q/n_1)/n$ for the n_1 units on the 2-nd to r-th visit of a cycle, (3.5)

(c) $w_{0,k} = 0$ for the N-n units not in the sample.

(2) For $a = -1, -2, \ldots, -u+1$,

(a) $w_{a,k} = Q^{-a}(1-Q)/n - Q^{-a}/n_1 = - Q^{-a}(n_2/n_1 + Q)/n$ for the n_2 units on the 1-st visit of a cycle,

(b) $w_{a,k} = Q^{-a}(1-Q)/n + Q^{-a+1}/n_1 - Q^{-a}/n_1 = Q^{-a} n_2(Q-1)/nn_1$ for the $n_1 - n_2$ units on the 2-nd to (r-1)-th visits of a cycle, (3.6)

(c) $w_{a,k} = Q^{-a}(1-Q)/n + Q^{-a+1}/n_1 = Q^{-a}(n_2 Q/n_1 + 1)/n$ for the n_2 units on the r-th visit of a cycle,

(d) $w_{a,k} = 0$ for the N-n units not in the sample.

(3) For $a = -u$,

(a) $w_{-u,k} = Q^u(1/n - 1/n_1)$ for the n_1 units on 1-st to (r-1)-th visits of a cycle,

(b) $w_{-u,k} = Q^u/n$ for n_2 units on the r-th visit of a cycle, (3.7)

(c) $w_{-u,k} = 0$ for the N-n units not in the sample.

Let E denote expectation over the N! possible rotation patterns. Then

$$E(\bar{x}_0') = E\left(\sum_{a=0}^{-u}\sum_{k=1}^{N} w_{a,k}\, x_{a,k}\right) = \sum_{k=1}^{N} x_{0,k}E(w_{0,k}) + \sum_{a=-1}^{-u}\sum_{k=1}^{N} x_{a,k}E(w_{a,k}).$$

Now

$$E(w_{0,k}) = \sum_{k=1}^{N}(N-1)!\, w_{0,k}/N! = \sum_{k=1}^{N} w_{0,k}/N = 1/N,$$

$$E(w_{a,k}) = \sum_{k=1}^{N}(N-1)!\, w_{a,k}/N! = \sum_{k=1}^{N} w_{a,k}/N = 0, \quad a < 0.$$

Hence

$$E(\bar{x}_0') = \sum_{k=1}^{N} x_{0,k}/N = \bar{X}_0,$$

and $\bar{x}_0'$ is an unbiased estimator of $\bar{X}_0$.

C. The Variance of the Simple Composite Estimator of the Current Occasion Mean

1. A general variance formula

In order to simplify the derivation of the variance of $\bar{x}_0'$, $V(\bar{x}_0')$, it sill be assumed that u is large so that

$$\bar{x}_0' = \sum_{a=0}^{-u}\sum_{k=1}^{N} w_{a,k}\, x_{a,k} \doteq \sum_{a=0}^{-\infty}\sum_{k=1}^{N} w_{a,k}\, x_{a,k}.$$

The error so introduced into the variance function will be negligible if u is at least moderately large since the weight variables $w_{a,k}$ decrease exponentially as a becomes more and more negative. We will henceforth use strict equality signs assuming that -u is effectively at $-\infty$. Now

$$V(\bar{x}'_0) = E(\bar{x}'_0)^2 - \bar{X}_0^2$$

$$= \sum_{a=0}^{-\infty} \sum_{k=1}^{N} E(w^2_{a,k}) x^2_{a,k} + \sum_{\substack{a \neq a' \\ =0}}^{-\infty} \sum_{k=1}^{N} E(w_{a,k} w_{a',k}) x_{a,k} x_{a'k} - \bar{X}_0^2$$

$$+ \sum_{a=0}^{-\infty} \sum_{\substack{k \neq k' \\ =1}}^{N} E(w_{a,k} w_{a,k'}) x_{a,k} x_{a,k'} + \sum_{\substack{a \neq a' \\ =0}}^{-\infty} \sum_{\substack{k \neq k' \\ =1}}^{N} E(w_{a,k} w_{a',k'}) x_{a,k} x_{a',k'}. \tag{3.8}$$

Since $\sum_{k=1}^{N} w_{0,k} = 1$ and $\sum_{k=1}^{N} w_{a,k} = 0$ for $a < 0$, it follows that

$$E(w_{0,k} w_{0,k'}) = \sum_{\substack{k \neq k' \\ =1}}^{N} w_{0,k} w_{0,k'} / N(N-1) = 1/N(N-1) - E(w^2_{0,k})/(N-1),$$

$$E(w_{a,k} w_{a,k'}) = -E(w^2_{a,k})/(N-1), \quad a < 0, \tag{3.9}$$

$$E(w_{a,k} w_{a',k'}) = -E(w_{a,k} w_{a',k})/(N-1), \quad a = 0, -1, -2, \ldots, \quad .$$

Substituting (3.9) into (3.8) gives

$$V(\bar{x}'_0) = \sum_{a=0}^{-\infty} \sum_{k=1}^{N} x^2_{a,k} E(w^2_{a,k}) + \sum_{\substack{a \neq a' \\ =0}}^{-\infty} E(w_{a,k} w_{a',k}) \sum_{k=1}^{N} x_{a,k} x_{a',k}$$

$$- \sum_{a=-1}^{-\infty} E(w^2_{a,k}) \sum_{\substack{k \neq k' \\ =1}}^{N} x_{a,k} x_{a,k'} / (N-1) + \sum_{\substack{k \neq k' \\ =1}}^{N} x_{0,k} x_{0,k'} / N(N-1) - \bar{X}_0^2$$

$$- E(w_{0,k}^2) \sum_{\substack{k \neq k' \\ =1}}^{N} x_{0,k} x_{0,k'}/(N-1) - \sum_{\substack{a \neq a' \\ =0}}^{-\infty} E(w_{a,k} w_{a',k}) \sum_{\substack{k \neq k' \\ =1}}^{N} x_{a,k} x_{a',k'}/(N-1). \tag{3.10}$$

Let S_a^2 be the mean square for the a-th occasion and $S_{a,a'}$ be the mean product between occasions a and a',

$$S_a^2 = (\sum_{k=1}^{N} x_{a,k}^2 - N\overline{X}_a^2)/(N-1),$$

$$S_{a,a'} = (\sum_{k=1}^{N} x_{a,k} x_{a',k} - N\overline{X}_a \overline{X}_{a'})/(N-1), \tag{3.11}$$

where $\overline{X}_a$ is the population mean on the a-th occasion. Then (3.10) becomes

$$V(\overline{x}_0') = (NE(w_{0,k}^2) - 1/N)S_0^2 + N \sum_{a=-1}^{-\infty} E(w_{a,k}^2) S_a^2$$

$$+ N \sum_{\substack{a \neq a' \\ =0}}^{-\infty} E(w_{a,k} w_{a',k}) S_{a,a'} . \tag{3.12}$$

It is worthwhile noting that (3.12) is a general formula for the variance of any estimator that can be written in the form

$\sum_{a=0}^{-\infty} \sum_{k=1}^{N} w_{a,k} x_{a,k}$ where the $w_{a,k}$ are any weights which satisfy the

conditions $\sum_{k=1}^{N} w_{0,k} = 1$ and $\sum_{k=1}^{N} w_{a,k} = 0$ for $a < 0$.

2. Infinite cycle variance with unspecified correlation structure

The weights for the general infinite cycle rotation pattern are given in (3.5) and (3.6). It is easily verified that

$$NE(w_{0,k}^2) = \sum_{k=1}^{N} w_{0,k}^2 = (1 + n_2 Q^2 / n_1)/n ,$$

$$NE(w_{a,k}^2) = \sum_{k=1}^{N} w_{a,k}^2 = Q^{-2a}(Q^2 + 2n_2 Q/n_1 + 1)n_2/nn_1 , \quad a < 0.$$

A careful consideration of the rotation pattern is necessary when evaluating the cross-product expectations $E(w_{a,k} w_{a',k})$. The term $E(w_{0,k} w_{-\ell(r+m)-1,k})$, $\ell = 1, 2, \ldots,$ is developed in some detail in (a) below so that the procedure will be clear. Final results only are quoted for other terms.

(a) $$NE(w_{0,k} w_{-\ell(r+m)-1,k}) = \sum_{k=1}^{N} w_{0,k} w_{-\ell(r+m)-1,k}$$

$$= \Big[(1-Q)/n\Big]\Big[0\Big] n_2 - \Big[(1+n_2 Q/n_1)/n\Big]\Big[Q^{\ell(r+m)+1} (n_2/n_1 + Q)/n)\Big] n_2$$

$$+ \Big[(1+n_2 Q/n_1)/n\Big]\Big[Q^{\ell(r+m)+1} n_2(Q-1)/nn_1\Big] (n_1 - n_2)$$

$$= - Q^{\ell(r+m)+1} (1+n_2 Q/n_1)^2 n_2/n^2 , \qquad \ell = 0, 1, 2, \ldots,$$

for if (1) if a unit is on the first visit of a cycle on occasion 0 it is necessarily out of the sample on occasion $-\ell(r+m)-1$, (2) if a unit is on the 2nd visit of a cycle on occasion 0 it is necessarily on the first visit of a cycle on occasion $-\ell(r+m)-1$, (3) if a unit is on the 3rd, 4th, ..., r-th visits of a cycle on occasion 0, it is necessarily on the

2-nd, 3-rd, ..., (r-1)-th visit respectively of a cycle on occasion $-\ell(r+m)-1$. Thus

(a) $$N \sum_{\ell=0}^{\infty} E(w_{0,k} w_{-\ell(r+m)-1,k}) S_{0,-\ell(r+m)-1}$$

$$= - \sum_{\ell=0}^{\infty} Q^{\ell(r+m)+1} (1+n_2 Q/n_1)^2 \, n_2 \, S_{0,-\ell(r+m)-1}/n^2 .$$

(b) $$N \sum_{\ell=1}^{\infty} E(w_{0,k} w_{-\ell(r+m),k} S_{0,-\ell(r+m)}$$

$$= \sum_{\ell=1}^{\infty} n_2(n_2 + Qn_1) Q^{\ell(r+m)+1} \, S_{0,-\ell(r+m)}/nn_1^2 .$$

(c) $$N \sum_{\ell=0}^{\infty} \sum_{t=-2}^{-r+2} E(w_{0,k} w_{-\ell(r+m)+t,k}) S_{0,-\ell(r+m)+t}$$

$$= \sum_{\ell=0}^{\infty} Q^{\ell(r+m)} \sum_{t=-2}^{-r+2} Q^{-t} n_2^2(1+n_2 Q/n_1)(t(Q-1)-r) S_{0,-\ell(r+m)+t}/n_1 n^2 .$$

(d) $$N \sum_{\ell=0}^{\infty} E(w_{0,k} w_{-\ell(r+m)-r+1,k}) S_{0,-\ell(r+m)-r+1}$$

$$= - \sum_{\ell=0}^{\infty} Q^{\ell(r+m)+r-1} n_2(Q+n_2/n_1)(n_2 Q/n_1 +1) S_{0,-\ell(r+m)-r+1}/n^2 .$$

It may be verified that (a), (c) and (d) sum to

$$\sum_{\ell=0}^{\infty} Q^{\ell(r+m)} n_2^2(1+n_2 Q/n_1) \sum_{t=1}^{r-1} Q^t (t(1-Q)-r) S_{0,-\ell(r+m)-t}/n_1 n^2 .$$

(e) $N \sum_{\ell=0}^{\infty} E(w_{0,k} w_{-\ell(r+m)-m-1,k}) S_{0,-\ell(r+m)-m-1}$

$$= \sum_{\ell=0}^{\infty} Q^{\ell(r+m)+m+1} n_2(1+n_2 Q/n_1)(1-Q) S_{0,-\ell(r+m)-m-1}/n^2 .$$

(f) $N \sum_{\ell=0}^{\infty} \sum_{t=0}^{m-r} E(w_{0,k} w_{-\ell(r+m)-r-t,k}) S_{0,-\ell(r+m)-r+t} = 0$,

for if a unit is in the sample on occasion 0 it is necessarily out of the sample on occasions $-\ell(r+m)-r$, $-\ell(r+m)-r-1, \ldots, -\ell(r+m)-m$.

(g) $N \sum_{\ell=0}^{\infty} \sum_{t=2}^{r-1} E(w_{0,k} w_{-m-t-\ell(r+m),k}) S_{0,-m-t-\ell(r+m)}$

$$= \sum_{\ell=0}^{\infty} Q^{\ell(r+m)} n_2 \sum_{t=2}^{r-1} Q^{m+t} \Big[n(1+n_2 Q/n_1)/n_1 - n_2 n Q(Q-1)/n_1^2$$

$$+ n_2(Q-1)(1+n_2 Q/n_1)t/n_1 \Big] S_{0,-\ell(r+m)-m-t}/n^2 .$$

(h) $N \sum_{a=-1}^{-\infty} \sum_{\ell=0}^{\infty} \sum_{t=1}^{r-2} E(w_{a,k} w_{a-t-\ell(r+m),k}) S_{a,a-t-\ell(r+m)}$

$$= - \sum_{a=-1}^{-\infty} \sum_{\ell=0}^{\infty} \sum_{t=1}^{r-2} n_2^3 (1-Q)^2 \, t Q^{-2a+t} Q^{\ell(r+m)} S_{a,a-t-\ell(r+m)}/n^2 n_1^2 .$$

(i) $N \sum_{a=-1}^{-\infty} \sum_{\ell=0}^{\infty} E(w_{a,k} w_{a-r+1-\ell(r+m),k}) S_{a,a-r+1-\ell(r+m)}$

$$= - \sum_{a=-1}^{-\infty} \sum_{\ell=0}^{\infty} Q^{-2a+r-1} n_2(1+n_2 Q/n_1)(Q+n_2/n_1) Q^{\ell(r+m)}$$

$$S_{a,a-r+1-\ell(r+m)}/n^2 .$$

$$\text{(j)}\quad N \sum_{a=-1}^{-\infty} \sum_{\ell=0}^{\infty} \sum_{t=0}^{m-r} E(w_{a,k} w_{a-\ell(r+m)-r-t,k}) S_{a,a-\ell(r+m)-r+t} = 0 .$$

$$\text{(k)}\quad N \sum_{a=-1}^{-\infty} \sum_{\ell=0}^{\infty} E(w_{a,k} w_{a-m-1-\ell(r+m),k}) S_{a,a-m-1-\ell(r+m)}$$

$$= -n_2(1+n_2 Q/n_1)(Q+n_2/n_1) \sum_{a=-1}^{-\infty} \sum_{\ell=0}^{\infty} Q^{-2a+m+1+\ell(r+m)}$$

$$S_{a,a-m-1-\ell(r+m)} /n^2 .$$

$$\text{(l)}\quad N \sum_{a=-1}^{-\infty} \sum_{\ell=0}^{\infty} \sum_{t=2}^{r-1} E(w_{a,k} w_{a-m-t-\ell(r+m),k}) S_{a,a-m-t-\ell(r+m)}$$

$$= \sum_{a=-1}^{-\infty} \sum_{\ell=0}^{\infty} \sum_{t=2}^{r-1} n_2^3 (t-r)(Q-1)^2 Q^{-2a+m+t+\ell(r+m)}$$

$$S_{a,a-m-t-\ell(r+m)} /n^2 n_1^2 .$$

$$\text{(m)}\quad N \sum_{a=-1}^{-\infty} \sum_{\ell=1}^{\infty} E(w_{a,k} w_{a-\ell(r+m),k}) S_{a,a-\ell(r+m)}$$

$$= \sum_{a=-1}^{-\infty} n_2(Q^2+2n_2 Q/n_1+1) Q^{-2a} \sum_{\ell=1}^{\infty} Q^{\ell(r+m)}$$

$$S_{a,a-\ell(r+m)} /nn_1 .$$

Substitution of the preceding terms into (3.12) yields

$$V(\bar{x}_0') = (1/n - 1/N)S_0^2 + Q^2 n_2 S_0^2/(nn_1)$$

$$+ 2n_2(1+n_2 Q/n_1)(1-Q)Q^{m+1} \sum_{\ell=0}^{\infty} Q^{\ell(r+m)} S_{0,-\ell(r+m)-m-1} /n^2$$

$$+ 2n_2(n_2+Qn_1)Q \sum_{\ell=1}^{\infty} Q^{\ell(r+m)} S_{0,\,-\ell(r+m)}/(nn_1^2)$$

$$+ 2n_2^2 \sum_{t=2}^{r-1} Q^{m+t} \left[(rn_1+2nQ-nQ^2)/n_1 + (Q-1)(1+n_2Q/n_1)t \right]$$

$$\sum_{\ell=0}^{\infty} Q^{\ell(r+m)} S_{0,\,-m-t-\ell(r+m)}/(n^2 n_1)$$

$$+ 2 \sum_{\ell=0}^{\infty} n_2^2(1+n_2Q/n_1)Q^{\ell(r+m)} \sum_{t=1}^{r-1} Q^t(t(1-Q)-r)$$

$$S_{0,\,-\ell(r+m)-t}/(n_1 n^2)$$

$$+ n_2(Q^2+2n_2Q/n_1+1) \sum_{a=-1}^{-\infty} Q^{-2a} S_a^2/(nn_1)$$

$$+ 2 \sum_{a=-1}^{-\infty} n_2(Q^2+2n_2Q/n_1+1)Q^{-2a} \sum_{\ell=1}^{\infty} Q^{\ell(r+m)} S_{a,\,a-\ell(r+m)}/(nn_1)$$

$$- 2n_2^3(1-Q)^2 \sum_{a=-1}^{-\infty} \sum_{t=1}^{r-2} tQ^{-2a+t} \sum_{\ell=0}^{\infty} Q^{\ell(r+m)} S_{a,\,a-t-\ell(r+m)}/(n^2 n_1^2)$$

$$- 2n_2(1+n_2Q/n_1)(Q+n_2/n_1) \sum_{a=-1}^{-\infty} Q^{-2a+r-1} \sum_{\ell=0}^{\infty} Q^{\ell(r+m)}$$

$$S_{a,\,a-r+1-\ell(r+m)}/n^2$$

$$- 2n_2(1+n_2Q/n_1)(Q+n_2/n_1) \sum_{a=-1}^{-\infty} Q^{-2a+m+1} \sum_{\ell=0}^{\infty} Q^{\ell(r+m)}$$

$$S_{a,\,a-m-1-\ell(r+m)}/n^2$$

$$+ 2n_2^3(1-Q)^2 \sum_{a=-1}^{-\infty} \sum_{t=2}^{r-1} (t-r)Q^{-2a+m+t} \sum_{\ell=0}^{\infty} Q^{\ell(r+m)}$$

$$S_{a,\, a-m-t-\ell(r+m)}/(n^2 n_1^2) . \qquad (3.13)$$

Equation (3.13) will be valid for all $r \geq 2$ if the convention that $\sum_a^b (\ ) = 0$ if $b < a$ is adopted.

3. Infinite cycle variance with Markoff type correlation structure

Assume now that

$$S_a^2 = S_0^2, \quad S_{a,\, a+t} = S_{0,\, t} \qquad (3.14)$$

and also that a Markoff type correlogram holds, i.e., that

$$S_{0,\, t} = \rho^{-t} S_0^2, \quad (t = -1, -2, -3, \ldots) . \qquad (3.15)$$

Then, after a great deal of tedious algebra, (3.13) reduces to

$$V(\bar{x}_0') = (1/n - 1/N)S_0^2 + 2n_2^3 Q S_0^2 \Big\{ Q^2(r\rho^2 - (r^2+1)\rho + r)$$

$$+ Q(r(r-1)\rho^2 - 2(r-1)\rho + r(r-1)) - (r-1)^2\rho + Q^r \rho^{r-1} \Big[Q^2(-(r-1)\rho^2$$

$$+ r(r-1)\rho) + Q(-(r^2 - 2r + 2)\rho^2 + 2r\rho - r^2) - (r-1)\rho^2 + r(r-1)\rho \Big]$$

$$+ Q^m \rho^{m+1} \Big[Q^3(-r^2\rho^2 + r(r+1)\rho - r) + Q^2(2r(r-1)\rho - (r-1)(r+1))$$

$$+ Q(-r(r-1)\rho - (r-1)(r-2)) + (r-1)^2 \Big] + Q^{m+r} \rho^{m+r} \Big[Q^3(r(r-1)\rho^2$$

$$- r(r-1)\rho) + Q^2(r\rho^2 + (-2r^2 + r - 1)\rho + r^2) + Q((r-1)(r-2)\rho$$

$$+ r(r-1)) + (r-1)\rho + r(r-1) \Big] \Big\} / n^2 n_1^2 (1 - Q^2)(1 - Q\rho)^2 (1 - (Q\rho)^{r+m}),$$

(3.16)

which is valid for all $r \geq 2$.

It is noteworthy that in equations (3.13) and (3.16) the finite population correction factor effects only the current occasion term. This is in agreement with Tikkiwal (1955a) who reached the same conclusion using a super-population argument.

It should be emphasized that (3.13) and (3.16) are valid only when $m \geq r$. The case $m < r$ is more difficult and is dealt with in the Appendices.

Rather than constant $S_\alpha = S_0$ as in (3.14) one could more generally assume the presence of a geometric trend in S_α,

$$S_\alpha \equiv S_0 k^\alpha ,$$

where k is either greater than, equal to, or less than unity. The case of constant S_α^2 is therefore included. The variance formula appropriate to the case $k \neq 1$ is reproduced in the Appendices and no further reference will be made to it.

4. The static population check on the calculations

Because of the unduly heavy calculations involved in arriving at the final form (3.16) of $V(\overline{x}_0')$, it was deemed essential that some check be performed to ascertain the correctness of the formula. A rather obvious check is that when $Q = 0$ in (3.1) then $\overline{x}_0' = \overline{x}_0$ and hence $V(\overline{x}_0') = (1/n - 1/N)S_0^2$. Setting $Q = 0$ in (3.16) gives the same result.

A more searching check is achieved under the assumption that the population is static over time. Then $x_{0,k} = x_{-1,k} = x_{-2,k} = \dots,$ for $k = 1, 2, \dots, N$. Thus a stationary covariance structure holds good and the coefficient of correlation ρ is unity. It follows that $\rho^{\alpha} = 1$ for all α so that the exponential correlation model is valid in a trivial sense. $V(\bar{x}_0')$ may now be derived directly. Since $\bar{x}_{0,-1} = \bar{x}_{-1,0}$, it follows that

$$\bar{x}_0' = (1-Q) \sum_{t=0}^{\infty} Q^t \bar{x}_{-t} . \tag{3.17}$$

Hence

$$V(\bar{x}_0') = (1-Q)^2 \sum_{t=0}^{\infty} Q^{2t} V(\bar{x}_{-t}) + 2 \sum_{\substack{t<\tau \\ =0}}^{\infty} Q^t Q^{\tau} \operatorname{Cov}(\bar{x}_{-t}, \bar{x}_{-\tau}). \tag{3.18}$$

Now the variance of a random observation $x_{\alpha,k}$ from a population P of size N is by definition

$$V(x_{\alpha,k}) = \sigma_0^2 = (N-1) S_0^2 / N , \tag{3.19}$$

and the covariance between two observations $x_{\alpha,j}$ and $x_{\alpha,k}$, $(j \neq k)$, selected at random and without replacement from P is

$$\operatorname{Cov}(x_{\alpha,j}, x_{\alpha,k}) = -\sigma_0^2/(N-1) = -S_0^2/N . \tag{3.20}$$

The latter is a familiar result and no proof will therefore be given. With the aid of (3.19) and (3.20) it may be verified that

$$V(\bar{x}_{-t}) = (1/n - 1/N) S_0^2 , \qquad (t = 0, -1, -2, \dots,) ,$$

$$\mathrm{Cov}(\bar{x}_{-t}, \bar{x}_{-t-s-\ell(r+m)}) = (n_2(r-s)/n^2 - 1/N)S_0^2\,,$$

$$(1 \leq s < r,\ \ell = 0, 1, 2, \ldots,),$$

$$= -S_0^2/N, \qquad (r \leq s \leq m,\ \ell = 0, 1, 2, \ldots,),$$

$$= ((s-m)n_2/n^2 - 1/N)S_0^2, \qquad (m < s \leq r+m,\ \ell = 0, 1, 2, \ldots,).$$

Substitution of these values into (3.18) gives

$$V(\bar{x}_0') = S_0^2(1-Q)^2 \Big(\sum_{t=0}^{\infty} Q^{2t}(1/n - 1/N)$$

$$+ 2\sum_{t=0}^{\infty}\sum_{s=1}^{r-1}\sum_{\ell=0}^{\infty} Q^{2t+s+\ell(r+m)}((r-s)n_2/n^2 - 1/N)$$

$$- 2\sum_{t=0}^{\infty}\sum_{s=0}^{m-r}\sum_{\ell=0}^{\infty} Q^{2t+r+s+\ell(r+m)}/N$$

$$+ 2\sum_{t=0}^{\infty}\sum_{s=1}^{r}\sum_{\ell=0}^{\infty} Q^{2t+m+s+\ell(r+m)}(sn_2/n^2 - 1/N)\Big), \tag{3.21}$$

$$= (1/n - 1/N)S_0^2 + 2n_2S_0^2(-Q + 2Q^2 - Q^3 + Q^{r+1} - 2Q^{r+2} + Q^{r+3} + Q^{m+1}$$

$$-2Q^{m+2} + Q^{m+3} - Q^{m+r+1} + 2Q^{m+r+2} - Q^{m+r+3})/n^2(1-Q^2)(1-Q)^2(1-Q^{r+m})$$

$$= (1/n - 1/N)S_0^2 + 2n_2QS_0^2(-1+Q^r+Q^m-Q^{m+r})/n^2(1-Q^2)(1-Q^{r+m}). \tag{3.22}$$

Putting $\rho = 1$ in (3.16) will reduce it to (3.22), thereby providing the check.

The static population assumption is employed throughout this dissertation in verifying formulae which require extended algebraic simplifications. It is a penetrating check on the validity of the numerical coefficients and on the exponents of Q in such relationships. That the exponents of ρ are not established is perhaps the sole barrier to the infallibility of the method.

5. One cycle variance as a special case

A one cycle rotation pattern may be regarded as a special case of an infinite cycle rotation design by taking $m = \infty$. A rotation group therefore remains in the sample for r consecutive occasions and drops out for $m = \infty$ consecutive occasions before returning for a second cycle, i.e., it never returns. Under the covariance structure specified by (3.14) and (3.15) the variance of $\bar{x}_0'$, the composite estimator of the current occasion mean, is found by setting $m = \infty$ in (3.16). Thus

$$\begin{aligned} V(\bar{x}_0') = {} & (1/n - 1/N)S_0^2 + 2n_2^3\, QS_0^2 \Big\{ Q^2(r\rho^2 - (r^2+1)\rho + r) \\ & + Q(r(r-1)\rho^2 - 2(r-1)\rho + r(r-1)) - (r-1)^2\rho + Q^r\rho^{r-1}\Big[Q^2 \cdot \\ & \quad (-(r-1)\rho^2 + r(r-1)\rho) + Q(-(r^2 - 2r + 2)\rho^2 + 2r\rho - r^2) \\ & - (r-1)\rho^2 + r(r-1)\rho \Big] \Big\} \Big/ n^2 n_1^2 (1 - Q^2)(1 - Q\rho)^2 . \end{aligned} \tag{3.23}$$

This formula is strictly valid only when (a) the first enumeration took place on occasion $u = -\infty$, (b) the population N is infinite. If N is not large then it is not possible even to approximate a one cycle design.

It is of some interest for purposes of comparison to evaluate the variance of $\bar{x}_0'$ under the assumption that an arithmetic rather than an exponential correlation structure holds over time. In order to reduce a very tedious algebraic derivation, the investigation was restricted to a one cycle rotation design $(m = \infty)$. A stationary covariance structure as in (3.14) is still assumed but (3.15) is now replaced by

$$\begin{aligned} S_{0,t} &= (\rho + (t+1)d)S_0^2 \quad \text{when } -(t+1)d \leq \rho , \\ &= 0 \qquad\qquad\qquad\quad \text{when } -(t+1)d > \rho , \end{aligned} \tag{3.24}$$

where $d > 0$ and $t = -1, -2, \ldots$. Thus the correlation between measurements on the same character decreases according to $\rho, \rho - d, \rho - 2d, \ldots, 0$ as the number of occasions between observations increases. It is assumed in the variance formula given below that $\rho + (-r+1)d > 0$, i.e., that matching only takes place between units that are positively correlated. Setting $m = \infty$ in (3.13) and substituting (3.24) yields after considerable simplification

$$\begin{aligned} V(\bar{x}_0') &= (1/n - 1/N)S_0^2 + 2S_0^2 \,[\; Q^2(r(r-1) - r(3r-4)Q + 3r(r-2)Q^2 \\ &\quad - r(r-4)Q^3 - rQ^4) + Q\rho(rQ^5 + r(r-4)Q^4 - (2r^2 - 4r - 1)Q^3 + (2r-3)Q^2 \\ &\quad + (r-1)(2r-3)Q - (r-1)^2) + Q^2 d(rQ^4 + r(r-4)Q^3 - (r^2 - 2r - 2)Q^2 \\ &\quad - (r-2)^2 Q + (r-1)(r-2) \;] \;/(r^2(r-1)^2 n_2 (1-Q^2)(1-Q)^3) \\ &\quad + 2S_0^2 Q^{r+1} d((r^2 - 4r + 2)Q^3 - (3r^2 - 10r + 4)Q^2 + (3r^2 - 8r + 2)Q \\ &\quad - r(r-2))/(r^2(r-1)n_2(1-Q^2)(1-Q)^3) + 2S_0^2 Q^{r+1} \rho/(r^2 n_2 (1-Q^2)) . \end{aligned} \tag{3.25}$$

A numerical comparison between the two types of correlation models may be found later in Tables 4 and 5.

D. Generalization of the Sampling Procedure

The preceding discussions have been based on the premis that the rotation pattern is specified for all N units in the population P. A unit remains in the sample for r occasions, drops out for m occasions, returns for another r occasions, and so on. This rotation scheme necessarily requires that $N = (r+m)n_2$ where n_2 is the size of a rotation group such that $rn_2 = n$. It is thus requisite that after a given rotation group has dropped out of the sample, all other rotation groups must be represented in the sample on succeeding occasions before that same rotation group can return for a first visit of another cycle. Such a rotation plan is obviously lacking in generality.

Consider the following alternative sampling scheme. A random sample of size N^* is selected with equal probability and without replacement from the N units of P. A rotation design is then imposed on the N^* units so that, in effect, the N^* units become a new population P^* according to our former terminology. At any occasion α there are n units from P^* in the sample with n_2 on each of the 1st, 2nd, ..., r-th visits, $n = rn_2$.

Let $\bar{x}'_0$ denote the composite estimator of the population mean $\bar{X}_{0,N^*}$ of P^*. Now $\bar{x}'_0$ is an unbiased estimator of $\bar{X}_{0,N^*}$ and this implies that $\bar{x}'_0$ is also an unbiased estimator of $\bar{X}_0$. For, let E_{N^*} denote mathematical expectation over all possible samples of size N^*

from P, and let $E|_{N^*}$ denote expectation over all $N^*!$ possible rotation patterns within the N^* selected units. Then

$$E(\bar{x}_0') = \underset{N^*}{E}\,(E|_{N^*}(\bar{x}_0')) = \underset{N^*}{E}\,(\bar{X}_{0,N^*}) = \bar{X}_0 \,,$$

which proves the unbiasedness property. Further

$$V(\bar{x}_0') = \underset{N^*}{E}\,(V|_{N^*}(\bar{x}_0')) + \underset{N^*}{V}\,(E|_{N^*}(\bar{x}_0'))$$

where the conditional variance operators are interpreted in the same manner as the conditional expectation operators. Now

$$\underset{N^*}{V}\,(E|_{N^*}(\bar{x}_0')) = \underset{N^*}{V}\,(\bar{X}_{0,N^*}) = (1/N^* - 1/N)S_0^2 \,.$$

Further, $\underset{N^*}{E}\,(V|_{N^*}(\bar{x}_0'))$ is derived from $V|_{N^*}(\bar{x}_0)$ by replacing the subpopulation P^* variances and covariances by the population P variances and covariances because they are obviously unbiased estimators of these parameters. The leading term of this second expression will be $(1/n - 1/N^*)S_0^2$, so that the leading term of $V(\bar{x}_0')$ will be

$$(1/N^* - 1/N)S_0^2 + (1/n - 1/N^*)S_0^2 = (1/n - 1/N)S_0^2 \,.$$

Therefore the variance of the composite estimator of $\bar{X}_0$ is not dependent upon the size of the subpopulation P^* and the variance formulae previously derived are valid for the modified rotation design. Thus, in order to impose an infinite cycle rotation pattern such that any rotation group of n_2 units has r consecutive visits in the sample followed by m visits out of the sample, first select a random sample of size $N^* = (r+m)n_2$ from P and then proceed in the usual fashion to develop the rotation plan.

Note that for a given population of size N and specified r and n the maximum possible value of m is $m_0 = r(N/n - 1)$ and m may take on any positive integral value $\leq m_0$.

For convenience we shall continue to refer to the rotation plan as being superimposed on all N units of P. It will, however, be understood that all results so obtained are valid under the more general sampling procedure described above.

E. Composite Estimation of Change

1. The change estimator and its variance

An obvious unbiased estimator of the change in level between the previous and current occasions is the difference between the composite estimators of the current and previous population means, viz.,

$$d_0' = \bar{x}_0' - \bar{x}_{-1}' \; . \tag{3.26}$$

As mentioned in the Introduction, the sample data made available on the current occasion may be used to supply an estimator of $\bar{X}_{-1}$ which is more precise than $\bar{x}'_{-1}$. Since the resulting estimate of change would then be in discrepance with the individual estimates of level, study will be restricted to an estimator of the form (3.26). Now since

$$\bar{x}_0' = \sum_{\alpha=0}^{-\infty} Q^{-\alpha} W_\alpha ,$$

where

$$W_a = Q(\bar{x}_{a,a-1} - \bar{x}_{a-1,a}) + (1-Q)\bar{x}_a ,$$

it follows that

$$\bar{x}'_{-1} = \sum_{a=0}^{-\infty} Q^{-a} W_{a-1} = \sum_{a=-1}^{-\infty} Q^{-a-1} W_a .$$

Now

$$V(d'_0) = V(\bar{x}'_0) + V(\bar{x}'_{-1}) - 2\,\mathrm{Cov}(\bar{x}'_0, \bar{x}'_{-1}),$$

and

$$\mathrm{Cov}(\bar{x}'_0, \bar{x}'_{-1}) = \mathrm{Cov}\Big(\sum_{a=0}^{-\infty} Q^{-a} W_a, \sum_{a=-1}^{-\infty} Q^{-a-1} W_a\Big)$$

$$= \mathrm{Cov}\Big(W_0, \sum_{a=-1}^{-\infty} Q^{-a-1} W_a\Big) + Q\,\mathrm{Cov}\Big(\sum_{a=-1}^{-\infty} Q^{-a-1} W_a, \sum_{a=-1}^{-\infty} Q^{-a-1} W_a\Big)$$

$$= \sum_{a=-1}^{-\infty} Q^{-a-1}\,\mathrm{Cov}(W_0, W_a) + QV(\bar{x}'_{-1}).$$

Hence

$$V(d'_0) = V(\bar{x}'_0) + (1-2Q)V(\bar{x}'_{-1}) - 2\sum_{a=-1}^{-\infty} Q^{-a-1}\,\mathrm{Cov}(W_0, W_a). \qquad (3.27)$$

But

$$V(\bar{x}'_0) = V\Big(\sum_{a=0}^{-\infty} Q^{-a} W_a\Big) = V\Big(W_0 + \sum_{a=-1}^{-\infty} Q^{-a} W_a\Big)$$

$$= V(W_0) + Q^2 V(\bar{x}'_{-1}) + 2\sum_{a=-1}^{-\infty} Q^{-a}\,\mathrm{Cov}(W_0, W_a),$$

so that

$$V(\bar{x}'_{-1}) = (V(\bar{x}'_0) - V(W_0) - 2 \sum_{a=-1}^{-\infty} \mathrm{Cov}(W_0, W_a))/Q^2 ,$$

and so, substituting into (3.27),

$$V(d'_0) = \left[(Q-1)^2 V(\bar{x}'_0) + (2Q-1)V(W_0) + 2(Q-1) \sum_{a=-1}^{-\infty} Q^{-a}\mathrm{Cov}(W_0, W_a)\right] \Big/ Q^2 . \tag{3.28}$$

It is now necessary to evaluate $V(W_0)$ and $\mathrm{Cov}(W_0, W_a)$; $V(\bar{x}'_0)$ has already been dealt with in previous sections. Let

$$W_0 = Q(\bar{x}_{0,-1} - \bar{x}_{-1,0}) + (1-Q)\bar{x}_0 = \sum_{t=-1}^{0} \sum_{k=1}^{N} v_{t,k}\, x_{t,k} ,$$

where

(a) $v_{0,k} = (1-Q)/n$ for n_2 units on 1st visit of a cycle at $t = 0$,

(b) $v_{0,k} = (1 + n_2 Q/n_1)/n$ for n_1 units on 2nd to r-th visits of a cycle at $t = 0$,

(c) $v_{0,k} = 0$ for $(N-n)$ units not in the sample at $t = 0$,

(d) $v_{-1,k} = -Q/n_1$ for n_1 units on 1st to (r-1)-th visit of a cycle at $t = -1$,

(e) $v_{-1,k} = 0$ for the remaining $N-n$, units at $t = -1$.

Thus

$$E(W_0) = \sum_{k=1}^{N} E(v_{0,k})x_{0,k} + \sum_{k=1}^{N} E(v_{-1,k})x_{-1,k} = \overline{X}_0 - Q\overline{X}_{-1}.$$

Using the fact that

$$V(W_0) = E(W_0^2) - (E(W_0))^2 = E(\sum_{t=-1}^{0} \sum_{k=1}^{N} v_{t,k} x_{t,k})^2 - (\overline{X}_0 - Q\overline{X}_{-1})^2,$$

some straightforward algebra will lead to

$$V(W_0) = NE(v_{0,k}^2)S_0^2 + NE(v_{-1,k}^2)S_{-1}^2 + 2NE(v_{0,k}v_{-1,k})S_{0,-1}$$

$$- S_0^2/N - Q^2 S_{-1}^2/N + 2QS_{0,-1}/N. \tag{3.29}$$

With the aid of (a) to (e) above, (3.29) finally reduces to

$$V(W_0) = (1/n - 1/N)S_0^2 + Q^2(1/n_1 - 1/n)S_0^2 + Q^2(1/n_1 - 1/N)S_{-1}^2$$

$$- 2Q(1/n - 1/N)S_{0,-1} - 2Q^2(1/n_1 - 1/n)S_{0,-1}. \tag{3.30}$$

Next let

$$W_\alpha = Q(\overline{x}_{\alpha,\alpha-1} - \overline{x}_{\alpha-1,\alpha}) + (1-Q)\overline{x}_\alpha = \sum_{t=\alpha-1}^{\alpha} \sum_{k=1}^{N} u_{t,k} x_{t,k}, \quad (\alpha < 0),$$

where $u_{\alpha,k} = v_{0,k}$, $u_{\alpha-1,k} = v_{-1,k}$ for the same visit numbers on the respective occasions. Hence

$$E(W_\alpha) = \overline{X}_\alpha - Q\overline{X}_{\alpha-1}.$$

Analogous to (3.30) it may be shown that

$$Cov(W_0, W_\alpha) = (NE(v_{0,k}u_{\alpha,k}) - 1/N)S_{0,\alpha} + (NE(v_{0,k}u_{\alpha-1,k})$$

$$+ Q/N)S_{0,a-1} + (NE(v_{-1,k}u_{a-1,k}) - Q^2/N)S_{-1,a-1} + (NE(v_{-1,k}u_{a,k})$$

$$+ Q/N)S_{-1,a} . \qquad (3.31)$$

(3.31) will be evaluated for an infinite cycle rotation design with $m \geq r$. Then

$$NE(v_{0,k}u_{a,k}) = n_2(1 + n_2 Q/n_1)((r-t)n_1 - tn_2 Q)/n^2 n_1$$

$$\text{for } a = -\ell(r+m) - t, \ \ell = 0, 1, 2, \ldots, \ 1 \leq t \leq r-1,$$

$$= 0 \quad \text{for } a = -\ell(r+m) - r - t, \ \ell = 0, 1, 2, \ldots, \ 0 \leq t \leq m-r,$$

$$= n_2(n_1 + n_2 Q)(-nQ + t(n_1 + n_2 Q))/n^2 n_1^2$$

$$\text{for } a = -\ell(r+m) - m - t, \ \ell = 0, 1, 2, \ldots, \ 1 \leq t \leq r-1,$$

$$= (n_1 + n_2 Q^2)/nn_1 \quad \text{for } a = -\ell(r+m), \ \ell = 1, 2, \ldots .$$

$NE(v_{0,k}u_{a-1,k})$

$$= -Q(n_1 + n_2 Q)(n_1 - tn_2)/nn_1^2 \quad \text{for } a = -\ell(r+m) - t, \ \ell = 0, 1, 2, \ldots,$$

$$1 \leq t \leq r-1,$$

$$= 0 \quad \text{for } a = -\ell(r+m) - r - t, \ \ell = 0, 1, 2, \ldots, \ 0 \leq t \leq m-r,$$

$$= -Qn_2(-nQ + t(n_1 + n_2 Q))/nn_1^2 \quad \text{for } a = -\ell(r+m) - m - t, \ \ell = 0, 1, 2, \ldots,$$

$$1 \leq t \leq r-1,$$

$$= -Q(n_1 + n_2 Q)/nn_1 \quad \text{for } a = -\ell(r+m), \ \ell = 1, 2, 3, \ldots .$$

$NE(v_{-1,k}u_{a,k})$

$= -Q(nn_1 - tn_2(n_1 + n_2 Q))/nn_1^2$ for $a = -\ell(r+m) - t$ where $\ell = 0, 1, 2, \ldots,$

$1 \leq t \leq r-1$,

$= 0$ for $a = -\ell(r+m) - r - t$, $\ell = 0, 1, 2, \ldots,$ $0 \leq t \leq m-r$,

$= -Q(n_1 + n_2 Q)(t-1)n_2/nn_1^2$ for $a = -\ell(r+m) - m - t$, $\ell = 0, 1, 2, \ldots,$

$1 \leq t \leq r-1$,

$= -Q(n_1 + n_2 Q)/nn_1$ for $a = -\ell(r+m)$, $\ell = 1, 2, 3, \ldots$.

$NE(v_{-1,k} u_{a-1,k})$

$= Q^2 (n_1 - tn_2)/n_1^2$ for $a = -\ell(r+m) - t$, $\ell = 0, 1, 2, \ldots,$ $1 \leq t \leq r-1$,

$= 0$ for $a = -\ell(r+m) - r - t$, $\ell = 0, 1, 2, \ldots,$ $0 \leq t \leq m-r$,

$= Q^2 (t-1)n_2/n_1^2$ for $a = -\ell(r+m) - m - t$, $\ell = 0, 1, 2, \ldots,$ $1 \leq t \leq r-1$,

$= Q^2/n_1$ for $a = -\ell(r+m)$, $\ell = 1, 2, 3, \ldots$.

Thus, making use of (3.31) and the above relationships, we have that

$$\sum_{a=-1}^{-\infty} Q^{-a} \mathrm{Cov}(W_0, W_a) = \sum_{t=1}^{\infty} Q^t (-S_{0,-t}/N + QS_{0,-t-1}/N + QS_{-1,-t}/N$$

$$-Q^2 S_{-1,-t-1}/N) + \sum_{\ell=0}^{\infty} Q^{\ell(r+m)} \sum_{t=1}^{r-1} Q^t n_2(n_1 + n_2 Q)((r-t)n_1 - tn_2 Q)$$

$$S_{0,-\ell(r+m)-t}/(n^2 n_1^2) + \sum_{\ell=0}^{\infty} Q^{\ell(r+m)} \sum_{t=1}^{r-1} Q^{m+t} n_2(n_1 + n_2 Q)(-nQ$$

$$+t(n_1+n_2Q))S_{0,-\ell(r+m)-m-t}/(n^2n_1^2)+\sum_{\ell=1}^{\infty}Q^{\ell(r+m)}(n_1+n_2Q^2)$$

$$S_{0,-\ell(r+m)}/(nn_1)$$

$$-\sum_{\ell=0}^{\infty}Q^{\ell(r+m)}\sum_{t=1}^{r-1}Q^{t+1}(n_1+n_2Q)(n_1-tn_2)S_{0,-\ell(r+m)-t-1}/(nn_1^2)$$

$$-\sum_{\ell=0}^{\infty}Q^{\ell(r+m)}\sum_{t=1}^{r-1}Q^{m+t+1}n_2(-nQ+t(n_1+n_2Q))S_{0,-\ell(r+m)-m-t-1}/(nn_1^2)$$

$$-\sum_{\ell=1}^{\infty}Q^{\ell(r+m)}Q(n_1+n_2Q)S_{0,-\ell(r+m)-1}/(nn_1)$$

$$+\sum_{\ell=0}^{\infty}Q^{\ell(r+m)}\sum_{t=1}^{r-1}Q^{t+2}(n_1-tn_2)S_{-1,-\ell(r+m)-t-1}/n_1^2$$

$$+\sum_{\ell=0}^{\infty}Q^{\ell(r+m)}\sum_{t=1}^{r-1}Q^{m+t+2}(t-1)n_2S_{-1,-\ell(r+m)-m-t-1}/n_1^2$$

$$+\sum_{\ell=1}^{\infty}Q^{\ell(r+m)+2}S_{-1,-\ell(r+m)-1}/n_1$$

$$-\sum_{\ell=0}^{\infty}Q^{\ell(r+m)}\sum_{t=1}^{r-1}Q^{t+1}(n_1n-tn_2(n_1+n_2Q))S_{-1,-\ell(r+m)-t}/(nn_1^2)$$

$$-\sum_{\ell=0}^{\infty}Q^{\ell(r+m)}\sum_{t=1}^{r-1}Q^{m+t+1}(n_1+n_2Q)(t-1)n_2S_{-1,-\ell(r+m)-m-t}/(nn_1^2)$$

$$-\sum_{\ell=1}^{\infty}Q^{\ell(r+m)+1}(n_1+n_2Q)S_{-1,-\ell(r+m)}/(nn_1)\,. \tag{3.32}$$

Therefore, by virtue of (3.28) and making use of (3.30) and (3.32),

$$V(d_0') = \Big\{ (Q-1)^2 V(\bar{x}_0') + (2Q-1)\,((1/n - 1/N)S_0^2 + Q^2(1/n_1 - 1/n)S_0^2$$

$$+ Q^2(1/n_1 - 1/N)S_{-1}^2 - 2Q(1/n - 1/N)S_{0,-1} - 2Q^2(1/n_1 - 1/n)S_{0,-1})$$

$$+ 2(Q-1)\sum_{\ell=0}^{\infty} Q^{\ell(r+m)} \sum_{t=1}^{r-1} Q^t \Big[n_2(n_1 + n_2 Q)\,((r-t)n_1 - tn_2 Q)S_{0,-\ell(r+m)-t}$$

$$/(n^2 n_1^2) - Q(n_1 + n_2 Q)(n_1 - tn_2)S_{0,-\ell(r+m)-t-1}/(n_1^2 n)$$

$$+ Q^2(n_1 - tn_2)S_{-1,-\ell(r+m)-t-1}/n_1^2 - Q(n_1 n - tn_2(n_1 + n_2 Q))$$

$$S_{-1,-\ell(r+m)-t}/(n_1^2 n)\Big] + 2(Q-1)\sum_{t=1}^{\infty} Q^t(-S_{0,-t} + QS_{0,-t-1} + QS_{-1,-t}$$

$$- Q^2 S_{-1,-t-1})/N + 2(Q-1)\sum_{\ell=0}^{\infty} Q^{\ell(r+m)} \sum_{t=1}^{r-1} Q^{m+t}\Big[n_2(n_1 + n_2 Q)$$

$$(-nQ + t(n_1 + n_2 Q))S_{0,-\ell(r+m)-m-t}/(n^2 n_1^2) - n_2 Q(-nQ + t(n_1 + n_2 Q))$$

$$S_{0,-\ell(r+m)-m-t-1}/(nn_1^2) + Q^2(t-1)n_2 S_{-1,-\ell(r+m)-m-t-1}/n_1^2$$

$$- Q(n_1 + n_2 Q)(t-1)n_2 S_{-1,-\ell(r+m)-m-t}/(n_1^2 n)\Big]$$

$$+ 2(Q-1)\sum_{\ell=1}^{\infty} Q^{\ell(r+m)}\Big[(n_1 + n_2 Q^2)S_{0,-\ell(r+m)}/(nn_1) - Q(n_1 + n_2 Q)$$

$$S_{0,-\ell(r+m)-1}/(nn_1) + Q^2 S_{-1,-\ell(r+m)-1}/n_1 - Q(n_1 + n_2 Q)$$

$$S_{-1,-\ell(r+m)}/(nn_1)\Big]\Big\} \;/Q^2 \,. \qquad (3.33)$$

If a stationary Markoff type covariance structure, as given by (3.14) and (3.15), is assumed so that $V(\bar{x}_0')$ is given by (3.16), then after a considerable amount of algebraic manipulation (3.33) will reduce to

$$V(d_0') = 2(1/n - 1/N)(1-\rho)S_0^2 + 2(Q-1)S_0^2 \Big\{ -(r-1)^2\rho + (2r^2-5r+3)Q\rho$$

$$+ r(r-1)Q\rho^2 + (2r^2-r-3)Q^2\rho - (3r^2-4r)Q^2\rho^2 + (3r-2r^2)Q^2 - rQ^3$$

$$+ (3r^2-2r+1)Q^3\rho + (-3r^2+2r)Q^3\rho^2 + r(r-1)Q^3\rho^3 + (-r^2+r)Q^4\rho^2$$

$$+ r(r-1)Q^4\rho^3 + Q^r\rho^{r-1}\Big[r(r-1)\rho - (r-1)\rho^2 - r^2Q - r(r-3)Q\rho$$

$$- (r^2-3r+3)Q\rho^2 + r^2Q^2 + (r^2-3r)Q^2\rho + (r^2-3r+3)Q^2\rho^2 - r(r-1)Q^3\rho$$

$$+ (r-1)Q^3\rho^2\Big] + Q^m\rho^{m+1}\Big[(r-1)^2 + (-2r^2+5r-3)Q - r(r-1)Q\rho$$

$$- 3(r-1)Q^2 + 3r(r-1)Q^2\rho + (r^2-r-1)Q^3 - r(r-3)Q^3\rho - r^2Q^3\rho^2 + rQ^4$$

$$- r(r+1)Q^4\rho + r^2Q^4\rho^2\Big] + Q^{m+r}\rho^{m+r}\Big[(-r^2+r) + (r-1)\rho + (3r^2-3r)Q$$

$$+ (-3r+3)Q\rho + r^2Q^2 + (-6r^2+7r-3)Q^2\rho + (2r^2-r)Q^2\rho^2 - r^2Q^3$$

$$+ (-r^2+2r+1)Q^3\rho + (4r^2-5r)Q^3\rho^2 - r(r-1)Q^3\rho^3 + r(r-1)Q^4\rho$$

$$- r(r-1)Q^4\rho^3\Big]\Big\} / r^2(r-1)^2(1-Q\rho)^2(1-Q^2)(1-(Q\rho)^{r+m})n_2. \qquad (3.34)$$

2. The static population check

As a check on (3.34) again consider the case $x_{0,k} = x_{-1,k} = x_{-2,k} = \ldots$, $k = 1, 2, \ldots, N$, whence $\rho = 1$. Then

$$d_0' = (1-Q)\sum_{t=0}^{\infty} Q^t(\bar{x}_{-t} - \bar{x}_{-t-1}),$$

and so

$$V(d_0') = V(\bar{x}_0') + V(\bar{x}_{-1}') - 2(1-Q)^2 \operatorname{Cov}\left(\sum_{t=0}^{\infty} Q^t \bar{x}_{-t}, \sum_{t=0}^{\infty} Q^t \bar{x}_{-t-1} \right).$$

Since $V(\bar{x}_0') = V(\bar{x}_{-1}')$ when $\rho = 1$, it follows that

$$V(d_0') = 2(1-Q)V(\bar{x}_0') - 2(1-Q)^2 \operatorname{Cov}\left(\bar{x}_0, \sum_{t=0}^{\infty} Q^t \bar{x}_{-t-1}\right). \tag{3.35}$$

Now it may be shown that

$$\operatorname{Cov}(\bar{x}_0, \bar{x}_{-1} + Q\bar{x}_{-2} + Q^2 \bar{x}_{-3} + \dots)$$

$$= \sum_{\ell=0}^{\infty} Q^{\ell(r+m)} \sum_{s=1}^{r-1} Q^{s-1}(n_2(r-s)/n^s - 1/N)S_0^2 - \sum_{\ell=0}^{\infty} Q^{\ell(r+m)} \sum_{s=0}^{m-r} Q^{r-1+s} S_0^2/N$$

$$+ \sum_{\ell=0}^{\infty} Q^{\ell(r+m)} \sum_{s=1}^{r} Q^{m-1+s} (sn_2/n^2 - 1/N)S_0^2 ,$$

$$= -S_0^2/N(1-Q) + S_0^2 \Big[(1-Q^{r-1})/n(1-Q) + n_2(rQ^{r-1} - (r-1)Q^r - 1)/n^2(1-Q)^2$$

$$+ n_2 Q^{m-1}(Q - (r+1)Q^{r+1} + rQ^{r+2})/n^2(1-Q)^2 \Big] /(1-Q^{r+m}). \tag{3.36}$$

Substituting (3.21) and (3.36) into (3.35) and simplifying gives

$$V(d_0') = 2n_2 S_0^2 (1-Q)^2 (1 - Q^r - Q^m + Q^{m+r})/n^2 (1-Q^2)(1-Q^{r+m}). \tag{3.37}$$

Substituting $\rho = 1$ into (3.34) also yields (3.37) thereby providing a check.

3. Further comments

It might be felt that a somewhat devious route was taken in deriving

(3.33) and (3.34) and that a direct approach in the spirit of B and C might lead more directly to the desired results. It may be verified that three sets of weights $w_{\alpha, k}$ appropriate to $\alpha = 0$, $\alpha = -1$ and $\alpha < -1$ are then necessary. The additional set of weights introduces a forbidding amount of algebra into the subsequent derivation. Hence the method employed is to be preferred.

There may be reason to estimate the change in level between occasions that are not consecutive. For example, in a monthly panel survey estimates of quarterly and yearly change may as well be of some special interest. A composite estimator of $\overline{X}_0 - \overline{X}_{-a}$, $a \leq -1$, is

$$d'_{0,-a} = \overline{x}'_0 - \overline{x}'_{-a} ,$$

$$= \sum_{\alpha=0}^{\infty} Q^{-\alpha} (W_\alpha - W_{\alpha-a})$$

where W_α is given by (3.3). Full cognizance of the rotation plan is essential since the number of sample units (if any) in common with occasions 0 and -a is dependent upon m and r. The variance function $V(d'_{0,-a})$ will therefore vary with the choice of -a. It is a straightforward but extremely tedious job to develop $V(d'_{0,-a})$ for a given value of -a and no attempt at any explicit evaluation will be made here.

A composite estimate of the average level, $(\overline{X}_0 + \overline{X}_{-1})/2$, on the current and previous occasion is $(\overline{x}'_0 + \overline{x}'_{-1})/2 = s'_0$ with variance $V(s'_0) = (V(\overline{x}'_0) + V(\overline{x}'_{-1}) + 2\,\mathrm{Cov}(\overline{x}'_0, \overline{x}'_{-1}))/4$. The final form of this variance may therefore be obtained by replacing the minus covariance by a plus covariance in the derivation of $V(d'_0)$ and dividing by four.

F. Covariance Between Current Occasion and Change Estimators

The problem of choosing an optimum design for the joint estimation of $\bar{x}'_0$ and d'_0 will be considered later. The solution offered will depend upon the covariance between the two estimators $\bar{x}'_0$ and d'_0 as well as their individual variances. Hence $\mathrm{Cov}(\bar{x}'_0, d'_0)$ will now be evaluated.

$$\mathrm{Cov}(\bar{x}'_0, d'_0) = \mathrm{Cov}(\bar{x}'_0, \bar{x}'_0 - \bar{x}'_{-1}) = V(\bar{x}'_0) - \mathrm{Cov}(\bar{x}'_0, \bar{x}'_{-1}) .$$

Further,

$$V(d'_0) = V(\bar{x}'_0) + V(\bar{x}'_{-1}) - 2\,\mathrm{Cov}(\bar{x}'_0, \bar{x}'_{-1}) ,$$

so that

$$\mathrm{Cov}(\bar{x}'_0, \bar{x}'_{-1}) = (V(\bar{x}'_0) + V(\bar{x}'_{-1}) - V(d'_0))/2 .$$

Since it was shown earlier that

$$V(\bar{x}'_{-1}) = (V(\bar{x}'_0) - V(W_0) - 2 \sum_{\alpha=-1}^{-\infty} Q^{-\alpha}\, \mathrm{Cov}(W_0, W_\alpha))/Q^2 ,$$

it follows that

$$\mathrm{Cov}(\bar{x}'_0, \bar{x}'_{-1}) = \Big[V(x'_0) + (V(\bar{x}'_0) - V(W_0) - 2 \sum_{\alpha=-1}^{-\infty} Q^{-\alpha}\, \mathrm{Cov}(W_0, W_\alpha))/Q^2 - V(d'_0) \Big] /2 ,$$

and hence that

$$\mathrm{Cov}(\bar{x}_0', d_0') = ((Q^2-1)V(\bar{x}_0') + V(W_0) + 2\sum_{\alpha=-1}^{-\infty} Q^{-\alpha}\,\mathrm{Cov}(W_0, W_\alpha)$$

$$+ Q^2 V(d_0'))/2Q^2 .$$

With the help of (3.28) this becomes

$$\mathrm{Cov}(\bar{x}_0', d_0') = ((Q-1)^2 V(\bar{x}_0') - V(W_0) + Q^2 V(d_0'))/2Q(Q-1), \qquad (3.38)$$

and expressions for $V(\bar{x}_0')$, $V(W_0)$ and $V(d_0')$ are already available.

Under the covariance structure specified by (3.14) and (3.15), substitution of (3.16), (3.30) and (3.33) into (3.38) gives

$$\mathrm{Cov}(\bar{x}_0', d_0') = (1/n - 1/N)(1-\rho)S_0^2 + S_0^2 \{ (r-1)^2 \rho + (-r(r-1)\rho^2$$

$$- (3r-4)(r-1)\rho)Q + (r(4r-5)\rho^2 - 2(2r-3)\rho + r(2r-3))Q^2 + (-r(r-1)\rho^3$$

$$+ 2r\rho^2 + (-r^2+r-4)\rho - 2r(r-2))Q^3 + (r(-2r+1)\rho^2 + (3r^2-2r+1)\rho - r)Q^4$$

$$+ (r(r-1)\rho^3 - r(r-1)\rho^2)Q^5 + Q^r\rho^{r-1}[(r-1)\rho^2 - r(r-1)\rho + ((r-2)^2\rho^2$$

$$+ 2r(r-2)\rho + r^2)Q + (2(-r^2+3r-3)\rho^2 - 2r(r-3)\rho - 2r^2)Q^2$$

$$+ ((r-2)^2\rho^2 + 2r(r-2)\rho + r^2)Q^3 + ((r-1)\rho^2 - r(r-1)\rho)Q^4]$$

$$+ Q^m\rho^{m+1}[-(r-1)^2 + (r(r-1)\rho + (3r-4)(r-1))Q + (-4r(r-1)\rho$$

$$- 2(r-3)(r-1))Q^2 + (r^2\rho^2 + 2r(2r-3)\rho - r^2 - 2r + 4)Q^3 + (-2r^2\rho^2$$

$$+ 4r\rho + r^2 - 2r - 1)Q^4 + (r^2\rho^2 - r(r+1)\rho + r)Q^5] + Q^{m+r}\rho^{m+r}[r(r-1)$$

$$-(r-1)\rho + 4((r-1)\rho - r(r-1))Q + (r(-2r+1)\rho^2 + 2(3r^2-5r+3)\rho$$

$$+ r(2r-3))Q^2 + (r(r-1)\rho^3 - 2r(r-2)\rho^2 + (-5r^2+5r-4)\rho + 2r^2)Q^3$$

$$+ (r(4r-5)\rho^2 + (-2r^2+3r+1)\rho - r^2)Q^4 + (-r(r-1)\rho^3 + r(r-1)\rho)Q^5\Big]\Big\}$$

$$/ r^2(r-1)^2 n_2 (1-Q^2)(1-Q\rho)^2 (1-(Q\rho)^{r+m}) . \tag{3.39}$$

As a check on (3.39) take the case where $x_{0,k} = x_{-1,k} = x_{-2,k} = \cdots$ so that $\rho = 1$. Then

$$\mathrm{Cov}(\bar{x}_0', d_0') = (1-Q)V(\bar{x}_0') - (1-Q)^2 \mathrm{Cov}(\bar{x}_0, \sum_{t=1}^{\infty} Q^{t-1} \bar{x}_{-t}),$$

which leads to

$$\mathrm{Cov}(\bar{x}_0', d_0') = (Q-1)^2 S_0^2 (1 - Q^r - Q^m + Q^{m+r})/r^2 n_2 (1-Q^2)(1-Q^{r+m}). \tag{3.40}$$

Putting $\rho = 1$ in (3.39) and simplifying also yields (3.40).

G. Numerical Determination of Optimum Values of Design and Estimator Parameters

Some consideration should be given to the choice of the two design parameters, r and m, and the estimator parameter, Q, if full benefit is to be derived from the partial replacement sample design. The sample size n on each occasion is assumed to be constrained by budget considerations. The correlation structure over time and the population variances are, of course, beyond the control of the statistician. If the cost of enumerating all units is the same regardless of the visit or cycle number, then the minimum variance of the statistic under consideration will be the criterion for selecting the appropriate values for r, m and Q. A different criterion will be introduced when the joint estimation of $\overline{X}_0$ and $\overline{X}_0 - \overline{X}_{-1}$ is considered.

By an optimum value of a design or estimator parameter is meant that value of the parameter which minimizes the variance of the statistic under consideration with all other parameter values held fixed. For a given covariance structure there is obviously some optimum combination of r, m, and Q which provides an overall minimum variance. The usual differential calculus approach for the determination of minima was not employed because of the extreme complexity of the differentials of the functions involved. Rather, a numerical investigation of the variance functions $V(\overline{x}_0')$ and $V(d_0')$ and the covariance function $Cov(\overline{x}_0', d_0')$ was carried out for the purpose of determining approximately the individual and overall optimum parameter values. The assistance of a high-speed electronic computer brought this task within the realm of feasibility.

1. Infinite population results

In order to obtain specific numerical results the covariance conditions (3.14) and (3.15) were imposed upon the populations so that $V(\overline{x}_0')$, $V(d_0')$ and $Cov(\overline{x}_0', d_0')$ are given by (3.16), (3.34) and (3.39) respectively in an infinite cycle design. These functions are dependent upon S_0^2, n and N as well as ρ, r, m and Q. By assuming that N is large so that $1/N$ may be ignored, and limiting the study to the efficiency gain of the composite estimators $\overline{x}_0'$ or d_0' relative to the arithmetic mean $\bar{x}_0$ or difference in arithmetic means $\bar{x}_0 - \bar{x}_{-1}$ respectively, the three factors S_0^2, n and N are eliminated from the calculations.

In the computer tabulations the number of consecutive occasions m for which a rotation group is not in the sample was taken to be an integral multiple ℓ of the number of consecutive occasions r that a rotation group is in the sample, $m = \ell r$, for convenience. The values 2, 3, 4, 6 and 8 were assigned to r, and the values 1, 2, 3, 4, 6, 8 and 10 to ℓ. The tabulations revealed that the larger ℓ values were usually unnecessary. The coefficient of correlation ρ between observations on the same unit on consecutive occasions was assigned the values 0.5, 0.6, 0.7, 0.8 and 0.9. The estimator coefficient Q ranged from 0.1 to 0.9 at intervals of 0.1. The relative efficiency criterion was calculated for all possible combinations of r, ℓ, ρ and Q and a visual inspection of the computer output was thus sufficient to determine the approximate optimum values of these parameters.

In Table 1 may be found the per cent increase in efficiency in using $\bar{x}'_0$ rather than $\bar{x}_0$ as an estimator of $\bar{X}_0$ when Q is close to its optimum value:

$$\text{Efficiency gain} = \frac{V(\bar{x}_0) - V(\bar{x}'_0)}{V(\bar{x}'_0)} (100)\% \qquad (3.41)$$

where $V(\bar{x}_0) = S_0^2/n$ with N infinite and $V(\bar{x}'_0)$ is given by (3.16). $\hat{Q}$ is the approximate optimum Q value as determined from the computer tabulations for specified ρ, r and ℓ. Because the tabulation interval for Q was 0.1 this approximate optimum should be at least within 0.05 of the true optimum Q. The efficiency gains at Q values below and above the approximate optima have also been recorded. These additional entries serve to indicate the degree of robustness of the optimum Q, i.e., the extent to which a deviation from the optimum Q decreases the efficiency gain. They also permit the evaluation of an improved approximation to the true optimum Q by means of, for instance, quadratic interpolation. For a given ρ the tabulations have been terminated at the point where any increase in ℓ would not alter the figures recorded to two decimal places. To illustrate the use of Table 1, for $\rho = 0.8$, $r = 3$, $\ell = 2$ the approximate optimum Q is $\hat{Q} = 0.5$ where the efficiency gain is 22.30%. At $Q = 0.4$ the gain is 20.46% and at $Q = 0.6$ it is 18.61%. Thus the true optimum would appear to be somewhere between 0.4 and 0.5.

Several interesting conclusions may be drawn from Table 1.

(a) The optimum number of visits by a rotation group is $r = 2$ for all ρ. The efficiency gain of $\bar{x}_0'$ over $\bar{x}_0$ declines steadily as r increases beyond two.

(b) As the correlation coefficient ρ increases the efficiency gain of $\bar{x}_0'$ over $\bar{x}_0$ becomes more and more pronounced. At $\rho = 0.5$, $\ell = 1$, $r = 2$, $\hat{Q} = 0.2$ the efficiency gain is 5.16% whereas at $\rho = 0.9$, $\ell = 1$, $r = 2$, $\hat{Q} = 0.5$, the gain is 38.78%.

(c) The optimum Q value is greater for larger values of ρ.

(d) For fixed ρ, r and Q the efficiency gain of $\bar{x}_0'$ over $\bar{x}_0$

increases as ℓ increases. This is in agreement with the optimum $r = 2$ already observed in (a).

(e) As ℓ increases the efficiency gain of $\bar{x}'_0$, and therefore $V(\bar{x}'_0)$ as well, rapidly approaches a limiting value. When $\rho \leq 0.9$ there is little error introduced in taking the one cycle variance formula as an approximation to the infinite cycle variance formula when $\ell \geq 3$. This is due to the fact that the two additional terms introduced into $V(\bar{x}'_0)$ by the finite recurrence time for a rotation group are then sufficiently damped by the $(Q\rho)^m$ and $(Q\rho)^{m+r}$ factors so as to be negligible in comparison with the one cycle terms.

(f) The efficiency of $\bar{x}'_0$ relative to $\bar{x}_0$ decreases only slightly in the neighborhood of the optimum Q; thus Q is robust.

There are two meaningful efficiency comparisons for the composite change estimator $d'_0 = \bar{x}'_0 - \bar{x}'_{-1}$. If $\bar{x}_0$ and $\bar{x}_{-1}$ are the arithmetic means on the current and previous occasions as observed in the rotation design, then there are $n_1 = (r-1)n_2$ common units entering into the calculation of these means and hence

$$V(\bar{x}_0 - \bar{x}_{-1}) = \frac{2S_0^2(1 - (r-1)\rho/r)}{rn_2}, \tag{3.42}$$

again ignoring the finite population correction. The efficiency gain is calculated as in (3.41). The efficiency gain using d'_0 may also be computed relative to the variance of the difference in sample means resulting from two independent samples, whence

$$V(\bar{x}_0 - \bar{x}_{-1}) = 2S_0^2/rn_2 . \tag{3.43}$$

Table 2 is similar in structure to Table 1 except that the efficiency gain relative to a change estimate from independent samples is recorded in brackets to the right of the gain relative to the simple change estimate in the rotation design. In cases where $\hat{Q}$ is 0.9 the efficiency gains appropriate to $Q = 0.7$ and $Q = 0.8$ are given. Since efficiency gains for Q values greater than 0.9 were not computed, it is possible that the true optimum Q is greater than 0.95 in such situations.

The following features of Table 2 are noteworthy.

(a) Larger gains in efficiency are scored through the use of composite estimation of change than by composite estimation of level for given ρ, r and ℓ. For larger values of ρ the gains are indeed appreciable; for example with $\rho = 0.9$, $r = 2$, $\ell = 1$, there is a 208.74% gain in efficiency using d_0' over $\bar{x}_0 - \bar{x}_{-1}$ in the rotation design.

(b) The gains relative to the difference of independent sample means are even greater since the latter mode of estimation does not take advantage of the positive correlation among observations on the same unit over time.

(c) The optimum value of r for any ρ and ℓ is again two (as for composite estimation of level) when comparison with the difference of simple matched sample means is made. The efficiency relative to the difference of independent sample means, however, steadily increases as r increases. This is because in the first case although both $V(d_0')$ and $V(\bar{x}_0 - \bar{x}_{-1})$ decrease with increasing r, the second variance diminishes at a faster rate than the first so that the overall efficiency ratio decreases. In the second case $V(\bar{x}_0 - \bar{x}_{-1})$ is independent of r and hence the efficiency ratio increases as r increases.

(d) As opposed to Table 1, the efficiency gain of d_0' relative to the two alternative methods of change estimation mentioned decreases as ℓ increases. For moderately large ℓ values $V(d_0')$ becomes insensitive to increases in ℓ.

(e) The optimum Q values for change estimation are in all cases greater than the optimum Q values for estimating level for the same values of ρ and r.

(f) The efficiency gains are affected but little by small deviations from the optimum Q values.

(g) Still another estimator of the change between the previous and current occasions is given by the difference between the arithmetic means of the matched units only on these occasions, $\bar{x}_{0,-1} - \bar{x}_{-1,0}$. This will be a more efficient estimator than $\bar{x}_0 - \bar{x}_{-1}$ provided that

$$\frac{2S_0^2(1 - (r-1)\rho/r)}{rn_2} > \frac{2S_0^2(1-\rho)}{(r-1)n_2} ,$$

or $$\rho > r/(2r-1) .$$

Thus if $r = 2$, ρ must be greater than $2/3$, if $r = 3$, ρ must be greater than 3/5, and so on. But in practical situations $\bar{x}_{0,-1} - \bar{x}_{-1,0}$ would not likely be used as an estimator of $\bar{X}_0 - \bar{X}_{-1}$ since it is in general incompatible with the individual estimates given by $\bar{x}_0$ and $\bar{x}_{-1}$. For this reason no comparison of the relative efficiencies of $\bar{x}'_0 - \bar{x}'_{-1}$ with $\bar{x}_{0,-1} - \bar{x}_{-1,0}$ will be made.

Although the optimum values of r for the composite estimation of $\bar{X}_0$ and $\bar{X}_0 - \bar{X}_{-1}$ agree, it is apparent that the optimum Q values for these estimators differ. It is therefore necessary to make some compromise between the two requirements in order to arrive at a single Q. Although the weight Q is permitted to vary between characters, it is apparent that the Q values used in the composite estimator of $\bar{X}_0$ and $\bar{X}_0 - \bar{X}_{-1}$ must necessarily agree if the estimate of change is to be given by the difference of the composite estimates of level. One possible criterion is the selection of that value of Q which minimizes the generalized variance of the two estimators, i.e., of the determinant of the covariance matrix of $\bar{x}'_0$ and d'_0,

$$\Delta'_0 = \begin{vmatrix} V(\bar{x}'_0) & \mathrm{Cov}(\bar{x}'_0, d'_0) \\ \mathrm{Cov}(\bar{x}'_0, d'_0) & V(d'_0) \end{vmatrix}$$

$$= V(\bar{x}'_0)V(d'_0) - (\mathrm{Cov}(\bar{x}'_0, d'_0))^2 .$$

Since $V(\bar{x}'_0)$ and $V(d'_0)$ decrease with increasing n, an efficiency criterion involving a ratio of generalized variances was defined,

$$\lambda = \left(\frac{\Delta}{\Delta_0'}\right)(100)\% , \qquad (3.44)$$

where

$$\Delta = \begin{vmatrix} V(\bar{x}_0) & Cov(\bar{x}_0, \bar{x}_0 - \bar{x}_{-1}) \\ Cov(\bar{x}_0, \bar{x}_0 - \bar{x}_{-1}) & V(\bar{x}_0 - \bar{x}_{-1}) \end{vmatrix} .$$

Here $\bar{x}_0 - \bar{x}_{-1}$ is the difference in sample means in the rotation design, and hence

$$\Delta = \begin{vmatrix} \frac{S_0^2}{rn_2} & \frac{S_0^2(1 - \frac{(r-1)}{r}\rho)}{rn_2} \\ \frac{S_0^2(1 - \frac{(r-1)}{r}\rho)}{rn_2} & \frac{2S_0^2(1 - \frac{(r-1)}{r}\rho)}{rn_2} \end{vmatrix}$$

$$= \frac{S_0^4(1 - \frac{(r-1)^2}{r^2}\rho^2)}{r^2 n_2^2} .$$

The approximate optimum compromise choice of Q, $\hat{Q}'$, is then that value of Q which maximized λ for a given ρ, r and ℓ. Table 3 is similar in format to Tables 1 and 2 and presents the approximate optimum $\hat{Q}'$ with the corresponding optimum λ' value. The λ criterion may be essentially regarded as a relative efficiency index as well. It will be noted from Table 3 that:

(a) In agreement with Tables 1 and 2 the optimum choice of r is always two.

(b) The optimum $\hat{\Omega}'$ lies between the optimum choices of Ω in Tables 1 and 2 for specified ρ, r and ℓ.

(c) The index λ increases with increasing ℓ for fixed ρ and r, and with increasing ρ for fixed ℓ and r.

(d) The index λ is robust with respect to small deviations from the optimum Ω.

There might perhaps be some question as to the desirability of the generalized variance criterion for selecting a compromise Ω value. One might alternately attempt to proceed along the lines of Hartley (1961) and prescribe gauges B_0 and C_0 such that

$$V(\bar{x}_0') \leq B_0 , \quad V(d_0') \leq C_0 .$$

Then, subject to these constraints, choose the optimum parameter values which minimize the cost of conducting the survey. This is a difficult problem in mathematical programming and no attempt was made at a solution.

It should again be emphasized that the conclusions derived from Tables 1, 2 and 3 are strictly valid only under the assumption of a Markoff type correlogram with constant population variances and covariances.

Table 1. Efficiency gains in per cent of $\bar{x}'_0$ over $\bar{x}_0$ near the optimum Q values

ρ	ℓ	r	Q	Efficiency gain	$\hat{Q}$	Efficiency gain	Q	Efficiency gain
0.5	1	2	0.1	4.03	0.2	5.16	0.3	1.88
		3	0.1	2.80	0.2	4.23	0.3	3.58
		4	0.1	2.12	0.2	3.31	0.3	3.17
		6	0.2	2.26	0.3	2.33	0.4	1.27
		8	0.2	1.71	0.3	1.81	0.4	1.11
	2	2	0.1	4.05	0.2	5.26	0.3	2.18
0.6	1	2	0.1	5.36	0.2	8.50	0.3	7.76
		3	0.2	6.28	0.3	7.25	0.4	5.49
		4	0.2	4.77	0.3	5.78	0.4	5.10
		6	0.2	3.18	0.3	3.96	0.4	3.79
		8	0.2	2.39	0.3	2.99	0.4	2.94
	2	2	0.1	5.39	0.2	8.69	0.3	8.34
		3	0.2	6.29	0.3	7.32	0.4	5.70
		4	0.2	4.77	0.3	5.79	0.4	5.14
	3	2	0.1	5.39	0.2	8.70	0.3	8.36
		3	0.2	6.29	0.3	7.33	0.4	5.70
0.7	1	2	0.2	12.03	0.3	14.23	0.4	11.38
		3	0.3	11.31	0.4	11.97	0.5	8.79
		4	0.3	8.66	0.4	9.78	0.5	8.54
		6	0.3	5.75	0.4	6.70	0.5	6.43
		8	0.3	4.28	0.4	5.03	0.5	4.96
	2	2	0.2	12.35	0.3	15.27	0.4	13.46
		3	0.3	11.46	0.4	12.43	0.5	9.71
		4	0.3	8.68	0.4	9.88	0.5	8.80
	3	2	0.2	12.36	0.3	15.32	0.4	13.62
		3	0.3	11.46	0.4	12.44	0.5	9.75
0.8	1	2	0.3	21.37	0.4	22.85	0.5	18.02
		3	0.4	19.53	0.5	20.20	0.6	15.21
		4	0.4	15.25	0.5	17.12	0.6	15.37
		6	0.5	11.86	0.6	11.99	0.7	7.73
		8	0.5	8.83	0.6	9.23	0.7	6.75
	2	2	0.3	23.13	0.4	26.67	0.5	23.85
		3	0.4	20.46	0.5	22.30	0.6	18.61
		4	0.4	15.47	0.5	17.78	0.6	16.81
		6	0.5	11.93	0.6	12.22	0.7	8.26
		8	0.5	8.84	0.6	9.27	0.7	6.88

Table 1 (Continued)

ρ	ℓ	r	Q	Efficiency gain	$\hat{Q}$	Efficiency gain	Q	Efficiency gain
0.8	3	2	0.3	23.23	0.4	27.07	0.5	24.82
		3	0.4	20.49	0.5	22.44	0.6	18.99
		4	0.4	15.47	0.5	17.80	0.6	16.89
	4	2	0.3	23.24	0.4	27.11	0.5	24.97
		3	0.4	20.49	0.5	22.45	0.6	19.04
0.9	1	2	0.4	36.42	0.5	38.78	0.6	33.75
		3	0.5	34.55	0.6	36.65	0.7	30.62
		4	0.5	27.92	0.6	32.40	0.7	31.22
		6	0.6	23.43	0.7	26.02	0.8	19.74
		8	0.6	17.49	0.7	20.62	0.8	18.10
	2	2	0.4	43.18	0.5	50.64	0.6	49.38
		3	0.5	39.07	0.6	45.28	0.7	42.46
		4	0.6	36.39	0.7	38.59	0.8	25.46
		6	0.6	24.22	0.7	28.35	0.8	23.99
		8	0.6	17.65	0.7	21.33	0.8	20.05
	3	2	0.5	53.19	0.6	54.22	0.7	39.64
		3	0.5	39.49	0.6	46.71	0.7	45.57
		4	0.6	36.73	0.7	39.79	0.8	27.64
		6	0.6	24.24	0.7	28.50	0.8	24.60
		8	0.6	17.65	0.7	21.35	0.8	20.19
	4	2	0.5	53.71	0.6	55.65	0.7	41.99
		3	0.5	39.53	0.6	46.93	0.7	46.36
		4	0.6	36.76	0.7	39.99	0.8	28.23
		6	0.6	24.24	0.7	28.51	0.8	24.68
	6	2	0.5	53.84	0.6	56.20	0.7	43.30
		3	0.5	39.53	0.6	46.97	0.7	46.60
		4	0.6	36.77	0.7	40.02	0.8	28.43
	8	2	0.5	53.85	0.6	56.25	0.7	43.50
		3	0.5	39.53	0.6	46.98	0.7	46.62

Table 2. Efficiency gains in per cent of $\bar{x}'_0 - \bar{x}'_{-1}$ over the difference of overall means and over independent samples (values in brackets) in estimating change near the optimum Q values

ρ	ℓ	r	Q	Efficiency gain	$\hat{Q}$	Efficiency gain	Q	Efficiency gain
0.5	1	2	0.2	10.54 (47.39)	0.3	12.76 (50.35)	0.4	12.39 (49.86)
		3	0.3	10.51 (65.77)	0.4	11.18 (66.77)	0.5	10.42 (65.63)
		4	0.3	8.68 (73.88)	0.4	9.38 (75.00)	0.5	9.06 (74.50)
		6	0.3	6.33 (82.29)	0.4	6.91 (83.28)	0.5	6.83 (83.13)
		8	0.3	4.96 (86.59)	0.4	5.43 (87.43)	0.5	5.41 (87.39)
	2	2	0.2	10.30 (47.06)	0.3	12.37 (49.82)	0.4	11.96 (49.28)
		3	0.3	10.47 (65.70)	0.4	11.11 (66.67)	0.5	10.35 (65.52)
		4	0.3	8.67 (73.88)	0.4	9.36 (74.98)	0.5	9.05 (74.48)
	3	2	0.2	10.29 (47.06)	0.3	12.36 (49.81)	0.4	11.94 (49.25)
0.6	1	2	0.3	20.94 (72.77)	0.4	23.43 (76.33)	0.5	22.65 (75.22)
		3	0.4	19.18 (98.64)	0.5	20.23 (100.39)	0.6	19.46 (99.09)
		4	0.4	15.75 (110.45)	0.5	16.84 (112.44)	0.6	16.71 (112.20)
		6	0.5	12.36 (124.71)	0.6	12.45 (124.90)	0.7	11.76 (123.52)
		8	0.5	9.70 (130.95)	0.6	9.82 (131.20)	0.7	9.37 (130.25)
	2	2	0.3	20.10 (71.57)	0.4	22.44 (74.91)	0.5	21.82 (74.03)
		3	0.4	18.99 (98.31)	0.5	20.00 (100.01)	0.6	19.28 (98.81)
		4	0.4	15.71 (110.38)	0.5	16.79 (112.34)	0.6	16.66 (112.10)
	3	2	0.3	20.07 (71.53)	0.4	22.38 (74.83)	0.5	21.75 (73.92)
0.7	1	2	0.4	39.49 (114.60)	0.5	42.83 (119.74)	0.6	41.68 (117.96)
		3	0.5	34.96 (153.04)	0.6	36.87 (156.64)	0.7	36.02 (155.03)
		4	0.6	30.53 (174.80)	0.7	30.74 (175.25)	0.8	28.84 (171.25)
		6	0.6	22.27 (193.46)	0.7	22.78 (194.67)	0.8	22.04 (192.89)
		8	0.6	17.47 (203.14)	0.7	17.92 (204.30)	0.8	17.50 (203.23)

Table 2 (Continued)

ρ	ℓ	r	Q	Efficiency gain	$\hat{Q}$	Efficiency gain	Q	Efficiency gain
0.7	2	2	0.4	37.31 (111.25)	0.5	40.81 (116.64)	0.6	40.36 (115.93)
		3	0.5	34.30 (151.80)	0.6	36.29 (155.54)	0.7	35.74 (154.50)
		4	0.6	30.31 (174.34)	0.7	30.61 (174.96)	0.8	28.89 (171.34)
		6	0.6	22.25 (193.39)	0.7	22.75 (194.61)	0.8	22.05 (192.91)
		8	0.6	17.46 (203.13)	0.7	17.91 (204.29)	0.8	17.50 (203.23)
	3	2	0.4	37.14 (110.99)	0.5	40.57 (116.27)	0.6	40.13 (115.59)
		3	0.5	34.27 (151.75)	0.6	36.24 (155.46)	0.7	35.70 (154.44)
		4	0.6	30.30 (174.33)	0.7	30.60 (174.95)	0.8	28.89 (171.35)
	4	2	0.4	37.13 (110.97)	0.5	40.54 (116.22)	0.6	40.09 (115.53)
0.8	1	2	0.5	77.42 (195.70)	0.6	83.52 (205.88)	0.7	82.37 (203.95)
		3	0.6	68.10 (260.22)	0.7	72.17 (268.94)	0.8	71.10 (266.64)
		4	0.7	59.63 (299.08)	0.8	60.49 (301.22)	0.9	57.15 (292.87)
		6	0.7	43.30 (329.90)	0.8	44.79 (334.36)	0.9	43.51 (330.53)
		8	0.7	33.89 (346.30)	0.8	35.19 (350.64)	0.9	34.60 (348.67)
	2	2	0.6	79.64 (199.40)	0.7	80.43 (200.71)	0.8	74.61 (191.01)
		3	0.6	66.12 (255.97)	0.7	70.85 (266.12)	0.8	70.80 (265.99)
		4	0.7	58.92 (297.30)	0.8	60.28 (300.69)	0.9	57.40 (293.49)
		6	0.7	43.12 (329.35)	0.8	44.70 (334.11)	0.9	43.63 (330.90)
		8	0.7	33.84 (346.15)	0.8	35.16 (350.55)	0.9	34.66 (348.85)
	3	2	0.6	78.81 (198.02)	0.7	79.88 (199.80)	0.8	74.49 (190.82)
		3	0.7	70.63 (265.64)	0.8	70.72 (265.84)	0.9	65.54 (254.73)
		4	0.7	58.85 (297.12)	0.8	60.24 (300.61)	0.9	57.46 (293.65)
	4	2	0.6	78.63 (197.71)	0.7	79.72 (199.53)	0.8	74.45 (190.75)
		3	0.7	70.59 (265.56)	0.8	70.71 (265.80)	0.9	65.58 (254.81)
	6	2	0.6	78.57 (197.62)	0.7	79.65 (199.41)	0.8	74.42 (190.71)
		3	0.7	70.58 (265.54)	0.8	70.70 (265.78)	0.9	65.59 (254.85)
	8	2	0.6	78.57 (197.62)	0.7	79.64 (199.40)	0.8	74.42 (190.70)

Table 2 (Continued)

ρ	ℓ	r	Q	Efficiency gain	$\hat{Q}$	Efficiency gain	Q	Efficiency gain
0.9	1	2	0.7	205.05 (454.63)	0.8	208.74 (461.35)	0.9	196.80 (439.63)
		3	0.7	166.55 (566.36)	0.8	180.66 (601.66)	0.9	180.10 (600.25)
		4	0.8	149.28 (667.00)	0.9	152.65 (677.38)	0.7	134.44 (621.36)
		6	0.8	108.12 (732.49)	0.9	113.13 (752.53)	0.7	96.43 (685.71)
		8	0.8	84.37 (767.61)	0.9	89.01 (789.45)	0.7	75.63 (726.50)
	2	2	0.7	195.35 (437.00)	0.8	205.05 (454.63)	0.9	196.65 (439.37)
		3	0.8	177.23 (593.07)	0.9	179.93 (599.83)	0.7	159.47 (548.67)
		4	0.8	146.70 (659.07)	0.9	152.50 (676.91)	0.7	130.12 (608.05)
		6	0.8	106.89 (727.55)	0.9	113.02 (752.10)	0.7	95.01 (680.03)
		8	0.8	83.82 (765.02)	0.9	88.94 (789.12)	0.7	75.18 (724.39)
	3	2	0.7	192.17 (431.21)	0.8	203.56 (451.93)	0.9	196.59 (439.25)
		3	0.8	176.12 (590.30)	0.9	179.86 (599.66)	0.7	157.86 (544.64)
		4	0.8	146.06 (657.11)	0.9	152.44 (676.74)	0.7	129.47 (606.05)
		6	0.8	106.72 (726.88)	0.9	113.00 (751.99)	0.7	94.92 (679.68)
		8	0.8	83.78 (764.84)	0.9	88.92 (789.06)	0.7	75.17 (724.34)
	4	2	0.7	190.99 (429.07)	0.8	202.88 (450.69)	0.9	196.55 (439.18)
		3	0.8	175.73 (589.32)	0.9	179.83 (599.58)	0.7	157.46 (543.66)
		4	0.8	145.89 (656.60)	0.9	152.42 (676.67)	0.7	129.36 (605.74)
		6	0.8	106.70 (726.79)	0.9	112.99 (751.96)	0.7	94.91 (679.65)
	6	2	0.7	190.35 (427.92)	0.8	202.37 (449.77)	0.9	196.52 (439.12)
		3	0.8	175.53 (588.83)	0.9	179.81 (599.52)	0.7	157.34 (543.35)
	8	2	0.7	190.26 (427.74)	0.8	202.24 (449.54)	0.9	196.50 (439.10)
		3	0.8	175.50 (588.76)	0.9	179.80 (599.50)	0.7	157.33 (543.33)
	10	2	0.7	190.24 (427.71)	0.8	202.21 (449.47)	0.9	196.50 (439.09)

Table 3. Values of the efficiency criterion λ in per cent near the optimum Q values

ρ	ℓ	r	Q	λ	$\hat{Q}'$	λ'	Q	λ
0.5	1	2	0.1	109.10	0.2	112.94	0.3	108.60
		3	0.2	110.86	0.3	110.99	0.4	105.73
		4	0.2	108.78	0.3	109.60	0.4	106.86
		6	0.2	106.25	0.3	107.15	0.4	105.96
		8	0.2	104.84	0.3	105.62	0.4	104.92
	2	2	0.1	109.09	0.2	113.01	0.3	108.90
		3	0.2	110.86	0.3	111.02	0.4	105.82
		4	0.2	108.78	0.3	109.60	0.4	106.87
0.6	1	2	0.2	121.43	0.3	123.10	0.4	114.52
		3	0.2	116.64	0.3	120.81	0.4	119.83
		4	0.3	117.03	0.4	117.55	0.5	113.19
		6	0.3	112.20	0.4	113.12	0.5	111.25
		8	0.3	109.46	0.4	110.31	0.5	109.23
	2	2	0.2	121.49	0.3	123.60	0.4	115.66
		3	0.2	116.63	0.3	120.85	0.4	120.02
		4	0.3	117.04	0.4	117.58	0.5	113.29
	3	2	0.2	121.49	0.3	123.61	0.4	115.73
0.7	1	2	0.2	131.64	0.3	141.72	0.4	141.63
		3	0.3	133.53	0.4	139.09	0.5	137.14
		4	0.4	132.19	0.5	133.03	0.6	126.18
		6	0.4	123.04	0.5	124.62	0.6	121.93
		8	0.4	117.84	0.5	119.26	0.6	117.82
	2	2	0.3	142.41	0.4	143.76	0.5	131.50
		3	0.3	133.55	0.4	139.43	0.5	138.15
		4	0.4	132.24	0.5	133.28	0.6	126.75
		6	0.4	123.04	0.5	124.63	0.6	122.00
	3	2	0.3	142.44	0.4	143.94	0.5	131.93
		3	0.3	133.55	0.4	139.44	0.5	138.19
		4	0.4	132.24	0.5	133.29	0.6	126.77
	4	2	0.3	142.44	0.4	143.95	0.5	131.98
0.8	1	2	0.4	182.68	0.5	183.12	0.6	163.75
		3	0.4	167.85	0.5	178.62	0.6	176.75
		4	0.5	164.94	0.6	168.24	0.7	156.90
		6	0.5	146.21	0.6	150.65	0.7	147.08
		8	0.5	135.62	0.6	139.33	0.7	137.99
	2	2	0.4	186.15	0.5	190.48	0.6	172.73
		3	0.5	180.70	0.6	181.26	0.7	160.79
		4	0.5	165.45	0.6	169.95	0.7	159.75

Table 3 (Continued)

ρ	ℓ	r	Q	λ	$\hat{Q}'$	λ'	Q	λ
0.8	2	6	0.5	146.25	0.6	150.88	0.7	147.75
		8	0.5	135.62	0.6	139.36	0.7	138.15
	3	2	0.4	186.54	0.5	191.75	0.6	174.90
		3	0.5	180.83	0.6	181.77	0.7	161.72
		4	0.5	165.47	0.6	170.04	0.7	160.06
		6	0.5	146.25	0.6	150.89	0.7	147.77
	4	2	0.4	186.58	0.5	191.96	0.6	175.41
		3	0.5	180.84	0.6	181.83	0.7	161.88
	6	2	0.4	186.58	0.5	192.00	0.6	175.55
		3	0.5	180.84	0.6	181.84	0.7	161.92
0.9	1	2	0.5	297.02	0.6	316.39	0.7	295.02
		3	0.6	297.50	0.7	308.78	0.8	271.19
		4	0.6	262.65	0.7	283.41	0.8	265.48
		6	0.6	214.85	0.7	235.48	0.8	234.95
		8	0.7	204.25	0.8	208.81	0.9	165.49
	2	2	0.5	311.59	0.6	344.36	0.7	327.11
		3	0.6	309.57	0.7	332.20	0.8	293.20
		4	0.6	266.94	0.7	296.06	0.8	282.56
		6	0.7	238.64	0.8	242.64	0.9	176.34
		8	0.7	205.06	0.8	211.93	0.9	166.66
	3	2	0.5	315.11	0.6	353.67	0.7	340.54
		3	0.6	311.66	0.7	338.51	0.8	301.36
		4	0.6	267.32	0.7	298.17	0.8	287.26
		6	0.7	238.85	0.8	243.74	0.9	176.78
		8	0.7	205.08	0.8	212.16	0.9	166.87
	4	2	0.5	315.85	0.6	356.49	0.7	345.98
		3	0.6	311.99	0.7	340.12	0.8	304.41
		4	0.6	267.36	0.7	298.51	0.8	288.53
		6	0.7	238.86	0.8	243.90	0.9	176.90
		8	0.7	205.08	0.8	212.18	0.9	166.91
	6	2	0.5	316.03	0.6	357.56	0.7	349.02
		3	0.6	312.05	0.7	340.63	0.8	305.96
		4	0.6	267.36	0.7	298.57	0.8	288.96
		6	0.7	238.86	0.8	243.92	0.9	176.95
	8	2	0.5	316.04	0.6	357.65	0.7	349.50
		3	0.6	312.05	0.7	340.66	0.8	306.18
	10	2	0.5	316.04	0.6	357.66	0.7	349.57

In Tables 1, 2, and 3 it was assumed that the correlation between $x_{\alpha,k}$ and $x_{\alpha+t,k}$ is given by a Markoff type lag correlogram, $S_{\alpha,\alpha+t} = \rho^{-t} S_\alpha S_{\alpha+t}$ with $S_\alpha = S_0$. It is of some interest to investigate the behaviour of the variance of the composite estimator $\bar{x}_0'$ when such a pattern does not hold. This will provide some indication as to the degree of trust that can be placed in the numerical results already given, for an exponential correlation pattern will at best be only an approximation to the true situation. Therefore the variance of $\bar{x}_0'$ was derived under the assumption of an arithmetic type lag correlogram,

$$S_{0,t} = (\rho + (t+1)d)S_0^2 \quad \text{when} -(t+1)d \leq \rho \ ,$$
$$= 0 \quad \text{when} -(t+1)d > \rho \ ,$$

with $S_\alpha = S_0$. For a one cycle design $V(\bar{x}_0')$ is given by (3.25). To further reduce the computational burden, (3.25) was examined for three and four visit designs only, the two visit variance being independent of any correlogram assumptions as substitution of $r = 2$ into (3.25) will show. Numerical results for $r = 3$ visits are presented in Table 4 and for $r = 4$ visits in Table 5. Here $\hat{Q}$ is the approximate optimum Q for the specified values of ρ and d. The common difference d is allotted the values 0.05, 0.10, 0.15, 0.20 and 0.30 and ρ the values 0.6, 0.7, 0.8 and 0.9. The efficiency gains registered by $\bar{x}_0'$ relative to $\bar{x}_0$ at $\hat{Q}$ are given as well as the values at $\hat{Q} \pm 0.1$. The efficiency gains under the exponential correlation (exp.) pattern assumption are also provided to facilitate a comparison. It is observed that:

(a) As would be anticipated, the more rapid the decrease in correlation between the same units with increasing time interval between observations, the less is the efficiency gain of $\bar{x}_0'$ over $\bar{x}_0$.

(b) Approximately the same degree of robustness of the optimum Q holds for the arithmetic correlation structure as did for the exponential pattern.

(c) The variance of $\bar{x}_0'$ would seem to be fairly robust with respect to the correlation structure assumed for values of $\rho \leq 0.8$. Some distortion is evident when $\rho = 0.9$.

Tables 1 to 5 should suffice to provide an adequate picture of the efficiency gains that can be derived by using composite estimators in rotation sampling designs. It has been concluded that two visits and one cycle provide an optimum design for estimating level. Thus if the minimum variance of the estimator $\bar{x}_0'$ was the sole criterion for selecting an appropriate design, the statistician would always choose a two visit one cycle rotation pattern. But such factors as cost, respondent cooperation, the estimation of change, etc., warrant a further study of other designs.

2. Finite population results

In the foregoing tabulations it has been assumed that N, the population size on any occasion, is large with respect to n, the sample size. Thus the finite population corrections (f.p.c.'s) in the variance and covariance formulas were safely ignored. But when the sampling

Table 4. Efficiency gains in per cent of $\bar{x}_0'$ over $\bar{x}_0$ with an arithmetic correlation pattern near the optimum Q values with r = 3 visits

ρ	d	Q	Efficiency gain	$\hat{Q}$	Efficiency gain	Q	Efficiency gain
0.9	0.05	0.6	52.79	0.7	56.12	0.8	41.62
	0.10	0.5	39.00	0.6	46.77	0.7	44.42
	0.15	0.5	36.36	0.6	41.22	0.7	34.35
	0.20	0.5	33.83	0.6	36.06	0.7	25.60
	0.30	0.4	25.54	0.5	29.03	0.6	26.81
	exp.	0.5	39.53	0.6	46.98	0.7	46.62
0.8	0.05	0.5	27.21	0.6	28.68	0.7	20.43
	0.10	0.4	21.70	0.5	25.00	0.6	24.39
	0.15	0.4	20.69	0.5	22.87	0.6	20.37
	0.20	0.4	19.70	0.5	20.81	0.6	16.61
	0.30	0.3	14.96	0.4	17.77	0.5	16.88
	exp.	0.4	20.49	0.5	22.45	0.6	19.04
0.7	0.05	0.4	15.28	0.5	15.38	0.6	11.14
	0.10	0.3	12.30	0.4	14.38	0.5	13.56
	0.15	0.3	11.92	0.4	13.49	0.5	11.80
	0.20	0.3	11.54	0.4	12.61	0.5	10.09
	0.30	0.3	10.79	0.4	10.90	0.5	6.82
	exp.	0.3	11.46	0.4	12.44	0.5	9.75
0.6	0.05	0.3	8.67	0.4	8.70	0.5	5.57
	0.10	0.2	6.63	0.3	8.31	0.4	7.89
	0.15	0.2	6.51	0.3	7.96	0.4	7.10
	0.20	0.2	6.39	0.3	7.61	0.4	6.32
	0.30	0.2	6.15	0.3	6.91	0.4	4.79
	exp.	0.2	6.29	0.3	7.33	0.4	5.70

Table 5. Efficiency gains in per cent of $\bar{x}_0'$ over $\bar{x}_0$ with an arithmetic correlation pattern near the optimum Q values with r = 4 visits

ρ	d	Q	Efficiency gain	$\hat{Q}$	Efficiency gain	Q	Efficiency gain
0.9	0.05	0.6	41.16	0.7	49.28	0.8	45.69
	0.10	0.6	35.32	0.7	36.97	0.8	23.00
	0.15	0.5	26.39	0.6	29.94	0.7	26.54
	0.20	0.5	23.96	0.6	24.97	0.7	17.58
	0.30	0.4	17.60	0.5	19.37	0.6	16.09
	exp.	0.6	36.77	0.7	40.02	0.8	28.43
0.8	0.05	0.5	22.34	0.6	25.91	0.7	24.84
	0.10	0.5	20.07	0.6	21.24	0.7	16.11
	0.15	0.4	15.53	0.5	17.87	0.6	16.90
	0.20	0.4	14.55	0.5	15.76	0.6	12.87
	0.30	0.4	13.58	0.5	13.71	0.6	9.10
	exp.	0.4	15.47	0.5	17.80	0.6	16.89
0.7	0.05	0.4	12.58	0.5	14.35	0.6	13.64
	0.10	0.4	11.65	0.5	12.35	0.6	9.82
	0.15	0.3	8.38	0.4	10.72	0.5	10.43
	0.20	0.3	8.00	0.4	9.82	0.5	8.57
	0.30	0.3	7.63	0.4	8.92	0.5	6.77
	exp.	0.3	8.68	0.4	9.88	0.5	8.80
0.6	0.05	0.3	6.40	0.4	8.01	0.5	7.33
	0.10	0.3	6.04	0.4	7.15	0.5	5.57
	0.15	0.3	5.68	0.4	6.30	0.5	3.87
	0.20	0.3	5.32	0.4	5.46	0.5	2.22
	0.30	0.2	4.72	0.3	4.97	0.4	4.64
	exp.	0.2	4.77	0.3	5.79	0.4	5.14

rate is not small one might inquire as to how the efficiency gains are affected. The current occasion variance, $V(\bar{x}'_0)$, is reduced by a factor S_0^2/N and the change variance, $V(d'_0)$, by a factor $2(1-\rho)S_0^2/N$ when the finite population is taken into account. Similar reductions also occur in $V(\bar{x}_0)$ and $V(\bar{x}_0 - \bar{x}_{-1})$ where $\bar{x}_0 - \bar{x}_{-1}$ refers to the change in level as estimated from the rotation design. Table 6 cites the efficiency gains using $\bar{x}'_0$ and d'_0 for several infinite cycle rotation designs with small N at the approximate optimum $\dot{Q}$ values. The optimum Q, r and ℓ values for a given ρ are, of course, not affected by the inclusion of the f.p.c. For convenience in comparison the efficiency gains when N is assumed infinite are also provided.

It is noted that in all cases the efficiency gain is greater when the f.p.c. is taken into account. This is indeed obvious for the same quantity is being subtracted from both numerator and denominator in the approximate efficiency ratio. For fixed N, n, r and ℓ the ratio of the exact efficiency gain to the approximate efficiency gain is increasing as ρ increases. For the change the situation is reversed; for increasing ρ the ratio decreases. The ratio is not subject to as much variation as ρ proceeds from 0.5 to 0.9 for the change estimate as it is for the level estimate. As the sampling fraction decreases the difference between the approximate and exact gains for a given ρ rapidly approaches zero. It is therefore concluded that Tables 1 and 2 which give the efficiency gains using $\bar{x}'_0$ and d'_0 are always on the conservative side and when the sampling fraction is not small the gains are considerably larger than those tabulated.

Table 6. Efficiency gains in per cent using $\bar{x}_0'$ and d_0' in finite populations

N	n	r	ℓ	ρ	Gain with $\bar{x}_0'$ and f.p.c.	Gain with $\bar{x}_0'$ without f.p.c.	Gain with d_0' and f.p.c.	Gain with d_0' without f.p.c.
4	2	2	1	0.5	10.88	5.16	20.45	12.76
				0.6	18.58	8.50	36.19	23.43
				0.7	33.17	14.23	63.90	42.83
				0.8	59.24	22.85	120.33	83.52
				0.9	126.70	38.78	290.20	208.74
6	2	2	2	0.5	8.11	5.26	16.49	12.37
				0.6	13.63	8.69	29.26	22.44
				0.7	24.80	15.27	52.10	40.81
				0.8	46.17	26.67	100.59	80.43
				0.9	101.73	50.64	251.55	205.05
6	3	3	1	0.5	8.83	4.23	19.18	11.18
				0.6	15.64	7.25	33.77	20.23
				0.7	27.21	11.97	59.95	36.87
				0.8	50.63	20.20	114.37	72.17
				0.9	115.72	36.65	278.30	180.66
8	2	2	3	0.5	7.14	5.26	15.21	12.36
				0.6	11.94	8.70	27.12	22.38
				0.7	21.52	15.32	48.43	40.57
				0.8	39.68	27.07	93.97	79.88
				0.9	88.24	54.22	236.14	203.56
8	4	4	1	0.5	6.84	3.31	16.67	9.38
				0.6	12.26	5.78	29.28	16.84
				0.7	21.68	9.78	52.36	30.74

Table 6 (Continued)

N	n	r	ℓ	ρ	Gain with $\bar{x}'_0$ and f.p.c.	Gain with $\bar{x}'_0$ without f.p.c.	Gain with d'_0 and f.p.c.	Gain with d'_0 without f.p.c.
8	4	4	1	0.8	41.31	17.12	130.07	60.49
				0.9	95.87	32.40	249.70	152.65
9	3	3	2	0.5	6.49	4.23	15.39	11.11
				0.6	11.41	7.32	27.28	20.00
				0.7	19.88	12.43	48.74	36.29
				0.8	37.65	22.30	93.73	70.85
				0.9	87.80	45.28	234.68	179.93
10	2	2	4	0.5	6.67	5.26	14.54	12.36
				0.6	11.11	8.70	26.02	22.38
				0.7	19.91	15.32	46.59	40.54
				0.8	36.36	27.11	90.57	79.72
				0.9	80.81	55.65	227.99	202.88
10	6	2	1	0.5	13.98	5.16	23.25	12.76
				0.6	24.36	8.50	40.62	23.43
				0.7	45.22	14.23	70.86	42.83
				0.8	86.93	22.85	131.96	83.52
				0.9	231.80	38.78	314.76	208.74
12	3	3	3	0.5	5.73	4.23	14.03	11.11
				0.6	10.01	7.33	25.00	20.00
				0.7	17.30	12.44	44.83	36.24
				0.8	32.34	22.44	86.56	70.72
				0.9	73.76	46.71	218.00	179.86

Table 6 (Continued)

N	n	r	ℓ	ρ	Gain with $\bar{x}'_0$ and f. p. c.	Gain with $\bar{x}'_0$ without f. p. c.	Gain with d'_0 and f. p. c.	Gain with d'_0 without f. p. c.
12	4	4	2	0.5	5.04	3.31	13.22	9.36
				0.6	8.94	5.79	23.41	16.79
				0.7	15.59	9.88	42.21	30.61
				0.8	29.28	17.78	82.25	60.28
				0.9	71.72	38.59	205.79	152.50
12	6	6	1	0.5	4.77	2.33	12.76	6.91
				0.6	8.24	3.96	22.63	12.45
				0.7	14.36	6.70	40.83	22.78
				0.8	27.23	11.99	79.17	44.79
				0.9	70.34	26.02	197.19	113.13
14	2	2	6	0.5	6.19	5.26	13.84	12.36
				0.6	10.29	8.70	24.86	22.38
				0.7	18.34	15.32	44.68	40.54
				0.8	33.14	27.11	87.10	79.65
				0.9	72.34	56.20	219.63	202.37
15	3	3	4	0.5	5.35	4.23	13.33	11.11
				0.6	9.33	7.33	23.81	20.00
				0.7	16.05	12.44	42.80	36.24
				0.8	29.73	22.45	82.82	70.71
				0.9	66.47	46.93	209.09	179.83

H. Variance Estimation

This dissertation would be incomplete if no mention of the important problem of variance estimation was made. In practice S_{α}^{2} and $S_{\alpha,\alpha+t}$, the mean squares and mean products for occasions α and $\alpha+t$, are not known and must therefore be estimated from the sample data. Nor is a strict exponential or arithmetic correlation trend likely to be entirely realistic. In the single-stage sample designs already considered the estimation of S_{α}^{2} and $S_{\alpha,\alpha+t}$ is straightforward. Let s_{α}^{2} and $s_{\alpha,\alpha+t}$ denote the unbiased estimators of S_{α}^{2} and $S_{\alpha,\alpha+t}$.

$$s_{\alpha}^{2} = \sum_{k=1}^{n} (x_{\alpha,k} - \bar{x}_{\alpha})^{2} / (n-1),$$

$$s_{\alpha,\alpha+t} = \sum_{k=1}^{m} (x_{\alpha,k} - {}_{m}\bar{x}_{\alpha})(x_{\alpha+t,k} - {}_{m}\bar{x}_{\alpha+t})/(m-1),$$

where $\bar{x}_{\alpha} = \sum_{k=1}^{n} x_{\alpha,k}/n$, ${}_{m}\bar{x}_{\alpha} = \sum_{k=1}^{m} x_{\alpha,k}/m$,

and m is the number of units matched between occasions α and $\alpha+t$. The estimators so obtained are substituted into (3.13) for the corresponding population values. Now (3.13) was derived under the assumption that $-u$, the occasion on which the first enumeration in the rotation plan took place, is effectively at $-\infty$. This expression therefore involves variance and covariance parameters which refer to occasions prior to the initial interview series and for which estimates are not therefore available. Little bias would be introduced in practice if the infinite sums occurring in (3.13) were truncated at $\alpha = -u$ provided that u is at least moderately large. Since the terms so omitted are weighted by factors $Q^{\alpha} << 1$, their contribution should be small. As an alternative to truncation one could set $s_{\alpha}^2 = s_{-u}^2$, $s_{\alpha, \alpha-t} = s_{-u, -u+t}$ for $\alpha < -u$, $t = 1, 2, \ldots$, with negligible bias for the same reason as above. This would permit the calculation of the omitted infinite sums as a correction to the truncated solution. If a stationary covariance structure with an exponential correlation pattern is an adequate approximation to the true situation then formula (3.16) could be adopted. The variance and correlation parameters could be estimated by pooling the data from the available occasions in an obvious manner.

Variance stimation with respect to the change between the current and previous occasions may be similarly dealt with.

A strictly unbiased method of variance estimation is now given. The technique involved is of theoretical interest and for this reason it was deemed fit to describe the principle here. The practical utility of the final result is questionable as the ensuing discussion will reveal.

The variance of $\bar{x}_0'$ is by definition

$$V(\bar{x}_0') = E(\bar{x}_0' - \bar{X}_0)^2 = E(\bar{x}_0')^2 - \bar{X}_0^2 .$$

An estimator of $V(\bar{x}_0')$ is therefore

$$v(\bar{x}_0') = (\bar{x}_0')^2 - (\widehat{\bar{X}_0^2}) , \tag{3.45}$$

and $v(\bar{x}'_0)$ will be unbiased provided that an unbiased estimator of $\bar{X}_0^2$ is employed. Since

$$\bar{X}_0 = \sum_{k=1}^{N} x_{0,k}/N$$

it follows that

$$\bar{X}_0^2 = \Big(\sum_{k=1}^{N} x_{0,k}^2 + \sum_{k \neq k' = 1}^{N} x_{0,k}\, x_{0,k'}\Big)/N^2 \quad .$$

An unbiased estimator of $\sum_{k=1}^{N} x_{0,k}^2$ is obviously $\sum_{k=1}^{N} w_{0,k} x_{0,k}^2 / E(w_{0,k})$

and an unbiased estimator of $\sum_{k \neq k' = 1}^{N} x_{0,k}\, x_{0,k'}$ is

$$\sum_{k \neq k' = 1}^{N} (w_{0,k}\, w_{0,k'}\, x_{0,k}\, x_{0,k'})/E(w_{0,k}\, w_{0,k'}) \, .$$

From (3.5) and (3.9) it is seen that

$$E(w_{0,k}) = 1/N, \quad E(w_{0,k} w_{0,k'}) = (n-1-n_2 Q^2/n_1)/nN(N-1) \, ,$$

and so

$$N^2 \widehat{(\bar{X}_0^2)} = N \sum_{k=1}^{N} w_{0,k}\, x_{0,k}^2$$

$$+ (N(N-1)n \sum_{k \neq k' = 1}^{N} w_{0,k}\, w_{0,k'}\, x_{0,k}\, x_{0,k'})/(n-1-n_2 Q^2/n_1) \, .$$

Substitution of the weights $w_{0,k}$ into $(\widehat{\bar{X}_0^2})$, and summing finally gives, by virtue of (3.45),

$$v(\bar{x}_0') = (\bar{x}_0')^2 - \Big(\sum_{k=1}^{n_2}(1-Q)x_{0,k}^2 + \sum_{k=1}^{n_1}(1+n_2Q/n_1)x_{0,k}^2\Big)/(Nn)$$

$$+ (N-1)\Big(\sum_{k\neq k'=1}^{n_2}(1-Q)^2 x_{0,k}x_{0,k'} + \sum_{k\neq k'=1}^{n_1}(1+n_2Q/n_1)^2 x_{0,k}x_{0,k'}$$

$$+ \sum_{k=1}^{n_2}\sum_{k=1}^{n_1}(1-Q)(1+n_2Q/n_1)x_{0,k}x_{0,k'}\Big)/((n-1-n_2Q^2/n_1)Nn). \qquad (3.46)$$

The index n_2 in the summations of (3.46) refer to the n_2 units in the sample for the first visit of a cycle, and the index n_1 to the remaining sample units.

As a check on the procedure Q may be set equal to zero in (3.46) so that $\bar{x}_0' = \bar{x}_0$. Then (3.46) reduces to

$$V(\bar{x}_0') = (1/n - 1/N)s_0^2$$

which is an unbiased estimator of $V(\bar{x}_0)$.

Note that (3.46) is an unbiased estimator of $V(\bar{x}_0')$ which makes no assumptions about the correlation structure over time. It is disturbing that the estimator of the term $\bar{X}_0^2$ involves only the current occasion information. Hence one would expect that this estimator of $V(\bar{x}_0')$ is lacking in precision. Rao (1962b) employed the above principle to develop an unbiased estimator of the variance of a ratio which ignores the concomitant information in one component. Using the familiar cities data of Cochran (1953), he demonstrated that the unbiased estimator furnishes variance estimates which differ greatly from those provided by the usual approximate variance formula. Hence an estimator of $V(\bar{x}_0')$ which makes use of the sample data from previous occasions to a larger extent would be preferred.

IV. MULTI-STAGE DESIGNS AND COMPOSITE RATIO ESTIMATORS WHEN SAMPLING ON SUCCESSIVE OCCASIONS

A. Multi-stage Sample Designs

In the sampling of human populations the costs involved in travelling from one sampling unit to another to collect information might well be prohibitive under a simple random sample design. Multi-stage sample designs have therefore been developed so that a survey of a given size may be conducted more quickly and economically. It is anticipated that the savings so introduced would more than offset the loss in precision resulting from the clustering of sampled units.

We shall consider in some detail the theory of sampling on successive occasions with partial replacement of units as applied to two-stage sample designs. The extension to multi-stage designs will be clear and no elaboration will be required.

1. Simple two-stage sampling

A two-stage sample design is characterized by a two-stage hierarchy of sampling units. There are N primary sampling units (p.s.u.'s) in the population P and the k-th primary $(k = 1, 2, \ldots, N)$ contains M_k secondary sampling units (secondaries) where $\sum_{k=1}^{N} M_k = M$. A sample of n p.s.u.'s is selected from the N p.s.u.'s and from the j-th selected primary $(j = 1, 2, \ldots, n)$ a sample of m_j secondaries is selected, $\sum_{j=1}^{n} m_j = m$. The selection at either stage may be made either with equal probability or with unequal probability for the respective units and the form of the estimator for some character x will vary accordingly. It will be assumed that the reader is acquainted with the structure of these estimators and their variances; they are readily available in, e.g., Sukhatme (1954) and Cochran (1953). When sampling on successive occasions it will be further assumed that there is no immigration or migration of units into or from P. Thus M_k $(k = 1, 2, \ldots, M)$ is constant on all sampling occasions $\alpha = 0, -1, -2, \ldots,$.

2. Rotation of primary sampling units

A rotation plan for p.s.u.'s is established in precisely the same manner as that described in Chapter III for one-stage rotation sampling designs. On any occasion α $(= 0, -1, -2, \ldots,)$ there are n p.s.u.'s in the sample of which n_1 are matched with the previous occasion and n_2 have entered the sample for the first visit of some cycle, $n_1 + n_2 = n$.

It will be assumed that m_k secondaries are selected for observation from the k-th primary with equal probability and without replacement. In order to give statistical validity to the discussion below the m_k secondaries to be sampled from the k-th primary are specified for all N primaries in P. This is the technique followed by Hartley (1959) in his study of simple multi-stage designs.

The composite estimator of the population total X_0 on the current occasion is

$$\hat{X}'_0 = Q(\hat{X}'_{-1} + \hat{X}_{0,-1} - \hat{X}_{-1,0}) + (1-Q)\hat{X}_0 , \qquad (4.1)$$

where $$0 \leq Q \leq 1 ,$$

$$\hat{X}_0 = \frac{N}{n} \sum_{k=1}^{n} \frac{M_k}{m_k} \sum_{j=1}^{m_k} x_{0,k,j} ,$$

$$\hat{X}_{0,-1} = \frac{N}{n_1} \sum_{k=1}^{n_1} \frac{M_k}{m_k} \sum_{j=1}^{m_k} x_{0,k,j} , \tag{4.2}$$

$$\hat{X}_{-1,0} = \frac{N}{n_1} \sum_{k=1}^{n_1} \frac{M_k}{m_k} \sum_{j=1}^{m_k} x_{-1,k,j} ,$$

and $x_{\alpha,k,j}$ is the observation on the j-th secondary of the k-th primary on occasion α, and $\hat{X}'_{-1}$ is the composite estimator of the population total X_{-1} on the previous occasion. Thus $\hat{X}_0$ is the simple two-stage estimator of the population total on the current occasion with both primaries and secondaries selected with equal probability and without replacement. $\hat{X}_{0,-1}$ and $\hat{X}_{-1,0}$ are similarly defined as estimators for the population totals on the current and previous occasions respectively using only the n_1 matched primaries. Let

$$\begin{aligned} v_{\alpha,k,j} &= M_k/m_k \text{ with probability } m_k/M_k , \\ &= 0 \text{ with probability } 1 - m_k/M_k , \end{aligned} \tag{4.3}$$

where $j = 1, 2, \ldots, M_k$; $k = 1, 2, \ldots, N$; $\alpha \leq 0$. Then $E(v_{\alpha,k,j}) = 1$.

Let $(w_{\alpha,k})$ denote the set of rotation weight variables that are employed in one-stage composite estimation of the sample mean. With an infinite cycle design initiated on occasion $\alpha = -u < 0$ the set would be given by (3.5), (3.6) and (3.7). It follows that $\hat{X}'_0$ may be written in the alternative form

$$\hat{X}_0' = \sum_{a=0}^{-u} \sum_{k=1}^{N} N w_{a,k} \sum_{j=1}^{M_k} v_{a,k,j}\, x_{a,k,j} \, . \tag{4.4}$$

It is clear that this estimator is of the form of a one-stage composite estimator of the total X_0 in a rotation design using n estimated primary totals on each occasion as the units of observation. Let u be large so that the approximation

$$\hat{X}_0' = \sum_{a=0}^{-\infty} \sum_{k=1}^{N} N w_{a,k} \sum_{j=1}^{M_k} v_{a,k,j}\, x_{a,k,j} \tag{4.5}$$

is valid. The expectation of $\hat{X}_0'$ will be taken over the $N!$ possible rotation patterns for primary sampling units compounded with the $\binom{M_k}{m_k}$ possible selections of m_k secondaries from M_k secondaries for $k = 1, 2, \ldots, N$. The $w_{a,k}$'s and the $v_{a,k,j}$'s are independent because of the device of specifying the sampled secondaries in each of the N primaries. Thus

$$E(\hat{X}_0') = \sum_{a=0}^{-\infty} \sum_{k=1}^{N} NE(w_{a,k}) \sum_{j=1}^{M_k} E(v_{a,k,j})\, x_{a,k,j}$$

$$= \sum_{k=1}^{N} \sum_{j=1}^{M_k} x_{0,k,j} = X_0 \, ,$$

since $NE(w_{0,k}) = 1$, $NE(w_{a,k}) = 0$ for $a < 0$, and $E(v_{a,k,j}) = 1$. Thus $\hat{X}_0'$ is an unbiased estimator of X_0.

In order to obtain the variance of $\hat{X}_0'$, $V(\hat{X}_0')$, write

$$\hat{X}'_0 = \sum_{a=0}^{-\infty} \sum_{k=1}^{N} N w_{a,k} \sum_{j=1}^{M_k} v_{a,k,j} (x_{a,k,j} - \overline{X}_{a,k})$$

$$+ \sum_{a=0}^{-\infty} \sum_{k=1}^{N} N w_{a,k} \sum_{j=1}^{M_k} v_{a,k,j} \overline{X}_{a,k}, \qquad (4.6)$$

where $\overline{X}_{a,k}$ is the population mean of the k-th primary on the a-th occasion. On setting

$$U_{a,k} = \sum_{j=1}^{M_k} v_{a,k,j} x_{a,k,j} - X_{a,k} \qquad (4.7)$$

where $X_{a,k} = M_k \overline{X}_{a,k}$, (4.6) may be written as

$$\hat{X}'_0 = \sum_{a=0}^{-\infty} \sum_{k=1}^{N} N w_{a,k} U_{a,k} + \sum_{a=0}^{-\infty} \sum_{k=1}^{N} N w_{a,k} X_{a,k} = A + B. \qquad (4.8)$$

Now $U_{a,k}$ and $w_{a,k}$ are statistically independent since $U_{a,k}$ is a function of $v_{a,k,j}$ which is independent of $w_{a,k}$. Further

$$E(U_{a,k}) = 0, \qquad (4.9)$$

and

$$V(\hat{X}'_0) = V(A) + V(B) + 2\,\mathrm{Cov}(A, B). \qquad (4.10)$$

Now

$$V(B) = V\left(\sum_{a=0}^{-\infty} \sum_{k=1}^{N} N w_{a,k} X_{a,k} \right)$$

is N^2 times the variance of a one-stage composite estimator of the population mean in a rotation design with the unit of observation being the primary total. Referring to (3.12) and assuming an infinite cycle rotation design for primaries, it immediately follows that

$$V(B) = N^2\left(\frac{1}{n} - \frac{1}{N} + \frac{Q^2 n_2}{nn_1}\right) \sum_{k=1}^{N} (X_{0,k} - X_0^*)^2/(N-1)$$

$$+ N^2 \sum_{a=-1}^{-\infty} \frac{n_2}{nn_1} Q^{-2a}\left(Q^2 + \frac{2n_2 Q}{n_1} + 1\right) \sum_{k=1}^{N} (X_{a,k} - X_a^*)^2/(N-1)$$

$$+ 2N^2 \sum_{a=0}^{-\infty} \sum_{t=1}^{\infty} NE(w_{a,k} w_{a-t,k}) \sum_{k=1}^{N} (X_{a,k} - X_a^*)(X_{a-t,k} - X_{a-t}^*)/(N-1), \tag{4.11}$$

where

$$X_a^* = \sum_{k=1}^{N} X_{a,k}/N .$$

Further,

$$V(A) = V\left(\sum_{a=0}^{-\infty} \sum_{k=1}^{N} N w_{a,k} U_{a,k}\right)$$

$$= N^2 \sum_{a=0}^{-\infty} \sum_{k=1}^{N} V(w_{a,k} U_{a,k}) + N^2 \sum_{a=0}^{-\infty} \sum_{\substack{k \neq k' \\ =1}}^{N} \mathrm{Cov}(w_{a,k} U_{a,k}, w_{a,k'} U_{a,k'})$$

$$+ N^2 \sum_{\substack{a \neq a' \\ =0}}^{-\infty} \sum_{k=1}^{N} \mathrm{Cov}(w_{a,k} U_{a,k}, w_{a',k} U_{a',k})$$

$$+ N^2 \sum_{\substack{a \neq a' \\ =0}}^{-\infty} \sum_{\substack{k \neq k' \\ =1}}^{N} \mathrm{Cov}(w_{a,k} U_{a,k}, w_{a',k'} U_{a',k'}) . \tag{4.12}$$

The second and fourth terms on the right are zero due to the independence of $w_{a,k}$, $U_{a,k}$ and $U_{a,k'}$. The $U_{a,k}$ and $U_{a',k}$ are not, however, independent when the same secondaries are selected from a primary over time and hence the third term on the right does not vanish.

Using the product rule for the variance of uncorrelated random variables, the first term on the right of (4.12) is

$$N^2 \sum_{a=0}^{-\infty} \sum_{k=1}^{N} V(w_{a,k} U_{a,k}) = N^2 \sum_{a=0}^{-\infty} \sum_{k=1}^{N} (E^2(w_{a,k}) V(U_{a,k})$$

$$+ E^2(U_{a,k}) V(w_{a,k}) + V(w_{a,k}) V(U_{a,k}))$$

$$= N^2 \sum_{a=0}^{-\infty} \sum_{k=1}^{N} E(w_{a,k}^2) V(U_{a,k})$$

$$= N^2 \sum_{k=1}^{N} E(w_{0,k}^2) V(U_{0,k}) + \sum_{a=-1}^{-\infty} \sum_{k=1}^{N} E(w_{a,k}^2) V(U_{a,k}) \,. \qquad (4.13)$$

But $V(U_{a,k}) = V(\sum_{j=1}^{M_k} v_{a,k,j}\, x_{a,k,j})$ is the variance of the estimated total of the k-th primary on occasion a from a simple random sample of m_k secondaries selected with equal probability and without replacement from M_k secondaries, and hence

$$V(U_{a,k}) = M_k^2 \left(\frac{1}{m_k} - \frac{1}{M_k}\right) S_{a,k}^2 \,,$$

where

$$S^2_{a,k} = \sum_{j=1}^{M_k} (x_{a,k,j} - \overline{X}_{a,k})^2 / (M_k - 1) ,$$

and (4.13) is therefore known. Also

$$\mathrm{Cov}(w_{a,k} U_{a,k}, w_{a',k} U_{a',k}) = E(w_{a,k} U_{a,k} w_{a',k} U_{a',k})$$

$$- E(w_{a,k} U_{a,k}) E(w_{a',k} U_{a',k}) = E(w_{a,k} w_{a',k}) E(U_{a,k} U_{a',k}) .$$

But $E(U_{a,k} U_{a',k}) = E(\hat{X}_{a,k} - X_{a,k})(\hat{X}_{a',k} - X_{a',k})$ is the covariance between the estimated totals of the k-th primary on occasions a and a' and hence

$$E(U_{a,k} U_{a',k}) = M_k^2 \left(\frac{1}{m_k} - \frac{1}{M_k} \right) S_{a,a',k}$$

where

$$S_{a,a',k} = \sum_{j=1}^{M_k} (x_{a,k,j} - \overline{X}_{a,k})(x_{a',k,j} - \overline{X}_{a',k}) / (M_k - 1) .$$

Also

$$\mathrm{Cov}(A, B) = \mathrm{Cov}\left(\sum_{a=0}^{-\infty} \sum_{k=1}^{N} N w_{a,k} U_{a,k}, \sum_{a=0}^{-\infty} \sum_{k=1}^{N} N w_{a,k} X_{a,k} \right)$$

$$= 0$$

due to the independence of $w_{a,k}$ and $U_{a,k}$ and because $E(U_{a,k}) = 0$. Therefore, by virtue of (4.10),

$$V(\hat{X}'_0) = N^2(\frac{1}{n} - \frac{1}{N} + \frac{n_2 Q^2}{nn_1}) \sum_{k=1}^{N} (X_{0,k} - X^*_0)^2/(N-1)$$

$$+ N^2 \frac{n_2}{nn_1} (Q^2 + 2 \frac{n_2 Q}{n_1} + 1) \sum_{a=-1}^{-\infty} Q^{-2a} \sum_{k=1}^{N} (X_{a,k} - X^*_a)^2/(N-1)$$

$$+ 2N^2 \sum_{a=0}^{-\infty} \sum_{t=1}^{\infty} NE(w_{a,k} w_{a-t,k}) \sum_{k=1}^{N} (X_{a,k} - X^*_a)(X_{a-t,k} - X^*_{a-t})/(N-1)$$

$$+ \frac{N}{n} (1 + \frac{n_2 Q^2}{n_1}) \sum_{k=1}^{N} M_k^2 (\frac{1}{m_k} - \frac{1}{M_k}) S^2_{0,k}$$

$$+ \frac{Nn_2}{nn_1} (Q^2 + 2 \frac{n_2 Q}{n_1} + 1) \sum_{a=-1}^{-\infty} Q^{-2a} \sum_{k=1}^{N} M_k^2 (\frac{1}{m_k} - \frac{1}{M_k}) S^2_{a,k}$$

$$+ 2N \sum_{k=1}^{N} M_k^2 (\frac{1}{m_k} - \frac{1}{M_k}) \sum_{a=0}^{-\infty} \sum_{t=1}^{\infty} NE(w_{a,k} w_{a-t,k}) S_{a,a-t,k} . \qquad (4.14)$$

The last term of (4.14) is determined when the rotation pattern is precisely specified. The form under an infinite cycle rotation design for primaries is complex and therefore omitted.

As a specific example consider the one cycle rotation design with a stationary Markoff type lag correlogram holding both between and within primaries. That is

$$S_a^2 = \sum_{k=1}^{N} (X_{a,k} - X^*_a)^2/(N-1) = \sum_{k=1}^{N} (X_{0,k} - X^*_0)^2/(N-1) = S_0^2 ,$$

$$\sum_{k=1}^{N} (X_{0,k} - X^*_0)(X_{a,k} - X^*_a)/(N-1) = \rho^{-a} S_0^2 ,$$

$$S^2_{a,k} = S^2_{0,k} , \quad S_{0,a,k} = \rho_k^{-a} S^2_{0,k} \qquad (4.15)$$

for all α, where ρ is the correlation between primary totals one occasion apart and ρ_k is the correlation between observations on secondaries of the k-th primary one occasion apart. Referring to the derivation of (3.16) for the cross-product expectations of the $w_{\alpha,k}$'s it will be seen that (4.14) becomes

$$V(\hat{X}_0') = N^2 S_0^2 \Big[\frac{1}{n} - \frac{1}{N} + \frac{2Q}{r^2(r-1)^2 n_2 (1-Q^2)(1-Q\rho)^2} (-(r-1)^2 \rho$$

$$+ r(r-1)Q - 2(r-1)Q\rho + r(r-1)Q\rho^2 + rQ^2 - (r^2+1)Q^2\rho + rQ^2\rho^2$$

$$+ Q^r \rho^{r-1}(r(r-1)\rho - (r-1)\rho^2 - r^2 Q + 2rQ\rho - (r^2 - 2r + 2)Q\rho^2 + r(r-1)Q^2\rho$$

$$- (r-1)Q^2\rho^2))\Big] \quad + N \sum_{k=1}^{N} M_k^2 \left(\frac{1}{m_k} - \frac{1}{M_k}\right) S_{0,k}^2 \Big[\frac{1}{rn_2}$$

$$+ \frac{2Q}{r^2(r-1)^2 n_2(1-Q^2)(1-Q\rho_k)^2} (-(r-1)^2\rho_k + r(r-1)Q - 2(r-1)Q\rho_k$$

$$+ r(r-1)Q\rho_k^2 + rQ^2 - (r^2+1)Q^2\rho_k + rQ^2\rho_k^2 + Q^r \rho_k^{r-1}(r(r-1)\rho_k - (r-1)\rho_k^2$$

$$- r^2 Q + 2rQ\rho_k - (r^2 - 2r + 2)Q\rho_k^2 + r(r-1)Q^2\rho_k - (r-1)Q^2\rho_k^2)\Big] . \qquad (4.16)$$

When independent samples are selected on each occasion within each primary then $S_{\alpha, \alpha-t, k} \doteq 0$ and (4.16) simplifies somewhat. The second term of (4.16) will become approximately

$$N \sum_{k=1}^{N} M_k^2 \left(\frac{1}{m_k} - \frac{1}{M_k}\right)\left(1 + \frac{2Q^2(Q+r-1)}{(r-1)^2(1-Q^2)}\right) S_{0,k}^2 / n \qquad (4.17)$$

which is always greater than $N \sum_{k=1}^{N} M_k^2 (1/m_k - 1/M_k) S_{0,k}^2 /n$. Now the entire expression within the square brackets of (4.16) will possibly be less than $1/n$ and in such cases it would be desirable to maintain the same set of secondaries within each primary over time from an efficiency point of view.

When $Q = 0$, then

$$V(\hat{X}_0') = N^2 (\frac{1}{n} - \frac{1}{N}) S_0^2 + \frac{N}{n} \sum_{k=1}^{N} M_k^2 (\frac{1}{m_k} - \frac{1}{M_k}) S_{0,k}^2$$

$$= V(\hat{X}_0)$$

where $\hat{X}_0$ is the two-stage estimator of X_0 using current occasion information only. This conforms with the reduction of the composite estimator to $N \sum_{k=1}^{n} (M_k/m_k) \sum_{j=1}^{m_k} x_{0,k,j} /n$.

Unequal probability sampling of secondaries may be dealt with through a redefinition of the secondary weight variables $v_{a,k,j}$. The extension of the foregoing principles to several stages of sampling should be obvious as well.

3. Rotation of secondary sampling units

On the first sampling occasion $-u$ a random sample of n primaries is selected from the N p.s.u.'s in P with equal probability and without replacement. This sample remains fixed thereafter. Within each selected primary a rotation pattern for secondaries is established as in the Chapter III discussion. In the k-th primary of size M_k, m_k secondaries are observed on occasion α ($\alpha = 0, -1, -2, \ldots, -u$) with m_{k1} secondaries common to occasions α and $\alpha-1$ and m_{k2} new secondaries on occasion α, so that $m_{k1} + m_{k2} = m_k$. The composite estimator

$$_k\hat{X}'_0 = Q_k\left({}_k\hat{X}'_{-1} + {}_k\hat{X}_{0,-1} - {}_k\hat{X}_{-1,0}\right) + (1-Q_k)\,{}_k\hat{X}_0 \tag{4.18}$$

is used to estimate the total $_kX_0$ of the k-th primary on the current occasion, where

$$_k\hat{X}_{0,-1} = M_k \sum_{j=1}^{m_{k1}} x_{0,k,j}/m_{k1} \qquad _k\hat{X}_{-1,0} = M_k \sum_{j=1}^{m_{k1}} x_{-1,k,j}/m_{k1}$$

$$_k\hat{X}_0 = M_k \sum_{j=1}^{m_k} x_{0,k,j}/m_k\,. \tag{4.19}$$

The method of Hartley (1959) will again be utilized to construct the composite estimator of X_0 and to determine its variance. Define the primary weight variables as follows:

$$\begin{aligned} s_k &= N/n \quad \text{with probability } n/N\,, \\ &= 0 \quad\;\; \text{with probability } 1-n/N\,, \end{aligned}$$

so that $E(s_k) = 1$.

Assuming that $-u$ is effectively at $-\infty$ for variance determination purposes, the composite estimator of X_0 is

$$\hat{X}'_0 = \sum_{\alpha=0}^{-u} \sum_{k=1}^{N} s_k \sum_{j=1}^{M_k} M_k\, w_{\alpha,k,j}\, x_{\alpha,k,j}$$

$$\doteq \sum_{\alpha=0}^{-\infty} \sum_{k=1}^{N} s_k \sum_{j=1}^{M_k} M_k w_{\alpha,k,j} x_{\alpha,k,j} . \tag{4.20}$$

The $w_{\alpha,k,j}$ is the rotation weight variable associated with the observation on the j-th secondary of the k-th primary on the α-th occasion. Analogous to (3.5) and (3.6) these are

$$\begin{aligned} w_{0,k,j} &= (1-Q_k)/m_k \text{ with probability } m_{k2}/M_k , \\ &= (1+\frac{m_{k2}}{m_{k1}} Q_k)/m_k \text{ with probability } m_{k1}/M_k , \\ &= 0 \text{ with probability } 1-m_k/M_k , \end{aligned}$$

and for $\alpha < 0$

$$\begin{aligned} w_{\alpha,k,j} &= -Q_k^{-\alpha}(Q_k + \frac{m_{k2}}{m_{k1}})/m_k \text{ with probability } m_{k2}/M_k , \\ &= Q_k^{-\alpha} m_{k2}(Q_k - 1)/m_k m_{k1} \text{ with probability } (m_{k1}-m_{k2})/M_k , \\ &= Q_k^{-\alpha}(1+\frac{m_{k2}}{m_{k1}} Q_k)/m_k \text{ with probability } m_{k2}/M_k , \\ &= 0 \text{ with probability } 1-m_k/M_k . \end{aligned}$$

Thus $E(w_{0,k,j}) = 1/M_k$, $E(w_{\alpha,k,j}) = 0$ for $\alpha < 0$.

By conceptually specifying the rotation plan within every primary, the s_k and $w_{\alpha,k,j}$ are independent for $\alpha = 0, -1, -2, \ldots,$. It then follows that $\hat{X}_0'$ is unbiased. Also $w_{\alpha,k,j}$ and $w_{\alpha,k',j}$ are independent for $k \neq k'$, but $w_{\alpha,k,j}$ and $w_{\alpha',k,j}$ are obviously dependent.

Now (4.20) may be written as

$$\hat{X}'_0 = \sum_{\alpha=0}^{-\infty} \sum_{k=1}^{N} s_k \sum_{j=1}^{M_k} M_k w_{\alpha,k,j} (x_{\alpha,k,j} - \overline{X}_{\alpha,k})$$

$$+ \sum_{k=1}^{N} s_k \sum_{j=1}^{M_k} M_k w_{0,k,j} \overline{X}_{0,k}$$

$$= \sum_{\alpha=0}^{-\infty} \sum_{k=1}^{N} s_k U_{\alpha,k} + \sum_{k=1}^{N} s_k X_{0,k} \tag{4.21}$$

where

$$U_{\alpha,k} = \sum_{j=1}^{M_k} M_k w_{\alpha,k,j} x_{\alpha,k,j} - X_{\alpha,k}, \quad E(U_{\alpha,k}) = 0,$$

and $U_{\alpha,k}$ is independent of s_k. Now $V(\sum_{k=1}^{N} s_k X_{0,k})$ is the variance of the estimated population total from a simple random of n units from N units and hence is

$$V(\sum_{k=1}^{N} s_k X_{0,k}) = N^2 (\frac{1}{n} - \frac{1}{N}) S_0^2 . \tag{4.22}$$

Further, let $U_k = \sum_{\alpha=0}^{-\infty} U_{\alpha,k}$, whence $E(U_k) = 0$. Then

$$V(\sum_{\alpha=0}^{-\infty} \sum_{k=1}^{N} s_k U_{\alpha,k}) = V(\sum_{k=1}^{N} s_k U_k)$$

$$= \sum_{k=1}^{N} V(s_k U_k) = \sum_{k=1}^{N} E(s_k^2) V(U_k)$$

where $E(s_k^2) = N/n$ and $V(U_k)$ is the variance of the composite estimator of the k-th primary total which can be written down immediately by analogy with Chapter III. Since

$$\text{Cov}\left(\sum_{\alpha=0}^{-\infty} \sum_{k=1}^{N} s_k U_k, \sum_{k=1}^{N} s_k X_{0,k} \right) = 0$$

it follows that

$$V(\hat{X}_0') = N^2 \left(\frac{1}{n} - \frac{1}{N}\right) S_0^2 + \frac{N}{n} \sum_{k=1}^{N} M_k^2 \left[\left(\frac{1}{m_k} - \frac{1}{M_k} + \frac{m_{k2}}{m_k m_{k1}} Q_k^2\right) S_{0,k}^2 \right.$$

$$+ \sum_{\alpha=-1}^{-\infty} \frac{m_{k2}}{m_k m_{k1}} Q_k^{-2\alpha} \left(Q_k^2 + 2 \frac{m_{k2}}{m_{k1}} Q_k + 1\right) S_{\alpha,k}^2 +$$

$$\left. 2 \sum_{\alpha=0}^{-\infty} \sum_{t=1}^{\infty} M_k E(w_{\alpha,k,j} w_{\alpha-t,k,j}) S_{\alpha,\alpha-t,k} \right]$$

$$= N^2 \left(\frac{1}{n} - \frac{1}{N}\right) S_0^2 + \frac{N}{n} \sum_{k=1}^{N} M_k^2 V(\bar{x}_{0,k}') . \qquad (4.23)$$

Here $\bar{x}_{0,k}'$ is the composite estimator of the current occasion mean of the k-th primary.

In particular, suppose that within the k-th primary a secondary stays in the sample for r_k consecutive occasions and then drops out forever. Assuming the usual stationary Markoff type lag correlogram within each primary, i.e., $S_{\alpha,\alpha-t,k} = \rho_k^{-t} S_{0,k}^2$, (4.23) becomes

$$V(\hat{X}_0') = N^2 \left(\frac{1}{n} - \frac{1}{N}\right) S_0^2 + \frac{N}{n} \sum_{k=1}^{N} M_k^2 S_{0,k}^2 \left[\frac{1}{m_k} - \frac{1}{M_k} \right.$$

$$+ 2Q_k(-(r_k - 1)^2 \rho_k + r_k(r_k - 1)Q_k - 2(r_k - 1)Q_k \rho_k + r_k(r_k - 1)Q_k \rho_k^2$$

$$+ r_k Q_k^2 - (r_k^2 + 1)Q_k^2 \rho_k + rQ_k^2 \rho_k^2 + Q_k^{r_k} \rho_k^{r_{k-1}}(r_k(r_k - 1)\rho_k - (r_k - 1)\rho_k^2$$

$$- r_k^2 Q_k + 2r_k Q_k \rho_k - (r_k^2 - 2r_k + 2)Q_k \rho_k^2 + r_k(r_k - 1)Q_k^2 \rho_k$$

$$- (r_k - 1)Q_k^2 \rho_k^2)/r_k^2(r_k - 1)^2 m_{k2}(1 - Q_k^2)(1 - Q_k\rho_k)^2 \Big] \quad . \tag{4.24}$$

Note that when rotation of secondaries within a fixed sample of primaries is employed, the composite estimators serve only to reduce the within primary component of the variance function. If it is a reasonable assumption that the correlation of secondaries between occasions is approximately the same within each primary, then the tables of Chapter III can be used to construct composite estimators within each selected primary with the same approximate optimum weight factors $Q_k = Q$. With rotation of primaries, the between primary component of variation is reduced under a reasonable choice of Q. There may, however, be definite cost advantages associated with maintaining a fixed set of primaries. For this reason a rotation of higher-stage sampling units is usually preferable.

There is no problem with variance estimation when employing composite estimators in multi-stage rotation designs. When secondaries are rotated within p.s.u.'s, the variance estimator based on the within primary composite estimators will be an unbiased estimator of the true sampling variance if primaries are selected with replacement and a slight overestimate if primaries are selected without replacement.

When the number of primary sampling units in P is small, individual rotation plans might well be established in every primary. Such a design is equivalent to stratified rotation sampling with each p.s.u. serving as a stratum.

B. Theory of Composite Ratio Estimators in One-stage Rotation Designs

Consider a one-stage rotation design where information is collected as well on a concomitant character y which is positively correlated with the main character x. If the y population mean on the current occasion, $\overline{Y}_0$, is known then a composite ratio estimator of the x mean, $\overline{X}_0$, on the current occasion is

$$\bar{\bar{x}}'_0 = \frac{\bar{x}'_0}{\bar{y}'_0} \overline{Y}_0 , \tag{4.25}$$

where $\bar{x}'_0$ and $\bar{y}'_0$ are the simple composite estimators

$$\bar{x}'_0 = Q_1(\bar{x}'_{-1} + \bar{x}_{0,-1} - \bar{x}_{-1,0}) + (1 - Q_1)\bar{x}_0 , \tag{4.26}$$

$$\bar{y}'_0 = Q_2(\bar{y}'_{-1} + \bar{y}_{0,-1} - \bar{y}_{-1,0}) + (1 - Q_2)\bar{y}_0 , \tag{4.27}$$

$$0 \leq Q_1, \; Q_2 \leq 1 .$$

The approximate variance of the ratio u/v, where u and v are unspecified pairs, is shown, e.g., by Cochran (1953), to be

$$V(\frac{x}{y}) \doteq \frac{1}{(E(y))^2} (V(x) + \frac{(E(x))^2}{(E(y))^2} V(y) - 2\frac{E(x)}{E(y)} Cov(x, y)). \tag{4.28}$$

Therefore

$$V(\bar{\bar{x}}_0') \doteq V(\bar{x}_0') + \left(\frac{\bar{X}_0}{\bar{Y}_0}\right)^2 V(\bar{y}_0') - 2\left(\frac{\bar{X}_0}{\bar{Y}_0}\right) \mathrm{Cov}(\bar{x}_0', \bar{y}_0') \tag{4.29}$$

$$= V(\bar{x}_0' - R\,\bar{y}_0') \quad \text{where} \quad R = \frac{\bar{X}_0}{\bar{Y}_0} .$$

To find $\mathrm{Cov}(\bar{x}_0', \bar{y}_0')$ let

$$\bar{x}_0' = \sum_{\alpha=0}^{-\infty} \sum_{k=1}^{N} w_{\alpha,k}\, x_{\alpha,k} , \tag{4.30}$$

$$\bar{y}_0' = \sum_{\alpha=0}^{-\infty} \sum_{k=1}^{N} w_{\alpha,k}'\, y_{\alpha,k} , \tag{4.31}$$

where the weights $w_{\alpha,k}$ and $w_{\alpha,k}'$ are identical for the same unit on the same occasion except for Q_2 in $w_{\alpha,k}'$ replacing Q_1 in $w_{\alpha,k}$. Then

$$\mathrm{Cov}(\bar{x}_0', \bar{y}_0') = E(\bar{x}_0'\, \bar{y}_0') - \bar{X}_0\, \bar{Y}_0$$

$$= NE(w_{0,k}\, w_{0,k}' - \frac{1}{N}) S_{x_0, y_0} + N \sum_{\alpha=-1}^{-\infty} E(w_{\alpha,k}\, w_{\alpha,k}') S_{x_\alpha, y_\alpha}$$

$$+ N \sum_{\alpha=0}^{-\infty} \sum_{t=1}^{\infty} \left(E(w_{\alpha,k}\, w_{\alpha-t,k}') + E(w_{\alpha-t,k}\, w_{\alpha,k}')\right) S_{x_\alpha, y_{\alpha-t}} \tag{4.32}$$

where

$$S_{x_\alpha, y_{\alpha-t}} = \left(\sum_{k=1}^{N} x_{\alpha,k}\, y_{\alpha-t,k} - N \bar{X}_\alpha \bar{Y}_{\alpha-t} \right) / (N-1) . \tag{4.33}$$

Since $V(\bar{x}_0')$ and $V(\bar{y}_0')$ have already been given in Chapter III, (4.29) will be known once the rotation pattern is specified. In practice one might take $Q_1 = Q_2 = Q$ in (4.26) and (4.27). It is evident however that an optimum choice of Q_1 and Q_2 which minimizes $V(\bar{\bar{x}}_0')$ will give at least as efficient an estimator as an optimization on Q alone.

If $Q_1 = Q_2 = Q$, then $w_{a,k} = w'_{a,k}$ and (4.32) becomes

$$\mathrm{Cov}(\bar{x}_0', \bar{y}_0') = (NE(w_{0,k}^2) - \frac{1}{N}) S_{x_0, y_0} + N \sum_{a=-1}^{-\infty} E(w_{a,k}^2) S_{x_a, y_a}$$

$$+ N \sum_{\substack{a \neq a' \\ =0}}^{-\infty} E(w_{a,k} w_{a',k}) S_{x_a, y_{a'}} . \tag{4.34}$$

Thus to obtain $\mathrm{Cov}(\bar{x}_0', \bar{y}_0')$ when $Q_1 = Q_2$ one need only substitute $S_{x_a, y_{a'}}$ for $S_{a,a'}$ in the formula (3.12) for $V(\bar{x}_0')$. It follows that in such a case the approximate variance of $\bar{\bar{x}}_0'$ can be procured by substituting $x_{a,k} - Ry_{a,k}$ for $x_{a,k}$ in $V(\bar{x}_0')$.

$\bar{\bar{x}}_0'$ will be more efficient than $\bar{x}_0'$ if $V(\bar{\bar{x}}_0') - V(\bar{x}_0') < 0$, or, from (4.29), if

$$RV(\bar{y}_0') - 2\,\mathrm{Cov}(\bar{x}_0', \bar{y}_0') < 0 . \tag{4.35}$$

Assuming that $Q_1 = Q_2$, and under the special correlation structure

$$S_{x_a, y_{a+t}} = S_{x_0, y_t} = \rho^{-t} S_{x_0, y_0} ,$$
$$S_{y_a, y_{a+t}} = S_{y_0, y_t} = \rho^{-t} S_{y_0}^2 , \tag{4.36}$$

condition (4.35) reduces to

$$R S_{y_0}^2 < 2 S_{x_0, y_0} \tag{4.37}$$

in view of the earlier statement about $\mathrm{Cov}(\bar{x}_0', \bar{y}_0')$. If also

$$S_{x_0, y_0} = \rho_{x,y} S_{x_0} S_{y_0}$$

then (4.37) becomes

$$R S_{y_0} < 2 \rho_{x,y} S_{x_0} ,$$

or
$$\rho_{x,y} > \frac{R S_{y_0}}{2 S_{x_0}} , \tag{4.38}$$

and if $S_{x_0} = S_{y_0}$, and $R = 1$,

$$\rho_{x,y} > \frac{1}{2} \tag{4.39}$$

where $\rho_{x,y}$ is the correlation between the measurements on the x and y characters on the same occasion. Result (4.39) is the same condition under which a ratio estimator $(\bar{x}\,\bar{Y})/\bar{y}$ is a more efficient estimator of $\bar{X}$ then $\bar{x}$ when $S_x = S_y$ and $R = 1$ as shown in, e.g., Cochran (1953). It will not be generally true that the correlations involved in the two assumptions of (4.36) will be equal. Then no general statement as (4.39) will necessarily follow and one must resort to an arithmetic study to determine the optimum properties of $\bar{\bar{x}}'_0$.

C. Ratio Estimators in Two-stage Sample Designs

The problem of estimating the population mean per secondary sampling unit when primaries are selected with equal probability and without replacement and secondaries are rotated within selected primaries will now be discussed. An obvious estimator of $\bar{X}_0$ is

$$_1\hat{\bar{X}}'_0 = \frac{\hat{X}'_0}{M} \tag{4.40}$$

where $\hat{X}'_0$ is the composite multi-stage estimator (4.20) of the current occasion population total X_0 and $M = \sum_{k=1}^{N} M_k$ is the total number of secondaries in the population and is assumed known. It is evident that $_1\hat{\bar{X}}'_0$ is unbiased with variance equal to $(1/M)^2\, V(\hat{X}'_0)$ where $V(\hat{X}'_0)$ is spelled out in general in (4.23). The between primary component of variation of $V(\hat{\bar{X}}'_0)$ is thus dependent upon the variation among primary totals,

$$V(_1\hat{\bar{X}}'_0) = (N^2(\frac{1}{n} - \frac{1}{N})\, \frac{\sum_{k=1}^{N} (X_{0,k} - X^*_0)^2}{N-1}$$

$$+ \frac{N}{n} \sum_{k=1}^{N} M_k^2\, V(\bar{x}'_{0,k}))/M^2 . \tag{4.41}$$

A second estimator of $\bar{X}_0$ is

$$ {}_2\hat{\bar{X}}'_0 = \frac{\sum_{k=1}^{n} M_k \bar{x}'_{0,k}}{\sum_{k=1}^{n} M_k} . \tag{4.42} $$

If all $M_i = \overline{M} = \sum_{k=1}^{N} M_k/N$, then (4.42) reduces to (4.40). From (4.28) the variance of ${}_2\hat{\bar{X}}'_0$ is approximately

$$ V({}_2\hat{\bar{X}}'_0) = (V(\sum_{k=1}^{n} M_k \bar{x}'_{0,k}) + R^2 V(\sum_{k=1}^{n} M_k) $$

$$ - 2R\,\mathrm{Cov}(\sum_{k=1}^{n} M_k \bar{x}'_{0,k}, \sum_{k=1}^{n} M_k))/(E(\sum_{k=1}^{n} M_k))^2, \tag{4.43} $$

where

$$ R = \frac{E(\sum_{k=1}^{n} M_k \bar{x}'_{0,k})}{E(\sum_{k=1}^{n} M_k)} . $$

Now

$$ E(\sum_{k=1}^{n} M_k \bar{x}'_{0,k}) = \underset{n}{E}(E(\sum_{k=1}^{n} M_k \bar{x}'_{0,k} \mid n)) = \underset{n}{E}(\sum_{k=1}^{n} M_k \overline{X}_{0,k}) = \frac{nX_0}{N} , $$

$$ E(\sum_{k=1}^{n} M_i) = n\overline{M} , $$

and hence $R = \overline{X}_0$. Now

$$V\left(\sum_{k=1}^{n} M_k \bar{x}'_{0,k}\right) = \underset{n}{V}\left(E\left(\sum_{k=1}^{n} M_k \bar{x}'_{0,k} \,\Big|\, n\right)\right) + \underset{n}{E}\left(V\left(\sum_{k=1}^{n} M_k \bar{x}'_{0,k} \,\Big|\, n\right)\right)$$

$$= n^2\left(\frac{1}{n} - \frac{1}{N}\right) \frac{\sum_{k=1}^{N} (X_{0,k} - X_0^*)^2}{N-1} + \frac{n}{N} \sum_{k=1}^{N} M_k^2 \, V(\bar{x}'_{0,k}), \tag{4.44}$$

$$V\left(\sum_{k=1}^{n} M_k\right) = n^2\left(\frac{1}{n} - \frac{1}{N}\right) \frac{\sum_{k=1}^{N} (M_k - \bar{M})^2}{N-1}, \tag{4.45}$$

$$\mathrm{Cov}\left(\sum_{k=1}^{n} M_k \bar{x}'_{0,k}, \sum_{k=1}^{n} M_k\right) = \underset{n}{E}\, \mathrm{Cov}\left(\sum_{k=1}^{n} M_k \bar{x}'_{0,k}, \sum_{k=1}^{n} M_k \,\Big|\, n\right)$$

$$+ \underset{n}{\mathrm{Cov}}\left(E\left(\sum_{k=1}^{n} M_k \bar{x}'_{0,k} \,\Big|\, n\right), E\left(\sum_{k=1}^{n} M_k \,\Big|\, n\right)\right)$$

$$= \underset{n}{\mathrm{Cov}}\left(\sum_{k=1}^{n} M_k \bar{X}_{0,k}, \sum_{k=1}^{n} M_k\right)$$

$$= n^2\left(\frac{1}{n} - \frac{1}{N}\right) \frac{\sum_{k=1}^{N} (M_k \bar{X}_{0,k} - X_0^*)(M_k - \bar{M})}{N-1}. \tag{4.46}$$

Thus, in view of (4.43), collecting (4.44), (4.45) and (4.46) and simplifying gives

$$V({}_2\hat{\bar{X}}'_0) \doteq \left(N^2\left(\frac{1}{n} - \frac{1}{N}\right) \frac{\sum_{k=1}^{N} M_k^2 (\bar{X}_{0,k} - \bar{X}_0)^2}{N-1} + \frac{N}{n} \sum_{k=1}^{N} M_k^2 \, V(\bar{x}'_{0,k})\right) / M^2. \tag{4.47}$$

A comparison of (4.47) and (4.41) reveals that their within primary components of variation agree. The between primary component of (4.47) depends upon the variation of primary means whereas the between primary component of (4.41) depends upon the variation of primary totals. Thus if the primary sizes M_k vary appreciably then ${}_2\hat{\bar{X}}'_0$ would be subject to a considerably smaller variance than would ${}_1\hat{\bar{X}}'_0$ and would therefore be preferred even though it is slightly biased. ${}_2\hat{\bar{X}}'_0$ also has the desirable feature that the total number of secondaries in the population, M, need not be known. It requires only that the M_k be known for the sampled primaries. If primaries are sampled with probability proportional to size, it can be shown that there is little additional gain from using ratio estimators.

It would be straying too far from the main subject matter to discuss the ramifications of the foregoing problems. Our intent was to illustrate the methods of approaching the various topics.

V. ROTATION SAMPLE DESIGNS INVOLVING A FINITE NUMBER OF CYCLES

A. The ℓ Cycle Rotation Design

In Chapter III attention was primarily devoted to infinite cycle rotation sample designs. A rotation group remains in the sample for $r \geq 2$ consecutive occasions, withdraws for $m \geq r$ consecutive occasions, returns for another r consecutive occasions, and this process continues indefinitely. It is in such a design that it is entirely realistic to speak of a finite population. Upon setting the recurrence time $m = \infty$, the one cycle rotation design, with its associated estimators and their variances, are obtained as a special case. The concept of a finite population becomes somewhat artificial since, sooner or later, all population units will be depleted by sampling and a rotation group must therefore return.

The most general systematic rotation design is the ℓ cycle design. Instead of recurring infinitely often a rotation group performs a total of ℓ cycles of r consecutive visits each in the sample. After each cycle the rotation group withdraws from the sample for $m \geq r$ occasions, and after the ℓ-th cycle it does not return again. Thus after the permanent pattern has been established, the sample size on any occasion is $\ell r n_2$ units of which rn_2 are in the first cycle, rn_2 in the second cycle, ..., and rn_2 in the ℓ-th and final cycle. Within each of the cycles there is a rotation group on each of the first, second, ..., r-th visits. Hence there are effectively ℓ one cycle rotation patterns simultaneously taking place. These rotation patterns are identical in nature except that they lag or precede one another by a multiple $k(m+r)$, $(k = 1, 2, \ldots, \ell-1)$, of consecutive sampling occasions, as an examination of Figure 2 for the case $r = 4$, $\ell = 2$ will reveal.

The same remarks concerning the finite population assumption are still valid. We continue to maintain the finite population correction in succeeding variance formulas with the rationalization that when a rotation group is forced to ultimately return into the sample in a moderately large population, the values then assumed should be essentially uncorrelated with any earlier sample values.

The simple composite estimator of the current occasion mean, $\overline{X}_0$, in an ℓ cycle design is still

$$\bar{x}'_0 = Q(\bar{x}'_{-1} + \bar{x}_{0,-1} - \bar{x}_{-1,0}) + (1 - Q)\bar{x}_0 \tag{5.1}$$

$$\doteq \sum_{\alpha=0}^{-\infty} \sum_{k=1}^{N} v_{\alpha,k}\, x_{\alpha,k}, \tag{5.2}$$

with variance

$$V(\bar{x}'_0) = (NE(v^2_{0,k}) - \frac{1}{N})S^2_0 + N \sum_{\alpha=-1}^{-\infty} E(v^2_{\alpha,k})S^2_\alpha$$

$$+ N \sum_{\substack{\alpha \neq \alpha' \\ =0}}^{-\infty} E(v_{\alpha,k} v_{\alpha',k}) S_{\alpha,\alpha'} . \tag{5.3}$$

It is assumed that the permanent pattern is well established on the current occasion so that, for variance determination purposes, the survey was effectively instituted on occasion $-u = -\infty$.

The $v_{\alpha,k}$ of (5.2) are related to the $w_{\alpha,k}$ of (3.2) in a simple manner. Since on any occasion α there are ℓ rotation groups on each of the 1st, 2nd, ..., r-th visits of their cycle, it follows immediately that

$$v_{\alpha,k} = w_{\alpha,k}/\ell \, , \ NE(v_{\alpha,k}^2) = NE(w_{\alpha,k}^2)/\ell \quad (\alpha = 0, -1, -2, \ldots,) .$$

The cross-product expectations, $NE(v_{\alpha,k} v_{\alpha' k})$ are somewhat more tedious to evaluate in an ℓ cycle design as opposed to an infinite cycle design. For example, when $\ell = 2$ there is no contribution to $NE(v_{\alpha,k} v_{\alpha',k})$ when $\alpha' < \alpha - r + 1$, $\alpha \leq 0$, from those rotation groups already in their second cycle on occasion α'.

An explicit expression for the variance of $\bar{x}_0'$ in the general ℓ cycle rotation design need not be presented here for it is in fact a direct extension of the infinite cycle variance function (3.13). It is only required to truncate the appropriate infinite sums of (3.13) to correspond to the $\ell < \infty$ cycles involved, and to introduce a dummy variable s to account for the varying number of product terms $NE(w_{\alpha,k} w_{\alpha',k})$ between rotation groups in different cycles on occasions α and α'. Thus, for example,

$$\sum_{s=0}^{\infty} Q^{s(r+m)+m+1} n_2(1 + n_2 Q/n_1)(1 - Q) S_{0,-s(r+m)-,-1} /n^2$$

becomes

$$\sum_{s=0}^{\ell-2} Q^{s(r+m)+m+1} n_2(1 + n_2 Q/n_1)(\ell - s - 1)(1 - Q) S_{0,-s(r+m)-m-1} / \ell n^2 ,$$

and so on.

B. The Current Population Survey

1. The survey design

The Current Population Survey (C.P.S.) conducted monthly by the United States Bureau of the Census employs a (two cycle) rotation sample design. A given rotation group remains in the sample for four consecutive months, withdraws for the next eight months, and then returns for another four months. It then drops out of the sample and does not return again. Hence, within any month one-eighth of the sample segments are enumerated for the first time, another eighth for the second time, etc., and the last eighth are interviewed for the eighth and final time. As between any two consecutive months seventy-five per cent of the segments are in common, and between the same months of any two consecutive years fifty per cent of the segments are in common. The sample design is illustrated in Figure 2, the current occasion being year 0 and month 0.

An examination of Figure 2 reveals that, in effect, two rotation patterns are simultaneously taking place. These patterns are identical in nature save for the fact that one lags the other in time by twelve months. Assuming that the permanent patterns have been underway for some time, there are four rotation groups in the first pattern or cycle and four different rotation groups in the second pattern or cycle. Within each of the cycles there is, on any given occasion, one rotation group on each of the first, second, third and fourth visits of that cycle. If the sample size on any occasion is n and the number of units on each visit number of the first and second cycles combined is n_2, so that

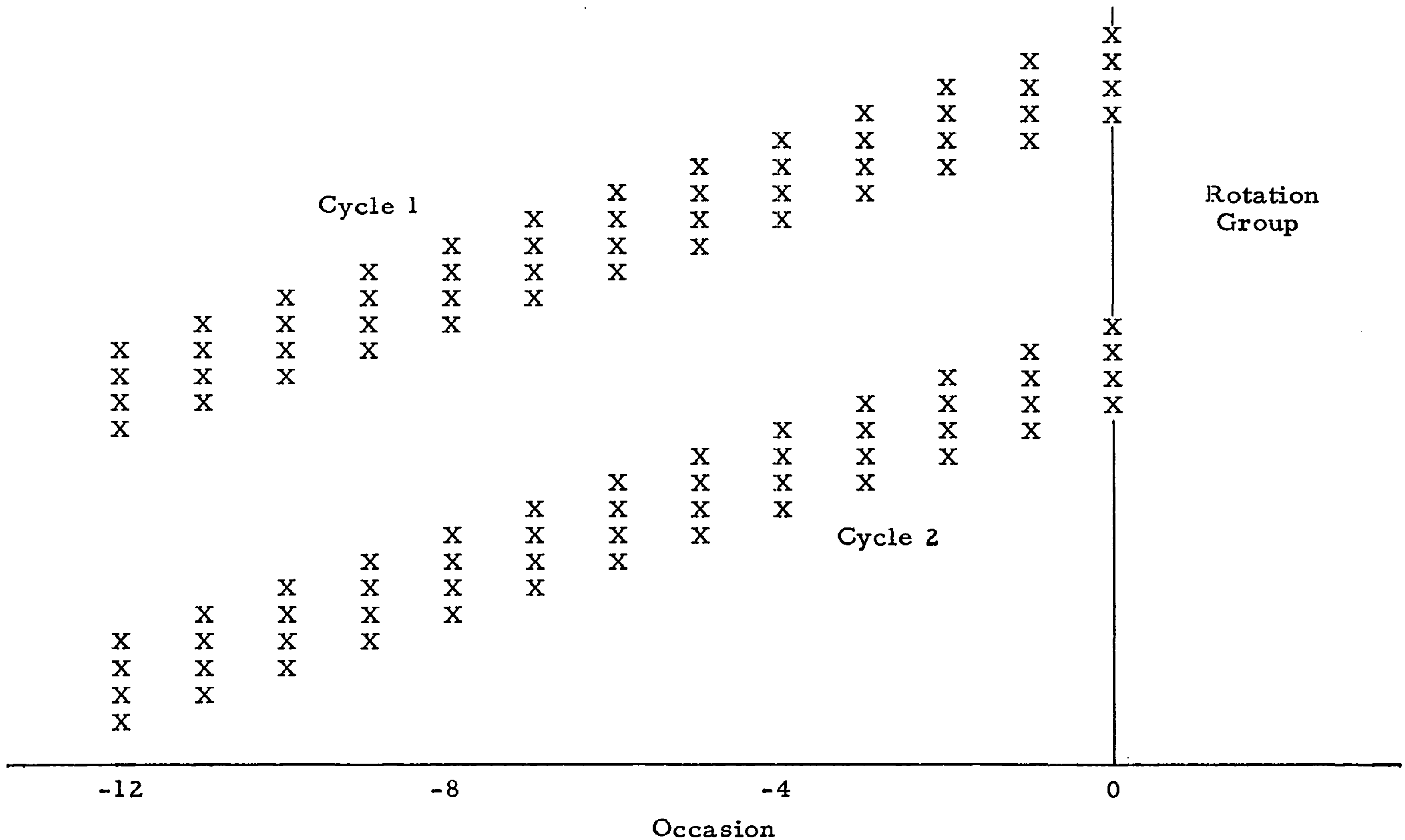

Figure 2. Rotation of sample in the U.S. Bureau of the Census Current Population Survey

$4n_2 = n$, then each cycle is individually identical to a four visit, one cycle $(m = \infty)$ design with sample size $2n_2$.

2. Correlogram assumption

The variance functions of Chapter III have been examined under the assumption that the correlation between $x_{\alpha, k}$ and $x_{\alpha', k}$ decreases steadily as $|\alpha - \alpha'|$ increases. There is evidence that such a state of affairs is not always the case. Eckler (1955) commented that in economic populations with month-to-month correlations ρ, the year-to-year correlation is often much larger than the ρ^{12} predicted by the exponential correlation model. Thus an underlying cyclic behavior of the population may upset the exponential correlation model unless the pattern length is a small part of the period. Tikkiwal (1956b) observed that in a quarterly series of livestock surveys, the correlations between quarters 1, 2, 3 and 4 and the 0-th (initial) quarter for the number of cattle on hand at the end of the quarter were respectively 0.97, 0.88, 0.65 and 0.90. Such a phenomenon would be anticipated in a stable agricultural economy where the holdings would be more-or-less uniform from year to year.

We are thereby led to consider the alternative correlation structure

$$\rho_{ij} = \rho_1^i \rho_2^j, \qquad (5.4)$$

where

$$i = 1, 2, \ldots, 10, 11; \quad j = 1, 2, 3, \ldots, .$$

ρ_{ij} is the population correlation coefficient between the value assumed by a sampling unit on a given occasion and that assumed by the same unit $12j+i$ months earlier, j being a year index and i a month index. It is assumed that $\rho_2 > \rho_1^{12}$ so that the correlogram is piecewise monotone decreasing within twelve month intervals starting from the current occasion. There is a positive discontinuity at each yearly point, the value of the saltus decreasing there as the time interval from the current occasion increases.

3. The simple composite estimator and its variance

The simple composite estimator of the current occasion mean $\overline{X}_0$ is given by (5.1) and alternately by (5.2). It follows that, for the special case $r = 4$, the rotation weight variables $v_{a,k}$ are as specified in Table 7. The $v_{a,k}$ are obviously independent of the cycle number and depend only upon the visit number within a cycle.

Since $NE(v_{0,k}) = 1$ and $NE(v_{a,k}) = 0$ for $a < 0$, $\overline{x}'_0$ is therefore an unbiased estimator of $\overline{X}_0$. The variance of $\overline{x}'_0$, $V(\overline{x}'_0)$, is obtained by evaluating the variance function (5.3) with the aid of Table 7. Some algebraic simplification will give

$$V(\overline{x}'_0) = (\frac{1}{n} - \frac{1}{N})S_0^2 + Q^2 S_0^2/(3n) + (3Q^2+2Q+3) \sum_{a=-1}^{-\infty} Q^{-2a} S_a^2/(9n)$$

$$+ (1+3Q)Q^{13} S_{0,-12}/(9n) - (3+Q)^2 Q(2S_{0,-1} + Q^{12} S_{0,-13})/(36n)$$

$$- (3+Q)(Q+1)Q^2 (2S_{0,-2} + Q^{12} S_{0,-14})/(18n)$$

$$- (3+Q)(1+3Q)Q^3 (2S_{0,-3} + Q^{12} S_{0,-15})/(36n) + (3+Q)(1-Q)Q^9 S_{0,-9}/(12n)$$

$$- (Q^2-6Q-3)Q^{10} S_{0,-10}/(18n) - (Q^2-14Q-3)Q^{11} S_{0,-11}/(36n)$$

$$- (Q-1)^2 Q \sum_{a=-1}^{-\infty} Q^{-2a} (2S_{a,a-1} + Q^{12} S_{a,a-13})/(36n)$$

$$- (Q-1)^2 Q^2 \sum_{a=-1}^{-\infty} Q^{-2a} (2S_{a,a-2} + Q^{12} S_{a,a-14})/(18n)$$

$$- (3+Q)(1+3Q) Q^3 \sum_{a=-1}^{-\infty} Q^{-2a} (2S_{a,a-3} + Q^{12} S_{a,a-15})/(36n)$$

$$- (3+Q)(1+3Q) Q^9 \sum_{a=-1}^{-\infty} Q^{-2a} S_{a,a-9}/(36n)$$

$$- Q^{10} (Q-1)^2 \sum_{a=-1}^{-\infty} Q^{-2a} S_{a,a-10}/(18n)$$

$$- Q^{11} (Q-1)^2 \sum_{a=-1}^{-\infty} Q^{-2a} S_{a,a-11}/(36n)$$

$$+ (3Q^2 + 2Q + 3) Q^{12} \sum_{a=-1}^{-\infty} Q^{-2a} S_{a,a-12}/(9n) . \tag{5.5}$$

Assuming that $S_a^2 = S_0^2$, $S_{a,a+t} = S_{0,t}$ and the correlation model (5.4) so that, for example, $S_{a,a-14} = \rho_1^2 \rho_2 S_0^2$, then (5.5) may be reduced to the form

$$V(\bar{x}_0') = (\frac{1}{n} - \frac{1}{N}) S_0^2 + S_0^2 \Big[8Q^2(Q+3) + (2+Q^{12}\rho_2)\{ 8Q^3$$

$$+ 7Q^2 - 6Q - 9 - 2Q\rho_1(-6Q^3 - Q^2 + 4Q + 3) - (3Q+1)(Q+3)Q^2\rho_1^2)Q\rho_1$$

$$+ 4(Q^2 + 6Q + 1)Q^{13}\rho_2 + (Q\rho_1)^9 (-4Q^3 - 15Q^2 - 6Q + 9 + 2Q\rho_1(-4Q^3 - 5Q^2$$

$$+ 6Q+3) - Q^2 \rho_1^2 (12Q^3 + 5Q^2 - 14Q - 3)) \,] \;/36n(1-Q^2). \qquad (5.6)$$

Table 7. Weights $v_{a,k}$ for $\bar{x}'_0$ with $r = 4$, $m = 8$, $\ell = 2$

Occasion a	Visit number within cycle	Weight $v_{a,k}$
0	1	$(1 - Q)/n$
	2 3 4	$(3 + Q)/3n$
	not present	0
< 0	1	$- Q^{-a}(1 + 3Q)/3n$
	2 3	$Q^{-a}(-1 + Q)/3n$
	4	$Q^{-a}(3 + Q)/3n$
	not present	0

When $x_{a,k} = x_{0,k}$, then $\rho_1 = \rho_2 = 1$, and

$$V(\bar{x}'_0) = (1-Q)^2 \, V(\sum_{t=0}^{\infty} Q^t \bar{x}_{-t})$$

$$= (1-Q)^2 \left(\sum_{t=0}^{\infty} Q^{2t} V(\bar{x}_{-t}) + 2 \sum_{\substack{t<s\\=0}}^{\infty} Q^{2t+s} \operatorname{Cov}(\bar{x}_{-t}, \bar{x}_{-t-s}) \right).$$

Substituting the relationships

$$V(\bar{x}_{-t}) = (\frac{1}{n} - \frac{1}{N})S_0^2$$

$$\begin{aligned}
Cov(\bar{x}_{-t}, \bar{x}_{-t-s}) &= (\frac{4-s}{4n} - \frac{1}{N})S_0^2 \quad \text{for} \quad s = 1, 2, 3, \\
&= -\frac{S_0^2}{N} \quad \text{for} \quad s = 4, 5, 6, 7, 8, \text{ and } s \geq 16, \\
&= (\frac{s-8}{8n} - \frac{1}{N})S_0^2 \quad \text{for} \quad s = 9, 10, 11, 12, \\
&= (\frac{16-s}{8n} - \frac{1}{N})S_0^2 \quad \text{for} \quad s = 13, 14, 15,
\end{aligned}$$

and carrying out the indicated simplifications gives

$$V(\bar{x}_0') = (\frac{1}{n} - \frac{1}{N})S_0^2 + S_0^2(-2Q + 2Q^5 + Q^9 - 2Q^{13} + Q^{17})/4n(1-Q^2). \tag{5.7}$$

On setting $\rho_1 = \rho_2 = 1$ in (5.6), it will reduce to (5.7), thereby verifying its correctness.

4. Numerical results for the efficiency of the simple composite estimator

The variance function (5.6) was explored numerically for nine selected pairs of (ρ_1, ρ_2) values, viz., $(\rho_1, \rho_2) = (0.9, 0.9)$, $(0.9, 0.8)$, $(0.9, 0.7)$, $(0.8, 0.9)$, $(0.8, 0.8)$, $(0.8, 0.7)$, $(0.7, 0.9)$, $(0.7, 0.8)$, $(0.7, 0.7)$, at intervals of 0.1 for Q. The relative efficiences of $\bar{x}_0'$ with respect to $\bar{x}_0$ ignoring the finite population correction, $-S_0^2/N$, are presented in Table 8. Because previous tables have ignored the behavior of the variance function when Q is not close to its optimum value, it was felt desirable to give a complete résumé here so that the

Table 8. Relative efficiency in per cent of $\overline{x}'_0$ with respect to $\overline{x}_0$ when $r = 4$, $m = 8$, $\ell = 2$ for selected (ρ_1, ρ_2) pairs

Ω	(ρ_1, ρ_2)								
	(0.9, 0.9)	(0.9, 0.8)	(0.9, 0.7)	(0.8, 0.9)	(0.8, 0.8)	(0.8, 0.7)	(0.7, 0.9)	(0.7, 0.8)	(0.7, 0.7)
0.2	109.73	109.73	109.73	107.97	107.97	107.97	106.32	106.32	106.32
0.3	115.50	115.50	115.50	111.90	111.90	111.90	108.68	108.68	108.68
0.4	122.14	122.14	122.14	115.47	115.47	115.47	109.88	109.88	109.88
0.5	129.52	129.52	129.52	117.79	117.79	117.79	108.80	108.80	108.80
0.6	136.54	136.54	136.55	116.80	116.81	116.81	103.45	103.46	103.47
0.7	138.76	138.85	138.93	108.03	108.10	108.17	90.55	90.60	90.66
0.8	122.17	122.92	123.68	83.23	83.66	84.11	65.63	65.95	66.27
0.9	63.06	65.46	68.05	37.70	38.73	39.81	28.75	29.43	30.14

reader might gain some concept of the efficiency losses that can occur through a poor choice of Q.

A comparison of the relative efficiencies of $\bar{x}'_0$ with respect to $\bar{x}_0$ at the near optimum Q levels from Tables 1 and 8 proves to be illuminating. It will be recalled that the underlying assumptions of Table 1 are an infinite cycle rotation pattern and a strict Markoff type lag correlogram. $\bar{x}'_0$ is seen to be insensitive to both the differences in sample design and correlation structures between the two tables. For example, when $\rho_1 = 0.9$ and $\rho_2 = 0.7$ the optimum relative efficiency from Table 8 is 138.93% whereas with $\rho = 0.9$, $r = 4$ and $m = 8$ a relative efficiency of 138.59% is recorded in Table 1. This is because $\bar{x}'_0$ is, to all intents and purposes, dependent only upon the most recently acquired sample values as the alternative form (3.2) ably demonstrates. The U.S. Bureau of the Census uses a value of $1/2$ for Q which is somewhat low for an efficient estimate of $\bar{X}_0$ if ρ is large; $Q = 0.7$ is close to optimum for $\rho_1 = 0.9$ according to Table 8.

There may be some question in Table 8 as to why, for fixed ρ_1, the relative efficiency of $\bar{x}'_0$ increases as ρ_2 decreases. Now we earlier concluded that the longer the recurrence time of a rotation group, the more efficient would be the rotation design when a simple composite estimator is used. Thus, a plausible explanation is that an observation behaves "more like" a new observation the smaller is the value of ρ_2 and the approximation to a longer recurrence time is thereby improved.

C. A Generalized Composite Estimator

1. The estimator and its variance

Consideration will now be given to an improved estimator of $\overline{X}_0$ in a C.P.S. design situation. The structure of this estimator is suggested by the assumed correlation model (5.4). The proposed generalized composite estimator is

$$\overset{*}{x}'_0 = \delta \overline{x}'_0 + (1 - \delta)\overline{x}'''_0 , \tag{5.8}$$

where

$$\overline{x}'_0 = Q_1(\overline{x}'_{-1} + \overline{x}_{0,-1} - \overline{x}_{-1,0}) + (1 - Q_1)\overline{x}_0 , \tag{5.9}$$

$$\overline{x}'''_0 = Q_2(\overline{x}'''_{-12} + \overline{x}_{0,-12} - \overline{x}_{-12,0}) + (1 - Q_2)\overline{x}_0 , \tag{5.10}$$

$$0 < Q_1, Q_2, \delta < 1 .$$

$\overset{*}{x}'_0$ is thus a weighted average of two estimators. The first, $\overline{x}'_0$, is the simple composite estimator (5.1) employed in a two cycle, four visit design. The second, $\overline{x}'''_0$, is a simple composite estimator appropriate to a one cycle rotation design with $r = 2$ visits. These two visits are conceptually separated by eleven occasions when no sampling occurs, i.e., a yearly survey. On any occasion the first of the two rotation groups is composed of those four original rotation groups in the first cycle of the two cycle design, the second of the two rotation groups being the other four rotation groups in the second cycle.

From (5.8) it follows that

$$V(\overset{*}{x}'_0) = \delta^2 V(\overline{x}'_0) + (1-\delta)^2 V(\overline{x}'''_0) + 2\delta(1-\delta)\,\mathrm{Cov}(\overline{x}'_0, \overline{x}'''_0) . \tag{5.11}$$

Now $V(\overline{x}'_0)$ is given in general by (5.5) and under the correlation structure (5.4) by (5.6). Let

$$\bar{x}_0''' = \sum_{k=1}^{N} \sum_{a=0}^{-\infty} u_{a,k}\, x_{a,k}, \tag{5.12}$$

with variance

$$V(\bar{x}_0''') = (NE(u_{0,k}^2) - \frac{1}{N}) S_0^2 - N \sum_{a=-1}^{-\infty} E(u_{a,k}^2) S_a^2$$

$$+ N \sum_{\substack{a \neq a' \\ =0}}^{-\infty} E(u_{a,k} u_{a',k}) S_{a,a'} . \tag{5.13}$$

With the two visit, one cycle interpretation in mind, it is readily seen that the system of rotation weight variables $u_{a,k}$ described in Table 9 are appropriate here. $\bar{x}_0'''$ is therefore unbiased and consequently $\overset{*}{x}{}_0'$ is unbiased as well.

Using the fact that $u_{a,k} = 0$ for $a \neq -12s$, $s = 0, 1, 2, \ldots$, (5.13) becomes, with the assistance of Table 9,

$$V(\bar{x}_0''') = (\frac{1}{n} - \frac{1}{N}) S_0^2 + Q_2^2 S_0^2/n + (1+Q_2)^2 \sum_{a=-1}^{-\infty} Q_2^{-2a} S_a^2/n$$

$$- Q_2(1+Q_2)^2 S_{0,-12}/n - Q_2(1+Q_2)^2 \sum_{a=-1}^{-\infty} Q_2^{-2a} S_{12a,12a-12}/n. \tag{5.14}$$

With a stationary covariance structure and the correlogram (5.4), (5.14) becomes

$$V(\bar{x}_0''') = (\frac{1}{n} - \frac{1}{N}) S_0^2 + S_0^2 Q_2(1+Q_2)(2Q_2 - \rho_2(1+Q_2))/n(1-Q_2^2). \tag{5.15}$$

Table 9. Weights $u_{a,k}$ for $\bar{x}_0'''$ with $r = 2$, $m = \infty$, $\ell = 1$

Occasion a	Cycle number	Weight $u_{a,k}$	Number of units
0	1	$(1-Q_2)/n$	$2n_2$
	2	$(1+Q_2)/n$	$2n_2$
	not present	0	$N-4n_2$
-12, -24, -36, ...	1	$-Q_2^{-a/12}(1+Q_2)/n$	$2n_2$
	2	$Q_2^{-a/12}(1+Q_2)/n$	$2n_2$
	not present	0	$N-4n_2$
$\neq 0, -12, -24, \ldots$	all units	0	N

The covariance term in (5.11), $\text{Cov}(\bar{x}_0', \bar{x}_0''')$, will next be evaluated. In terms of the rotation weight variables $u_{a,k}$ and $v_{a,k}$,

$$\text{Cov}(\bar{x}_0', \bar{x}_0''') = \text{Cov}\left(\sum_{a=0}^{-\infty} \sum_{k=1}^{N} v_{a,k}\, x_{a,k},\ \sum_{a=0}^{-\infty} \sum_{k=1}^{N} u_{a,k}\, x_{a,k} \right). \tag{5.16}$$

It may be shown, by a method similar to that employed in deriving the variance function (5.13), that (5.16) can be written as

$$\text{Cov}(\bar{x}_0', \bar{x}_0''') = (NE(u_{0,k}\, v_{0,k}) - \frac{1}{N})S_0^2$$

$$+ N \sum_{\alpha=-1}^{-\infty} E(u_{\alpha,k} v_{\alpha,k}) S_{\alpha}^2 + N \sum_{\substack{\alpha \neq \alpha' \\ =0}}^{-\infty} E(u_{\alpha,k} v_{\alpha',k}) S_{\alpha,\alpha'} . \qquad (5.17)$$

It may be verified from Tables7 and 9 that (5.17) becomes

$$\mathrm{Cov}(\bar{x}_0', \bar{x}_0''') = \left(\frac{1}{n} - \frac{1}{N}\right) S_0^2 + \sum_{t=0}^{2} Q_1^{t+1} (3 + Q_1 - (1-Q_1)t) \left[-2S_{0,-1-t} \right.$$

$$\left. + Q_1^8 (1+Q_2) S_{0,-9-t} - Q_1^{12} (1+Q_2) S_{0,-13-t} \right] /(24n) - Q_2 (1+Q_2) S_{0,-12}/(2n)$$

$$+ \sum_{\ell=1}^{\infty} Q_1^{12\ell} Q_2^{\ell} \sum_{t=0}^{2} Q_1^{9+t} (1+Q_2)(3 + Q_1 - (1-Q_1)t) \left[S_{-12\ell, -12\ell-9-t} \right.$$

$$\left. - Q_1^4 S_{-12\ell, -12\ell-13-t} \right] /(24n) + \sum_{\ell=1}^{\infty} Q_1^{12\ell} Q_2^{\ell} \sum_{t=0}^{2} Q_1^{-9-t} (1+Q_2)(1 + 3Q_1$$

$$+ t(1-Q_1)) \left[-S_{-12\ell, -12\ell+9+t} + Q_1^8 S_{-12\ell-12, -12\ell+t+1} \right] /(24n).$$

$$(5.18)$$

In deriving (5.18) it is observed that $NE(u_{\alpha,k} v_{\alpha,k}) = 0$ for all $\alpha < 0$

and that $N \sum_{\substack{\alpha \neq \alpha' \\ =0}}^{-\infty} E(u_{\alpha,k} v_{\alpha',k}) S_{\alpha,\alpha'} = N \sum_{\alpha=0}^{-\infty} \sum_{t=1}^{\infty} E(u_{\alpha,k} v_{\alpha-t,k}) S_{\alpha,\alpha-t}$

$+ N \sum_{\alpha=0}^{-\infty} \sum_{t=1}^{\infty} E(u_{\alpha-t,k} v_{\alpha,k}) S_{\alpha-t,\alpha}$, where the two terms on the right do

not contribute identical values as might first be anticipated.

Application of the stationary correlation structure and some simplification finally reduces (5.18) to

$$\mathrm{Cov}(\bar{x}_0', \bar{x}_0''') = (\frac{1}{n} - \frac{1}{N})S_0^2 - Q_2\rho_2(1+Q_2)S_0^2/(2n) + Q_1\rho_1 S_0^2(-2+Q_1^8\rho_1^8(1+Q_2)$$

$$- Q_1^{12}\rho_2(1+Q_2))(3+Q_1(1+2\rho_1) + Q_1^2\rho_1(2+\rho_1) + 3Q_1^3\rho_1^2)/(24n)$$

$$+ Q_1Q_2\rho_1(1+Q_2)S_0^2\ [\ Q_1^{20}(\rho_1^8 - Q_1^4\rho_2)(3+Q_1(1+2\rho_1) + Q_1^2\rho_1(2+\rho_1)$$

$$+ 3Q_1^3\rho_1^2) + (\rho_1^8 - Q_1^8 Q_2\rho_2)(3\rho_1^2 + Q_1\rho_1(2+\rho_1) + Q_1^2(1+2\rho_1)$$

$$+ 3Q_1^3)\ ]\ /\{24n(1 - Q_1^{12} Q_2)\}. \qquad (5.19)$$

As a check on (5.19) we evaluate $\mathrm{Cov}(\bar{x}_0', \bar{x}_0''')$ directly when $x_{a,k} = x_{0,k}$ so that $\rho_1 = \rho_2 = 1$. Then

$$\mathrm{Cov}(\bar{x}_0', \bar{x}_0''') = (1-Q_1)(1-Q_2)\ \mathrm{Cov}(\sum_{t=0}^{\infty} Q_1^t\,\bar{x}_{-t}, \sum_{t=0}^{\infty} Q_2^t\,\bar{x}_{-12t})$$

$$= (1-Q_1)(1-Q_2) \quad \sum_{s=0}^{\infty} Q_1^{12s} Q_2^s V(\bar{x}_{-12s}) + \sum_{s=0}^{\infty} Q_1^{12s} Q_2^s \sum_{t=1}^{\infty} Q_1^t \mathrm{Cov}(\bar{x}_{-12s-t},$$

$$\bar{x}_{-12s}) + \sum_{t=1}^{12} Q_1^{12-t} Q_2\ \mathrm{Cov}(\bar{x}_{-12+t}, \bar{x}_{-12})$$

$$+ \sum_{s=2}^{\infty} \sum_{t=1}^{\infty} Q_1^{12s-t} Q_2^{12s}\ \mathrm{Cov}(\bar{x}_{-12s+t}, \bar{x}_{-12s}) \quad , \qquad (5.20)$$

$$= (\frac{1}{n} - \frac{1}{N})S_0^2 - Q_2(1+Q_2)S_0^2/(2n) + Q_1(1+Q_1)(1+Q_1^2)S_0^2(-2+Q_2+Q_2^2 + Q_1^8$$

$$- Q_1^{12} + Q_1^8Q_2 - Q_1^8 Q_2^2 - Q_1^8 Q_2^3 - Q_1^{12} + Q_1^{12} Q_2)/(8n(1-Q_1^{12}Q_2)).$$

$$(5.21)$$

Setting $\rho_1 = \rho_2 = 1$ in (5.19) and simplifying the resulting expression also yields (5.21), thus providing a check.

Therefore, by virtue of (5.11), gathering together (5.6), (5.15) and (5.19) gives the final form

$$V(\hat{x}_0') = (\frac{1}{n} - \frac{1}{N})S_0^2 + \delta^2 S_0^2 \{ 8Q_1^2(3+Q_1) + (2+Q_1^{12}\rho_2)(8Q_1^3 + 7Q_1^2 - 6Q_1$$

$$- 9 - 2Q_1\rho_1(-6Q_1^3 - Q_1^2 + 4Q_1 + 3) - (3Q_1 + 1)(Q_1 + 3)Q_1^2\rho_1^2)Q_1\rho_1$$

$$+ 4(Q_1^2 + 6Q_1 + 1)Q_1^{13}\rho_2 + Q_1^9\rho_1^9(-4Q_1^3 - 15Q_1^2 - 6Q_1 + 9 + 2Q_1\rho_1(-4Q_1^3$$

$$- 5Q_1^2 + 6Q_1 + 3) - Q_1^2\rho_1^2(12Q_1^3 + 5Q_1^2 - 14Q_1 - 3)) \} / (36n(1 - Q_1^2))$$

$$+ (1-\delta)^2 S_0^2 Q_2(1+Q_2)(2Q_2 - \rho_2(1+Q_2))/(n(1 - Q_2^2))$$

$$+ \delta(1-\delta)S_0^2 \{ -12Q_2(1+Q_2)(1 - Q_1^{12}Q_2)\rho_2 + Q_1\rho_1(1 - Q_1^{12}Q_2)(-2$$

$$+ Q_1^8\rho_1^8(1+Q_2) - Q_1^{12}\rho_2(1+Q_2))(3 + Q_1(1+2\rho_1) + Q_1^2\rho_1(2+\rho_1) + 3Q_1^3\rho_1^2)$$

$$+ Q_1Q_2\rho_1(1+Q_2)[Q_1^{20}(\rho_1^8 - Q_1^4\rho_2)(3 + Q_1(1+2\rho_1) + Q_1^2\rho_1(2+\rho_1)$$

$$+ 3Q_1^3\rho_1^2) + (\rho_1^8 - Q_1^8 Q_2\rho_2)(3\rho_1^2 + Q_1\rho_1(2+\rho_1) + Q_1^2(1+2\rho_1)$$

$$+ 3Q_1^3)] \}/(12n(1 - Q_1^{12}Q_2)). \qquad (5.22)$$

2. <u>Numerical results for the efficiency of the generalized composite estimator</u>

The numerical investigation of $V(\bar{x}_0')$ cited earlier in this chapter was, in fact, only a portion of an overall study of the variance function

$V(\overset{*}{x}_0')$. It is evident that upon setting $\delta = 0$ in (5.22), $V(\bar{x}_0''')$ will arise as a special case, whereas setting $\delta = 1$ in (5.22) yields $V(\bar{x}_0')$. The values assumed by ρ_1, ρ_2 and Q_1 have already been specified in Table 8. In addition, Q_2 was assigned the same range of values as those of Q_1 and δ varied from 0.0 to 1.0 at intervals of 0.1. These calculations were made possible by the availability of a high-speed electronic computer.

Those values of Q_1, Q_2 and δ which approximately minimize $V(\overset{*}{x}_0')$ for the nine (ρ_1, ρ_2) pairs are given in Table 10. The resultant relative efficiencies with respect to the sample mean $\bar{x}_0$ are tabulated as well. Included directly below of these approximate optimum sets are those two combinations of parameter values which produce relative efficiencies that most nearly approach the indicated optimum. The relative efficiency of $\overset{*}{x}_0'$ with respect to both $\bar{x}_0'$ and $\bar{x}_0'''$ using the approximate optimum parameter values for each of the three estimators is presented in Table 11. The approximate optimum value of Q_2, i.e., $\overset{*}{Q}_2$, which minimizes $V(\bar{x}_0''')$ is cited as well; those for $\bar{x}_0'$ and $\overset{*}{x}_0'$ are already available in Tables 8 and 10 respectively.

Tables 10 and 11 reveal that appreciable efficiency gains may be achieved through the use of a generalized composite estimator rather than a simple composite estimator under the assumed correlation model. Although the tables virtually speak for themselves, the following observations are of some importance:

(a) The figures reported in Tables 10 and 11 may again only be

Table 10. Relative efficiency in per cent of $\overset{*}{x}'_0$ with respect to $\bar{x}_0$ at approximate optimum Q_1, Q_2, δ values for selected (ρ_1, ρ_2) pairs, $r = 4$, $m = 8$, $\ell = 2$

ρ_1	ρ_2	Q_1	Q_2	δ	Relative efficiency
0.9	0.9	0.8	0.7	0.5	240.12
		0.8	0.7	0.4	239.28
		0.9	0.7	0.3	235.85
0.9	0.8	0.8	0.6	0.5	188.14
		0.8	0.7	0.6	186.70
		0.8	0.6	0.6	186.06
0.9	0.7	0.8	0.5	0.6	166.88
		0.8	0.6	0.6	166.78
		0.8	0.5	0.5	166.03
0.8	0.9	0.8	0.7	0.4	210.45
		0.8	0.7	0.3	209.37
		0.7	0.7	0.4	206.42
0.8	0.8	0.7	0.6	0.5	162.65
		0.8	0.6	0.4	162.54
		0.7	0.6	0.4	160.82
0.8	0.7	0.7	0.5	0.5	144.62
		0.7	0.6	0.6	144.42
		0.7	0.5	0.6	143.11
0.7	0.9	0.8	0.7	0.3	189.09
		0.7	0.7	0.4	188.32
		0.7	0.7	0.3	188.20
0.7	0.8	0.7	0.6	0.4	149.26
		0.6	0.6	0.5	147.90
		0.6	0.6	0.4	147.63
0.7	0.7	0.7	0.5	0.4	132.71
		0.6	0.5	0.5	132.61
		0.6	0.6	0.6	132.25

Table 11. Relative efficiency in per cent of $\hat{x}'_0$ with respect to $\bar{x}'_0$ and $\bar{x}'''_0$ at the respective approximate optimum Q_1, Q_2, δ values for selected (ρ_1, ρ_2) pairs, $r = 4$, $m = 8$, $\ell = 2$

ρ_1	ρ_2	Relative efficiency $\hat{x}'_0$ with respect to $\bar{x}'_0$	Relative efficiency $\hat{x}'_0$ with respect to $\bar{x}'''_0$	$\hat{Q}_2$
0.9	0.9	173.05	153.67	0.6
0.9	0.8	135.51	148.01	0.4
0.9	0.7	120.11	144.71	0.3
0.8	0.9	178.66	134.69	0.6
0.8	0.8	138.08	127.95	0.4
0.8	0.7	122.78	125.41	0.3
0.7	0.9	172.09	121.02	0.6
0.7	0.8	135.84	117.42	0.4
0.7	0.7	120.78	115.08	0.3

regarded as approximate optima, for $V(\hat{x}'_0)$ was investigated for combinations of Q_1, Q_2 and δ at discrete intervals. Further, the stationary assumption for the variances and covariances, and the assumed correlation structure are at best only inexact models of the true conditions in a specific sampling situation.

(b) Since the specification of a two cycle design necessarily demands that the population size N be effectively infinite, the deletion of the finite population corrections from the calculations involved in Tables 10 and 11 is wholly justified.

(c) The efficiency gains in employing $\overset{*}{x}'_0$ rather than $\bar{x}_0$ as an estimator of $\bar{X}_0$ are well worthwhile for all of the nine (ρ_1, ρ_2) combinations explored. From Table 11 significant gains over the simple composite estimators $\bar{x}'_0$ and $\bar{x}'''_0$ are also in evidence. For example, when $\rho_1 = 0.8$ and $\rho_2 = 0.9$, the relative efficiencies of $\overset{*}{x}'_0$ with respect to $\bar{x}_0$, $\bar{x}'_0$ and $\bar{x}'''_0$ are 210.45%, 178.66% and 134.69%.

(d) Thus $\bar{x}'''_0$, which ignores all sampled information but that collected on occasions $-12s$, $s = 0, 1, 2, \ldots$, can be a more efficient estimator of $\bar{X}_0$ than $\bar{x}'_0$. This is because the structure of $\bar{x}'_0$ is such as to weight the data gathered on the prior occasions $-12s$ by factors (approximately equal to) Q_1^{12s}. This is desirable under a strict exponential correlation model but inefficient when high year-to-year correlations exist.

(e) The simultaneous optimum values of Q_1 and Q_2 in $\overset{*}{x}'_0$ are not in agreement with the individual optimum values of Q_1 in $\bar{x}'_0$ and of Q_2 in $\bar{x}'''_0$. This discrepancy is due to the (positive) correlation existing between $\bar{x}'_0$ and $\bar{x}'''_0$, and this correlation varies with the choice of Q_1 and Q_2.

(f) Some latitude in the selection of Q_1, Q_2 and δ is permitted without appreciable efficiency losses from the optimum attainable efficiency resulting.

The reader is justified in asking why the estimator

$$\bar{x}_0'' = Q_1'\bar{x}_0 + Q_2'(\bar{x}_{-1}'' + \bar{x}_{0,-1} - \bar{x}_{-1,0}) + Q_3'(\bar{x}_{-12}'' + \bar{x}_{0,-12} - \bar{x}_{-12,0}) \tag{5.23}$$

where $0 < Q_1', Q_2' < 1$, $Q_1' + Q_2' + Q_3' = 1$, was not considered in preference to $\hat{x}_0'$. It apparently possesses all of the advantages of $\hat{x}_0'$ and is a function of the two parameters Q_2' and Q_3' (for $Q_1' = 1 - Q_2' - Q_3'$) compared with the three of $\hat{x}_0'$. The functional form (5.23) for a composite estimator has been termed a "multi-component estimator" and will be examined in detail in Chapter VI. The discussion there will reveal that to obtain an exact general variance function for (5.23) appears to be an intractable problem. Further results of Chapter VI will lead us to the inference that an optimum choice of Q_1, Q_2 and δ in $\hat{x}_0'$ will give an estimator which is almost as precise as $\bar{x}_0''$ under an optimum choice of Q_1' and Q_2' in $\bar{x}_0''$.

This chapter has indeed emphasized the important role that the choice of estimator as well as the choice of design plays in exploiting a sampling situation to the utmost.

VI. MULTI-COMPONENT ESTIMATION

A. The Estimator and Its Exact Variance

The possibility of improving the composite estimator

$$\bar{x}'_0 = Q(\bar{x}'_{-1} + \bar{x}_{0,-1} - \bar{x}_{-1,0}) + (1-Q)\bar{x}_0 \tag{6.1}$$

by the addition of a third term which explicitly utilizes the estimate of change between the current occasion and a previous occasion $-a < -1$ will be examined. This multi-component estimator is

$$\bar{x}''_0 = Q_1\bar{x}_0 + Q_2(\bar{x}''_{-1} + \bar{x}_{0,-1} - \bar{x}_{-1,0}) + Q_3(\bar{x}''_{-a} + \bar{x}_{0,-a} - \bar{x}_{-a,0}) \tag{6.2}$$

where $Q_1 + Q_2 + Q_3 = 1$. It is, of course necessary that there be a matching of some sample units between the current occasion and occasion $-a$ if such an estimator is even to be considered.

Estimator (6.2) is intuitively appealing in sampling situations where the assumption of an exponential correlation model would be unrealistic. It is indeed possible that the trend of the correlation coefficient of a sampling unit characteristic over time may not be monotonely decreasing due to, perhaps, a seasonality in the characteristic being estimated. For example, consider a survey conducted during June and December, say, of each year to determine employment in the construction trade in the suburban areas of a city. It is evident that the value assumed by a sampling unit twelve months ago yields more pertinent information concerning the value associated with that same unit on the current occasion than would an observation collected six months ago. Thus the number of days a carpenter is unemployed during December of the current year would be suspected to be more highly correlated with the number of days unemployed twelve months ago rather than with the previous June figure. Empirical evidence of such a state of affairs has already been cited in Chapter V. Estimator (6.2) with $a = 2$ would appear to exploit this specific situation.

It is not, however, obvious that a multi-component estimator will produce any appreciable gains in efficiency over a simple composite estimator such as (6.1) since the information supplied by the third component of (6.2) is implicitly contained in the second component. The efficiency will certainly not be less if optimum values of Q_1, Q_2 and Q_3 are used to minimize the variance of (6.2) for the optimization would necessarily select $Q_3 = 0$ rather than permit losses in efficiency.

In order to determine the variance of (6.2) it is necessary to express it first as a linear function of the sample observations. Accordingly, let

$$\bar{x}_t'' = Q_1 \bar{x}_t + Q_2(\bar{x}_{t-1}'' + \bar{x}_{t,t-1} - \bar{x}_{t-1,t}) + Q_3(\bar{x}_{t-a}'' + \bar{x}_{t,t-a} - \bar{x}_{t-a,t}) \tag{6.3}$$

where $a > 1$. Multiplying (6.3) through by $Q_2^t \; Q_3^t$ and putting

$$U_t = Q_2^t \, Q_3^t \, \bar{x}_t'' ,$$

yields

$$U_t = Q_2^2 \, Q_3 \, U_{t-1} + Q_2^a \, Q_3^{a+1} \, U_{t-a} + Q_2^t \, Q_3^t \, (Q_1 \bar{x}_t + Q_2 \, \delta_{t,t-1} + Q_3 \, \delta_{t,t-a}) \tag{6.4}$$

where

$$\delta_{t,t-1} = \bar{x}_{t,t-1} - \bar{x}_{t-1,t},$$

$$\delta_{t,t-a} = \bar{x}_{t,t-a} - \bar{x}_{t-a,t}.$$

Let

$$\phi(t) = Q_2^t Q_3^t (Q_1 \bar{x}_t + Q_2 \delta_{t,t-1} + Q_3 \delta_{t,t-a}). \tag{6.5}$$

Then (6.4) becomes

$$U_t - Q_2^2 Q_3 U_{t-1} - Q_2^a Q_3^{a+1} U_{t-a} = \phi(t). \tag{6.6}$$

This resembles a linear autoregressive scheme of order a. The solution of such a difference equation is discussed, for example, by Kendall (1948). The autoregressive equation

$$X_t + b_1 X_{t-1} + \dots + b_p X_{t-p} = Y_t, \tag{6.7}$$

subject to the conditions that (a) $|Y_t| < \infty$ as $t \to -\infty$ and (b) X_t is defined for $t \to -\infty$, has the solution

$$X_t = \sum_{s=0}^{\infty} g_s Y_{t-s}, \tag{6.8}$$

where

$$g_s = \Sigma\, \xi_1^{s_1} \dots \xi_p^{s_p},$$

the summation being over all $s_1, s_2, \dots, s_p$ subject to $\sum_{i=1}^{p} s_i = s$, and $\xi_1, \dots, \xi_p$ are the roots of the characteristic equation

$$\xi^p + b_1 \xi^{p-1} + \dots + b_{p-1}\xi + b_p = 0,$$

with $|\xi_1|, \dots, |\xi_p| < 1$.

Hence for $a > 2$ we are forced to solve a higher order polynomial equation. Since (6.6) involves unspecified coefficients Q_2 and Q_3 it is not in general possible to give an explicit expression for U_t. The problem is therefore limited to the case where $a = 2$, thereby requiring the solution of only a quadratic equation. With $a = 2$

$$g_s = \xi_1^s + \xi_1^{s-1}\xi_2 + \dots + \xi_1 \xi_2^{s-1} + \xi_2^s$$

$$= (\xi_1^{s+1} - \xi_2^{s+1})/(\xi_1 - \xi_2), \tag{6.9}$$

where ξ_1 and ξ_2 are the roots of $\xi^2 + b_1\xi + b_2 = 0$, providing that $|\xi_1| < 1$, $|\xi_2| < 1$.

Hence, corresponding to the multi-component estimator

$$\bar{x}_t'' = Q_1 \bar{x}_t + Q_2(\bar{x}_{t-1}'' + \delta_{t,t-1}) + Q_3(\bar{x}_{t-2}'' + \delta_{t,t-2}), \tag{6.10}$$

there arises the difference equation

$$U_t - Q_2^2 Q_3 U_{t-1} - Q_2^2 Q_3^3 U_{t-2} = \emptyset(t). \tag{6.11}$$

The roots of the associated characteristic equation

$$\xi^2 - Q_2^2 Q_3 \xi - Q_2^2 Q_3^3 = 0$$

are

$$\xi_1 = (Q_2^2 Q_3 + Q_2 Q_3\sqrt{Q_2^2 + 4Q_3})/2 = Q_2 Q_3(Q_2 + B)/2, \tag{6.12}$$

$$\xi_2 = (Q_2^2 Q_3 - Q_2 Q_3 \sqrt{Q_2^2 + 4Q_3})/2 = Q_2 Q_3 (Q_2 - B)/2, \tag{6.13}$$

where
$$B = \sqrt{Q_2^2 + 4Q_3}\ . \tag{6.14}$$

It is easily verified that $|\xi_1| < 1$, $|\xi_2| < 1$ in all cases, for $Q_1 + Q_2 + Q_3 = 1$ implies that $Q_2 Q_3 \leq 1/4$ (an obvious extension of the fact that $PQ \leq 1/4$ in simple binomial sampling) and therefore that $Q_2^2 Q_3 \leq 1/4$. An upper bound for ξ_1 is therefore found by setting $Q_2 Q_3 = Q_2^2 Q_3 = 1/4$, $Q_2 = Q_3 = 1$ in (6.12) and this yields $\xi_1 < 0.404 < 1.0$. Since ξ_1 is non-negative, therefore $|\xi_1| < 1$. Similarly an upper bound for ξ_2 is determined by setting $Q_2 = Q_3 = 0$ in (6.13), giving $\xi_2 < 1/8$. A lower bound is given by putting $Q_2^2 Q_3 = 0$ and $Q_2^2 = Q_3 = 1$ in (6.13), so that $\xi_2 > -5^{1/2}/8 > -1$. Hence $|\xi_2| < 1$. The solution of (6.11) is thus

$$U_t = \sum_{a=0}^{-\infty} \left[\frac{(Q_2 Q_3 (Q_2 + B)/2)^{-a+1} - (Q_2 Q_3 (Q_2 - B)/2)^{-a+1}}{Q_2 Q_3 B} \right] \phi(t+a), \tag{6.15}$$

from (6.8) and (6.9).

By putting $t = 0$ in (6.15), it follows that

$$\bar{x}_0'' = Q_1 \bar{x}_0 + Q_2(\bar{x}_{-1}'' + \bar{x}_{0,-1} - \bar{x}_{-1,0}) + Q_3(\bar{x}_{-2}'' + \bar{x}_{0,-2} - \bar{x}_{-2,0}) = U_0$$

$$= \frac{1}{B} \sum_{a=0}^{-\infty} \left[\left(\frac{Q_2 + B}{2}\right)^{-a+1} - \left(\frac{Q_2 - B}{2}\right)^{-a+1} \right] (Q_1 \bar{x}_a + Q_2 \delta_{a,a-1} + Q_3 \delta_{a,a-2}) \tag{6.16}$$

It is assumed for purposes of variance determination that the survey was effectively initiated at $\alpha = -\infty$.

Following the procedure developed for the simple composite estimator $\bar{x}'_0$ in Chapter III, (6.16) is now expressed as the sum of products of weight variables $w_{\alpha,k}$, $\alpha = 0, -1, -2, \ldots$; $k = 1, 2, \ldots, N$, and the values $x_{\alpha,k}$ assumed by each unit in the population irrespective of whether or not it is sampled on occasion α. If the k-th unit is not in the sample on occasion α, then $w_{\alpha,k} = 0$. Thus (6.16) may be written in the alternative form

$$\bar{x}''_0 = \sum_{\alpha=0}^{-\infty} \sum_{k=1}^{N} w_{\alpha,k}\, x_{\alpha,k}\,. \tag{6.17}$$

Consider the one cycle rotation design $(m = \infty)$ of Chapter III wherein a sampling unit remains in the sample for $r \geq 2$ consecutive occasions. It then drops out of the sample and does not return. The sample size on any occasion is $n = rn_2$, n_2 units being rotated out of the sample to be replaced by n_2 units on their first visit.

By expressing (6.16) explicitly in terms of the $x_{\alpha,k}$ as in (6.17) it is seen that for occasions $\alpha \leq -2$ the following five sets of weights are appropriate, where, for brevity

$$\frac{Q_2 + B}{2} = C, \quad \frac{Q_2 - B}{2} = D. \tag{6.18}$$

(a) Units in for a first visit at α:

$$w_{\alpha,k} = \frac{1}{B}\left(C^{-\alpha-1}\left(\frac{-Q_3}{n-2n_2} - \frac{CQ_2}{n-n_2} + \frac{C^2 Q_1}{n}\right)\right.$$

$$- D^{-a-1}\left(\frac{-Q_3}{n-2n_2} - \frac{DQ_2}{n-n_2} + \frac{D^2Q_1}{n}\right)) .$$

(b) Units in for a second visit at a:

$$w_{a,k} = \frac{1}{B}\left[C^{-a-1}\left(\frac{-Q_3}{n-2n_2} - \frac{CQ_2}{n-n_2} + C^2\left(\frac{Q_1}{n} + \frac{Q_2}{n-n_2}\right)\right) - D^{-a-1}\left(\frac{-Q_3}{n-2n_2}\right.\right.$$

$$\left.\left. - \frac{DQ_2}{n-n_2} + D^2\left(\frac{Q_1}{n} + \frac{Q_2}{n-n_2}\right)\right)\right] .$$

(c) Units in for an $(r-1)$-th visit at a:

$$w_{a,k} = \frac{1}{B}\left[C^{-a}\left(\frac{-Q_2}{n-n_2} + C\left(\frac{Q_1}{n} + \frac{Q_2}{n-n_2} + \frac{Q_3}{n-2n_2}\right)\right) - D^{-a}\left(\frac{-Q_2}{n-n_2}\right.\right.$$

$$\left.\left. + D\left(\frac{Q_1}{n} + \frac{Q_2}{n-n_2} + \frac{Q_3}{n-2n_2}\right)\right)\right] .$$

(d) Units in for an r-th visit at a:

$$w_{a,k} = \frac{(C+D)^{-a+1}}{B}\left(\frac{Q_1}{n} + \frac{Q_2}{n-n_2} + \frac{Q_3}{n-2n_2}\right) .$$

(e) Units in for a 3rd, 4th, ..., $(r-3)$rd, $(r-2)$nd visit at a:

$$w_{a,k} = \frac{1}{B}\left[C^{a-1}\left(\frac{-Q_3}{n-2n_2} - \frac{CQ_2}{n-n_2} + C^2\left(\frac{Q_1}{n} + \frac{Q_2}{n-n_2} + \frac{Q_3}{n-2n_2}\right)\right)\right.$$

$$\left. - D^{a-1}\left(\frac{-Q_3}{n-2n_2} - \frac{DQ_2}{n-n_2} + D^2\left(\frac{Q_1}{n} + \frac{Q_2}{n-n_2} + \frac{Q_3}{n-2n_2}\right)\right)\right] .$$

When $a = 0$ the corresponding weights are as follows.

(f) Units in for a first visit at $\alpha = 0$:

$$w_{0,k} = \frac{Q_1}{n} .$$

(g) Units in for a second visit at $\alpha = 0$:

$$w_{0,k} = \frac{Q_1}{n} + \frac{Q_2}{n-n_2} .$$

(h) Units in for a 3rd, 4th, ..., (r-1)-th, r-th visit at $\alpha = 0$:

$$w_{0,k} = \frac{Q_1}{n} + \frac{Q_2}{n-n_2} + \frac{Q_3}{n-2n_2} .$$

When $\alpha = -1$ the appropriate weights are as follows.

(i) Units in for a first visit at $\alpha = -1$:

$$w_{-1,k} = \frac{Q_1 Q_2}{n} - \frac{Q_2}{n-n_2} .$$

(j) Units in for a second visit at $\alpha = -1$:

$$w_{-1,k} = \frac{Q_1 Q_2}{n} + \frac{Q_2^2}{n-n_2} - \frac{Q_2}{n-n_2} .$$

(k) Units in for 3rd, 4th, ..., (r-2)nd, (r-1)-th visit at $\alpha = -1$:

$$w_{-1,k} = \frac{Q_1 Q_2}{n} + \frac{Q_2^2}{n-n_2} + \frac{Q_2 Q_3}{n-2n_2} - \frac{Q_2}{n-n_2} .$$

(l) Units in for r-th visit at $\alpha = -1$:

$$w_{-1,k} = \frac{Q_1 Q_2}{n} + \frac{Q_2^2}{n-n_2} + \frac{Q_2 Q_3}{n-2n_2} .$$

The complexity of the above system of weight variables when $\alpha \leq -2$ leads us to consider a further simplification of the problem in order to obtain an explicit expression for the variance, $V(\bar{x}_0'')$, of the estimator $\bar{x}_0''$. Accordingly let $r = 3$ so that a unit remains in the sample for three consecutive occasions before dropping out forever. Table 12 exhibits the entire system of weights $w_{\alpha, k}$ appropriate to this situation.

It may be verified that

$$NE(w_{0,k}) = 1, \quad NE(w_{\alpha,k}) = 0 \text{ for } \alpha < 0,$$

so that $\bar{x}_0''$ is an unbiased estimator of $\bar{X}_0$.

The variance of $\bar{x}_0''$ is given by the general formula developed in Chapter III, viz.,

$$V(\bar{x}_0'') = (NE(w_{0,k}^2) - \frac{1}{N})S_0^2 + N \sum_{\alpha=-1}^{-\infty} E(w_{\alpha,k}^2)S_\alpha^2$$

$$+ N \sum_{\substack{\alpha \neq \alpha' \\ =0}}^{-\infty} E(w_{\alpha,k} w_{\alpha',k})S_{\alpha,\alpha'}, \qquad (6.19)$$

where S_0^2, S_α^2, $S_{\alpha,\alpha'}$ are defined as in (3.11).

Table 12 of weights is employed in evaluating the terms of (6.19). The algebraic manipulations involved are extremely tedious. Some of the intermediate calculations are presented here to assist the reader who might have reason to follow the algebra himself.

$$NE(w_{0,k}^2)S_0^2 = (2 + Q_2^2 + 4Q_3^2 + 2Q_2 Q_3)S_0^2/6n_2, \qquad (6.20)$$

Table 12. Weights $w_{\alpha, k}$ for the multi-component estimator with $r = 3$

Occasion α	Visit number	Weight $w_{\alpha, k}$
0	1	$(1 - Q_2 - Q_3)/3n_2$
	2	$(2 + Q_2 - 2Q_3)/6n_2$
	3	$(2 + Q_2 + 4Q_3)/6n_2$
-1	1	$Q_2(-1 - 2Q_2 - 2Q_3)/6n_2$
	2	$Q_2(-1 + Q_2 - 2Q_3)/6n_2$
	3	$Q_2(2 + Q_2 + 4Q_3)/6n_2$
≤ -2	1	$(C^{-\alpha-1}(-6Q_3 - 3CQ_2 + 2C^2 Q_1) - D^{-\alpha-1}(-6Q_3 - 3DQ_2 + 2D^2 Q_1))/6Bn_2$
	2	$(C^{-\alpha}(-3Q_2 + 2CQ_1 + 3CQ_2) - D^{-\alpha}(-3Q_2 + 2DQ_1 + 3DQ_2))/6Bn_2$
	3	$(C^{-\alpha+1} - D^{-\alpha+1})(2 + Q_2 + 4Q_3)/6Bn_2$

$$NE(w^2_{-1,k})S^2_{-1} = Q_2^2(1 + Q_2^2 + 4Q_3^2 + Q_2 + 4Q_3 + 2Q_2 Q_3)S^2_{-1}/6n_2. \qquad (6.21)$$

In evaluating the second and third terms on the right of (6.19) the assumption of constant variance $S^2_\alpha = S^2_0$, $\alpha = -1, -2, \ldots,$ is made. Then

$$\sum_{\alpha=-2}^{-\infty} NE(w^2_{\alpha,k})S^2_\alpha = S^2_0 \sum_{\alpha=-2}^{-\infty} \Big[(C^{-\alpha+1} - D^{-\alpha+1})^2 (2 + Q_2 + 4Q_3)^2$$

$$+ (C^{-\alpha}(-3Q_2 + 2CQ_1 + 3CQ_2) - D^{-\alpha}(-3Q_2 + 2DQ_1 + 3DQ_2))^2$$

$$+ (C^{-\alpha-1}(-6Q_3 - 3CQ_2 + 2C^2 Q_1) - D^{-\alpha-1}(-6Q_3 - 3DQ_2 + 2D^2 Q_1))^2 \Big]/36B^2 n_2$$

$$(6.22)$$

The following relationships were of use in simplifying this component.

$$C + D = Q_2, \quad C^2 + D^2 = Q_2^2 + 2Q_3, \quad CD = -Q_3,$$

$$C^3 + D^3 = Q_2(Q_2^2 + 3Q_3), \quad C^4 + D^4 = Q_2^4 + 4Q_2^2 Q_3 + 2Q_3^2,$$

$$C^5 + D^5 = Q_2^5 + 5Q_2^3 Q_3 + 5Q_2 Q_3^2, \quad C^6 + D^6 = Q_2^6 + 6Q_2^4 Q_3 + 9Q_2^2 Q_3^2 + 2Q_3^3,$$

$$(1 - C^2)(1 - D^2) = 1 - Q_2^2 - 2Q_3 + Q_3^2, \quad \sum_{\alpha=-2}^{-\infty} C^{-2\alpha-2} = C^2/(1 - C^2),$$

$$\sum_{\alpha=-2}^{-\infty} D^{-2\alpha-2} = D^2/(1 - D^2), \quad \sum_{\alpha=-2}^{-\infty} (-Q_3)^{-\alpha-1} = Q_3/(1 + Q_3). \qquad (6.23)$$

The final result is

$$\sum_{\alpha=-2}^{-\infty} NE(w^2_{\alpha,k})S^2_\alpha = S^2_0 \Big[Q_2(16Q_3^3 - 8Q_3^4 - 8Q_3^5) + Q_2^2(12Q_3^2 + 40Q_3^3$$

$$+ 40Q_3^4 - 4Q_3^5 - 16Q_3^6) + Q_2^3(24Q_3^2 + 18Q_3^3 + 2Q_3^4 - 8Q_3^5) + Q_2^4(6Q_3 + 38Q_3^2$$

$$+ 46Q_3^3 + 12Q_3^4 - 4Q_3^5) + Q_2^5(9Q_3 + 17Q_3^2 + 9Q_3^3 - 2Q_3^4) + Q_2^6(1 + 11Q_3 + 13Q_3^2$$

$$+ 3Q_3^3) + Q_2^7(1 + 3Q_3 + 2Q_3^2) + Q_2^8(1 + Q_3) + 16Q_3^3 - 16Q_3^6 \Big] / 6B^2 n_2(1 + Q_3) \cdot$$

$$(1 - Q_2^2 - 2Q_3 + Q_3^2) . \tag{6.24}$$

Adding (6.24) to (6.20) and (6.21) gives

$$\sum_{a=0}^{-\infty} NE(w_{a,k}^2) S_a^2 = S_0^2/3n_2 + S_0^2(Q_2(8Q_3^2 + 8Q_3^3 - 16Q_3^4) + Q_2^2(8Q_3 + 24Q_3^2$$

$$+ 12Q_3^3 - 4Q_3^4) + Q_2^3(6Q_3 + 18Q_3^2 - 4Q_3^3) + Q_2^4(2 + 4Q_3 + 4Q_3^2) + Q_2^5(1 + 4Q_3)$$

$$+ 32Q_3^3 - 16Q_3^4 - 16Q_3^5)/6n_2(Q_2^2 + 4Q_3)(1 + Q_3)(1 - Q_2^2 - 2Q_3 + Q_3^2) . \tag{6.25}$$

Since $r = 3$ it follows that

$$N \sum_{\substack{a \neq a' \\ = 0}}^{-\infty} E(w_{a,k} w_{a',k}) S_{a,a'} = 2N \sum_{a=0}^{-\infty} E(w_{a,k} w_{a-1,k}) S_{a,a-1}$$

$$+ 2N \sum_{a=0}^{-\infty} E(w_{a,k} w_{a-2,k}) S_{a,a-2} .$$

Let

$$S_{a,a-1} = \rho_1 S_0^2 , \quad S_{a,a-2} = \rho_2 S_0^2 . \tag{6.26}$$

Then

$$2NE(w_{0,k} w_{-1,k}) S_{0,-1} = 2\rho_1 S_0^2 (Q_2(-4 - 10Q_3 - 4Q_3^2)$$

$$+ Q_2^2(-4+4Q_3) - Q_2^3)/36n_2 . \qquad (6.27)$$

$$2NE(w_{-1,k}w_{-2,k})S_{-1,-2} = 2\rho_1 S_0^2(Q_2(8Q_3+14Q_3^2-4Q_3^3) + Q_2^2(2Q_3+4Q_3^2)$$

$$+ Q_2^3(-1-5Q_3-4Q_3^2) + Q_2^4(2+4Q_3) - Q_2^5)/36n_2, \qquad (6.28)$$

where the relationships

$$C - D = B, \quad C^2 - D^2 = BQ_2, \quad C^3 - D^3 = B(Q_2^2+Q_3), \qquad (6.29)$$

were used. Further

$$2N\sum_{a=-2}^{-\infty} E(w_{a,k}w_{a-1,k})S_{a,a-1} = 2\rho_1 S_0^2 \sum_{a=-2}^{-\infty}\Big[(C^{-a+1} - D^{-a+1})(2+Q_2$$

$$+ 4Q_3)(C^{-a+1}(-3Q_2+2CQ_1+3CQ_2) - D^{-a+1}(-3Q_2+2DQ_1+3DQ_2))$$

$$+(C^{-a}(-3Q_2+2CQ_1+3CQ_2) - D^{-a}(-3Q_2+2DQ_1+3DQ_2))\ (C^{-a}(-6Q_3-3CQ_2$$

$$+ 2C^2Q_1) - D^{-a}(-6Q_3-3DQ_2+2D^2Q_1))\Big]\ /36^2Bn_2. \qquad (6.30)$$

Using the relationships in (6.23) and the additional result

$$C^7 + D^7 = Q_2^7 + 7Q_2^5Q_3 + 14Q_2^3Q_3^2 + 7Q_2Q_3^3, \qquad (6.31)$$

(6.30) reduces to

$$2N\sum_{a=-2}^{-\infty} E(w_{a,k}w_{a-1,k})S_{a,a-1} = 2\rho_1 S_0^2\ \Big[\ Q_2(-32Q_3^3+40Q_3^4+32Q_3^5$$

$$-56Q_3^6+16Q_3^7) + Q_2^2(-32Q_3^3+56Q_3^4-8Q_3^5-16Q_3^6) + Q_2^3(4Q_3^2+30Q_3^3+4Q_3^4$$

$$- 26Q_3^5 + 20Q_3^6) + Q_2^4(-8Q_3^2 + 78Q_3^3 + 22Q_3^4 - 20Q_3^5) + Q_2^5(-Q_3 - 27Q_3^2 - 41Q_3^3$$

$$- 15Q_3^4 + 4Q_3^5) + Q_2^6(8Q_3 + 40Q_3^2 + 22Q_3^3 - 4Q_3^4) + Q_2^7(-1 - 12Q_3 - 14Q_3^2 - 3Q_3^3)$$

$$+ Q_2^8(2 + 6Q_3 + 4Q_3^2) + Q_2^9(-1 - Q_3)\Big] / 36B^2 n_2(1 + Q_3)(1 - Q_2^2 - 2Q_3 + Q_3^2) . \tag{6.32}$$

Adding (6.32) to (6.27) and (6.28) finally results in

$$2N \sum_{\alpha=0}^{-\infty} E(w_{\alpha,k} w_{\alpha-1,k}) S_{\alpha,\alpha-1} = 2\rho_1 S_0^2 \Big[\ Q_2(-16Q_3 + 8Q_3^2 + 32Q_3^3 - 24Q_3^4)$$

$$+ Q_2^2(-16Q_3 + 40Q_3^2 - 24Q_3^3) + Q_2^3(-4 + 10Q_3 + 32Q_3^2 - 6Q_3^3) + Q_2^4(-4 + 34Q_3$$

$$- 6Q_3^2) + Q_2^5(2 + 6Q_3) + 6Q_2^6\Big] / 36n_2(1 + Q_3)(Q_2^2 + 4Q_3)(1 - Q_2^2 - 2Q_3 + Q_3^2) . \tag{6.33}$$

Again,

$$2NE(w_{o,k} w_{-2,k}) S_{0,-2} = 2\rho_2 S_0^2(2 + Q_2 + 4Q_3)\ (C(-6Q_3 - 3CQ_2 + 2C^2 Q_1)$$

$$- D(-6Q_3 - 3DQ_2 + 2D^2 Q_1))/36Bn_2,$$

$$= 2\rho_2 S_0^2(2 + Q_2 + 4Q_3)(-2Q_2 Q_3 - Q_2^2(1 + 2Q_3) - 2Q_2^3 - 4Q_3 - 2Q_3^2)/36Bn_2 . \tag{6.34}$$

$$2NE(w_{-1,k} w_{-3,k}) S_{-1,-3} = 2\rho_2 S_0^2(2 + Q_2 + 4Q_3) Q_2\ (C^2(-6Q_3 - 3CQ_2$$

$$+ 2C^2 Q_1) - D^2(-6Q_3 - 3DQ_2 + 2D^2 Q_1))/36Bn_2,$$

$$= 2\rho_2 S_0^2(2 + Q_2 + 4Q_3)(Q_2^2(-5Q_3 - 4Q_3^2) - 4Q_2^2 Q_3 - Q_2^4(1 + 2Q_3) - 2Q_2^5)/36Bn_2, \tag{6.35}$$

where, in addition to (6.29), the relationship

$$C^4 - D^4 = B(Q_2^3 + 2Q_2Q_3) \tag{6.36}$$

was employed. Also

$$2N \sum_{a=-2}^{-\infty} E(w_{a,k}w_{a-2,k})S_{a,a-2} = 2\rho_2 S_0^2 \sum_{a=-2}^{-\infty} (C^{-a+1} - D^{-a+1})(2+Q_2$$

$$+ 4Q_3)\{C^{-a+1}(-6Q_3 - 3CQ_2 + 2C^2Q_1) - D^{-a+1}(-6Q_3 - 3DQ_2 + 2D^2Q_1)\}$$

$$/36B^2 n_2 .$$

Using the relationships in (6.23) and (6.31) and also

$$C^8 + D^8 = Q_2^8 + 8Q_2^6Q_3 + 20Q_2^4Q_3^2 + 16Q_2^2Q_3^3 + 2Q_3^4, \tag{6.37}$$

we obtain

$$2N \sum_{a=-2}^{-\infty} E(w_{a,k}w_{a-2,k})\rho_2 S_0^2 = 2\rho_2 S_0^2(2+Q_2+4Q_3)(Q_2(-8Q_3^4+8Q_3^5)$$

$$+ Q_2^2(-44Q_3^3 - 46Q_3^4 + 14Q_3^5 + 16Q_3^6) + Q_2^3(-34Q_3^3 - 6Q_3^4 + 16Q_3^5) + Q_2^4(-38Q_3^2$$

$$- 68Q_3^3 - 17Q_3^4 + 12Q_3^5) + Q_2^5(-40Q_3^2 - 26Q_3^3 + 12Q_3^4) + Q_2^6(-11Q_3 - 26Q_3^2$$

$$- 13Q_3^3 + 2Q_3^4) + Q_2^7(-16Q_3 - 14Q_3^2 + 2Q_3^3) + Q_2^8(-1 - 3Q_3 - 2Q_3^2) + Q_2^9(-2$$

$$- 2Q_3) - 16Q_3^4 + 8Q_3^5 + 8Q_3^6)/36n_2(1+Q_3)(Q_2^2 + 4Q_3)(1 - Q_2^2 - 2Q_3 + Q_3^2) .$$

$$\tag{6.38}$$

Adding (6.38) to (6.34) and (6.35) yields

$$2N \sum_{a=0}^{-\infty} E(w_{a,k}w_{a-2,k})S_{a,a-2} = 2\rho_2 S_0^2(2+Q_2+4Q_3)\ (Q_2(-8Q_3^2+8Q_3^3)$$

$$+Q_2^2(-8Q_3-6Q_3^2+2Q_3^3)+Q_2^3(-10Q_3+2Q_3^2)-Q_2^4(1+2Q_3)-2Q_2^5-16Q_3^2$$

$$+8Q_3^3+8Q_3^4)/36n_2(1+Q_3)(Q_2^2+4Q_3)(1-Q_2^2-2Q_3+Q_3^2). \tag{6.39}$$

Therefore, by virtue of (6.19), collecting (6.25), (6.33) and (6.39) gives finally the variance formula

$$V(\bar{x}_0'') = (\frac{1}{n}-\frac{1}{N})S_0^2 + S_0^2\Big[\ 24Q_2(Q_3^2+Q_3^3-2Q_3^4)+12Q_2^2(2Q_3+6Q_3^2+3Q_3^3$$

$$-Q_3^4)+6Q_2^3(3Q_3+9Q_3^2-2Q_3^3)+6Q_2^4(1+2Q_3+2Q_3^2)+3Q_2^5(1+4Q_3)+48(2Q_3^3$$

$$-Q_3^4-Q_3^5)+\rho_1(8Q_2(-2Q_3+Q_3^2+4Q_3^3-3Q_3^4)+8Q_2^2(-2Q_3+5Q_3^2-3Q_3^3)$$

$$+2Q_2^3(-2+5Q_3+16Q_3^2-3Q_3^3)+2Q_2^4(-2+17Q_3-3Q_3^2)+2Q_2^5(1+3Q_3)+6Q_2^6)$$

$$+(2+Q_2+4Q_3)\rho_2(8Q_2(-Q_3^2+Q_3^3)+2Q_2^2(-4Q_3-3Q_3^2+Q_3^3)+2Q_2^3(-5Q_3$$

$$+Q_3^2)+Q_2^4(-1-2Q_3)-2Q_2^5+8(-2Q_3^2+Q_3^3+Q_3^4))\Big]\ /6n(1+Q_3)(Q_2^2+4Q_3)\cdot$$

$$(1-Q_2^2-2Q_3+Q_3^2). \tag{6.40}$$

Because of the heavy algebraic calculations involved in arriving at (6.40), an independent check on its correctness was made. As a special case the populations on all occasions are assumed to be identical, so that $\rho_1 = \rho_2 = 1$, and $\delta_{a,a-1} = \delta_{a,a-2} = 0$. The solution of the corresponding difference equation analogous to (6.4) is then

$$\bar{x}_0'' = Q_1 \sum_{s=0}^{\infty}\Big[\ (\frac{Q_2+B}{2})^{s+1} - (\frac{Q_2-B}{2})^{s+1}\ \Big]\ \bar{x}_{-s}/B. \tag{6.41}$$

Directly from (6.41) we have that

$$V(\bar{x}_0'') = \frac{Q_1^2}{B^2} \sum_{s=0}^{\infty} \left[\left(\frac{Q_2+B}{2}\right)^{s+1} - \left(\frac{Q_2-B}{2}\right)^{s+1} \right]^2 V(\bar{x}_{-s})$$

$$+ \frac{2Q_1^2}{B^2} \sum_{s=0}^{\infty} \left[\left(\frac{Q_2+B}{2}\right)^{s+1} - \left(\frac{Q_2-B}{2}\right)^{s+1} \right] \left[\left(\frac{Q_2+B}{2}\right)^{s+2} - \left(\frac{Q_2-B}{2}\right)^{s+2} \right] \mathrm{Cov}(\bar{x}_{-s}, \bar{x}_{-s-1})$$

$$+ \frac{2Q_1^2}{B^2} \sum_{s=0}^{\infty} \left[\left(\frac{Q_2+B}{2}\right)^{s+1} - \left(\frac{Q_2-B}{2}\right)^{s+1} \right] \left[\left(\frac{Q_2+B}{2}\right)^{s+3} - \left(\frac{Q_2-B}{2}\right)^{s+3} \right] \mathrm{Cov}(\bar{x}_{-s}, \bar{x}_{-s-2})$$

$$+ \frac{2Q_1^2}{B^2} \sum_{s=0}^{\infty} \left[\left(\frac{Q_2+B}{2}\right)^{s+1} - \left(\frac{Q_2-B}{2}\right)^{s+1} \right] \sum_{t=3}^{\infty} \left[\left(\frac{Q_2+B}{2}\right)^{s+1+t} - \left(\frac{Q_2-B}{2}\right)^{s+1+t} \right] \mathrm{Cov}(\bar{x}_{-s}, \bar{x}_{-s-t}). \qquad (6.42)$$

Now

$$V(\bar{x}_{-s}) = \left(\frac{1}{3n_2} - \frac{1}{N}\right) S_0^2, \quad \mathrm{Cov}(\bar{x}_{-s}, \bar{x}_{-s-1}) = \left(\frac{2}{9n_2} - \frac{1}{N}\right) S_0^2,$$

$$\mathrm{Cov}(\bar{x}_{-s, -s-2}) = \left(\frac{1}{9n_2} - \frac{1}{N}\right) S_0^2, \quad \mathrm{Cov}(\bar{x}_{-s}, \bar{x}_{-s-t}) = -\frac{S_0^2}{N}, \quad t = 3, 4, 5, \ldots . \qquad (6.43)$$

Substitution of the values in (6.43) into (6.42) and considerable simplification gives

$$V(\bar{x}_0'') = (\frac{1}{n} - \frac{1}{N})S_0^2 + S_0^2 \Big[Q_2(-8Q_3 + 24Q_3^3 - 16Q_3^4) + Q_2^2(-4Q_3 + 30Q_3^2 - 2Q_3^4) + Q_2^3(-2 + 22Q_3^2 - 4Q_3^3) + 14Q_2^4 Q_3 + 4Q_2^5 Q_3 + 2Q_2^6 - 16Q_3^2 + 24Q_3^3 - 8Q_3^5 \Big] / 3n(1+Q_3)(Q_2^2 + 4Q_3)(1 - Q_2^2 - 2Q_3 + Q_3^2). \qquad (6.44)$$

It will be seen that on setting $\rho_1 = \rho_2 = 1$, (6.40) reduces to (6.44), thereby providing at least a partial check on the accuracy of the calculations.

As a second partial check we may set $Q_3 = 0$ in $\bar{x}_0''$, thus reducing it to $\bar{x}_0'$. Equation (6.40) then becomes

$$V(\bar{x}_0'') = V(\bar{x}_0') = (\frac{1}{n} - \frac{1}{N})S_0^2 + S_0^2 \Big[Q_2^2 \frac{(2+Q_2)}{2} + \rho_1 Q_2 \frac{(-2 - 2Q_2 + Q_2^2 + 3Q_2^3)}{3} - \rho_2 \frac{Q_2^2(2 + 5Q_2 + 2Q_2^2)}{6} \Big] / n(1 - Q_2^2). \qquad (6.45)$$

The substitution of $r = 3$, $m = \infty$, $S_{0,-1} = \rho_1 S_0^2$, $S_{0,-2} = \rho_2 S_0^2$ into the general formula (3.13) produces the expression (6.45), thus providing the check.

A second simple composite estimator is obtained from (6.16) by setting $Q_2 = 0$,

$$\bar{x}_0''' = Q_1 \bar{x}_0 + Q_3(\bar{x}_{-2}''' + \bar{x}_{0,-2} - \bar{x}_{-2,0}). \qquad (6.46)$$

It explicitly ignores the estimate of change available from the units common to occasions 0 and -1, and is intuitively more appealing than

$\bar{x}_0'$ when $\rho_2 > \rho_1$. Putting $Q_2 = 0$ in (6.40) gives

$$V(\bar{x}_0''') = (\frac{1}{n} - \frac{1}{N})S_0^2 + S_0^2\{6(2Q_3^3 - Q_3^4 - Q_3^5) + \rho_2(2 + 4Q_3)(-2Q_3^2 + Q_3^3 + Q_3^4)\}/3nQ_3(1 + Q_3)(1 - 2Q_3 + Q_3^2) . \qquad (6.47)$$

B. Numerical Results

A numerical investigation of the variance function $V(\bar{x}_0'')$ given by (6.40) for various designated values of the coefficients Q_1, Q_2 and Q_3 and coefficients of correlation ρ_1 and ρ_2 was undertaken. Both Q_1 and Q_2 were permitted to vary from 0.0 to 0.9 at intervals of 0.1 with the restriction that $Q_2 + Q_3 \leq 0.9$; it is evident that one would never in practice set $Q_1 = 0.0$ and thereby completely ignore the information obtained from unmatched units on the current occasion. The correlation coefficients ρ_1 and ρ_2 were assigned the following nine selected pairs of values: (0.9, 0.9), (0.9, 0.8), (0.9, 0.7), (0.8, 0.9), (0.8, 0.8), (0.8, 0.7), (0.7, 0.9), (0.7, 0.8), (0.7, 0.7).

Since both S_0^2 and the sample size are unspecified the relative efficiency, given by the variance of the simple current occasion mean expressed as a per cent of the variance of $\bar{x}_0''$, was considered, i.e.,

$$R.E_{\bar{x}_0''} = \frac{V(\bar{x}_0)}{V(\bar{x}_0'')}(100)\% .$$

The finite population correction term, $-S_0^2/N$, was ignored in both numerator and denominator since N is necessarily large with respect to n. Similar definitions and comments for the relative efficiency of

$\bar{x}_0'$ over $\bar{x}_0$ and of $\bar{x}_0'''$ over $\bar{x}_0$ also apply.

Tables 13 to 21 present the results of the calculations for the nine pairs of (ρ_1, ρ_2) values specified above. They are somewhat abbreviated here because values of Q_3 greater than 0.6 did not give results of any particular interest. As an illustration of the use of the tables, from Table 14 the maximum gain in efficiency of $\bar{x}_0''$ over $\bar{x}_0$ with $\rho_1 = 0.9$ and $\rho_2 = 0.8$ is scored when $Q_2 = 0.5$, $Q_3 = 0.1$ and hence $Q_1 = 0.4$, the gain being 47.5%.

Table 22 summarizes the approximate optimum values $\hat{Q}_2$ and $\hat{Q}_3$ of Q_2 and Q_3 which give the maximum increase in efficiency of $\bar{x}_0''$ over $\bar{x}_0$ for each of the nine assumed correlation structures. Also tabulated are the approximate optimum Q_2 and Q_3 values using $\bar{x}_0'$ and $\bar{x}_0'''$ respectively. Finally the last two columns present the relative efficiency of $\bar{x}_0''$ with respect to $\bar{x}_0'$ and $\bar{x}_0'''$ using the optimum Q values cited for each correlation situation.

Table 13. Relative efficiency in per cent of $\overline{x}_0''$ with respect to $\overline{x}_0$ with $\rho_1 = 0.9$, $\rho_2 = 0.9$

Q_2 \ Q_3	0.0	0.1	0.2	0.3	0.4	0.5	0.6
0.0	100.0	112.4	125.3	137.6	147.1	149.9	141.2
0.1	106.3	120.9	136.2	150.6	161.0	161.8	146.4
0.2	113.5	130.5	148.1	163.7	171.8	163.1	129.4
0.3	122.1	141.6	161.0	174.8	173.3	142.2	78.9
0.4	132.5	154.3	173.0	178.3	152.2	85.8	
0.5	144.5	167.2	179.2	159.5	92.4		
0.6	158.2	176.4	164.7	98.9			
0.7	169.8	167.5	105.5				
0.8	167.2	111.9					
0.9	117.9						

Table 14. Relative efficiency in per cent of $\overline{x}_0''$ with respect to $\overline{x}_0$ with $\rho_1 = 0.9$, $\rho_2 = 0.8$

Q_2 \ Q_3	0.0	0.1	0.2	0.3	0.4	0.5	0.6
0.0	100.0	110.3	118.9	123.8	122.4	112.5	93.9
0.1	106.3	118.2	127.9	132.6	129.2	115.2	91.2
0.2	113.3	126.6	136.4	138.9	129.7	107.1	74.0
0.3	121.2	135.1	143.3	139.7	119.5	84.0	53.1
0.4	130.0	143.1	145.6	129.4	92.9	45.5	
0.5	139.1	147.5	137.0	101.3	49.7		
0.6	145.8	142.0	109.2	54.2			
0.7	144.5	116.6	59.1				
0.8	122.9	64.4					
0.9	70.2						

Table 15. Relative efficiency in per cent of $\bar{x}_0''$ with respect to $\bar{x}_0$ with $\rho_1 = 0.9$, $\rho_2 = 0.7$

Q_2 \ Q_3	0.0	0.1	0.2	0.3	0.4	0.5	0.6
0.0	100.0	108.2	113.1	112.5	104.8	90.0	70.3
0.1	106.3	115.5	120.5	118.3	107.8	89.4	66.3
0.2	113.0	122.7	126.3	120.5	104.2	79.7	51.8
0.3	120.3	129.2	129.0	116.1	91.2	59.6	33.0
0.4	127.7	133.5	125.5	101.4	66.9	31.0	
0.5	133.9	131.9	110.9	74.2	34.0		
0.6	135.3	119.0	81.6	37.3			
0.7	125.6	89.4	41.0				
0.8	97.1	45.2					
0.9	50.0						

Table 16. Relative efficiency in per cent of $\bar{x}_0''$ with respect to $\bar{x}_0$ with $\rho_1 = 0.8$, $\rho_2 = 0.9$

Q_2 \ Q_3	0.0	0.1	0.2	0.3	0.4	0.5	0.6
0.0	100.0	112.4	125.3	137.6	147.1	149.9	141.2
0.1	105.5	119.8	134.8	148.8	159.0	159.7	144.7
0.2	111.5	127.9	144.5	159.5	166.8	158.7	126.6
0.3	118.2	136.4	154.3	166.9	165.6	137.2	77.3
0.4	125.9	145.6	162.3	166.7	143.9	83.1	
0.5	134.4	153.8	163.9	147.5	88.3		
0.6	142.7	157.5	148.6	92.9			
0.7	148.1	146.8	97.1				
0.8	142.0	100.4					
0.9	102.6						

Table 17. Relative efficiency in per cent of $\bar{x}_0''$ with respect to $\bar{x}_0$ with $\rho_1 = 0.8$, $\rho_2 = 0.8$

Ω_2 \ Ω_3	0.0	0.1	0.2	0.3	0.4	0.5	0.6
0.0	100.0	110.3	118.9	123.8	122.4	112.5	93.9
0.1	105.5	117.1	126.6	131.2	127.9	114.0	90.6
0.2	111.2	123.9	133.3	135.7	126.9	105.2	73.1
0.3	117.4	130.4	137.9	134.6	115.7	82.2	40.9
0.4	123.8	135.7	137.7	123.2	89.8	44.8	
0.5	129.5	137.0	127.9	96.3	48.5		
0.6	132.6	129.7	101.8	52.4			
0.7	128.5	106.2	56.3				
0.8	108.8	60.4					
0.9	64.4						

Table 18. Relative efficiency in per cent of $\bar{x}_0''$ with respect to $\bar{x}_0$ with $\rho_1 = 0.8$, $\rho_2 = 0.7$

Ω_2 \ Ω_3	0.0	0.1	0.2	0.3	0.4	0.5	0.6
0.0	100.0	108.2	113.1	112.5	104.8	90.0	70.3
0.1	105.4	114.5	119.3	117.4	106.8	88.7	66.0
0.2	111.0	120.3	123.8	118.2	102.4	78.6	51.4
0.3	116.6	125.0	124.7	112.6	89.0	58.7	27.8
0.4	121.7	126.9	119.6	97.7	65.2	30.6	
0.5	125.0	123.5	104.8	71.5	33.5		
0.6	123.8	110.3	77.5	36.4			
0.7	113.4	83.1	39.7				
0.8	88.1	43.2					
0.9	47.0						

Table 19. Relative efficiency in per cent of $\bar{x}_0''$ with respect to $\bar{x}_0$ with $\rho_1 = 0.7$, $\rho_2 = 0.9$

Q_2 \ Q_3	0.0	0.1	0.2	0.3	0.4	0.5	0.6
0.0	100.0	112.4	125.3	137.6	147.1	149.9	141.2
0.1	104.7	118.8	133.3	147.3	157.0	157.7	143.1
0.2	109.5	125.2	141.2	155.3	162.3	154.8	123.9
0.3	114.7	131.6	148.1	160.0	158.5	132.3	75.7
0.4	120.0	137.7	152.7	156.7	136.4	80.6	
0.5	125.5	142.5	151.3	137.2	84.5		
0.6	130.0	142.5	135.3	87.6			
0.7	131.4	130.7	89.8				
0.8	123.6	91.0					
0.9	90.7						

Table 20. Relative efficiency in per cent of $\bar{x}_0''$ with respect to $\bar{x}_0$ with $\rho_1 = 0.7$, $\rho_2 = 0.8$

Q_2 \ Q_3	0.0	0.1	0.2	0.3	0.4	0.5	0.6
0.0	100.0	110.3	118.9	123.8	122.4	112.5	93.9
0.1	104.6	116.0	125.3	129.9	126.6	113.0	89.9
0.2	109.3	121.5	130.4	132.6	124.2	103.3	72.2
0.3	113.9	126.1	133.0	129.9	112.2	80.4	40.5
0.4	118.1	128.9	130.9	117.6	86.8	44.0	
0.5	121.2	127.9	119.9	91.7	47.4		
0.6	121.5	119.3	95.4	50.6			
0.7	115.6	88.0	53.8				
0.8	97.6	56.9					
0.9	59.6						

Table 21. Relative efficiency in per cent of $\bar{x}_0''$ with respect to $\bar{x}_0$ with $\rho_1 = 0.7$, $\rho_2 = 0.7$

Q_2 \ Q_3	0.0	0.1	0.2	0.3	0.4	0.5	0.6
0.0	100.0	108.2	113.1	112.5	104.8	90.0	70.3
0.1	104.6	113.5	118.2	116.3	105.9	88.0	65.6
0.2	109.1	117.9	121.2	115.9	100.6	77.6	61.8
0.3	113.1	120.9	120.6	109.3	86.9	57.7	27.6
0.4	116.1	121.1	114.4	94.2	63.7	30.3	
0.5	117.2	116.0	99.4	69.0	32.9		
0.6	114.2	102.6	73.7	35.6			
0.7	103.3	71.6	38.4				
0.8	80.6	41.3					
0.9	44.3						

Table 22. Optimum choices of weight coefficients for $\bar{x}_0', \bar{x}_0'', \bar{x}_0'''$ and relative efficiences of $\bar{x}_0''$ with respect to $\bar{x}_0'$ and $\bar{x}_0'''$ in per cent

		$\bar{x}_0''$		$\bar{x}_0'$	$\bar{x}_0'''$	Relative efficiency of $\bar{x}_0''$ with respect to	
ρ_1	ρ_2	$\hat{Q}_2$	$\hat{Q}_3$	$\hat{Q}_2$	$\hat{Q}_3$	$\bar{x}_0'$	$\bar{x}_0'''$
0.9	0.9	0.5	0.2	0.7	0.5	105.4	119.4
0.9	0.8	0.5	0.1	0.6	0.3	101.2	119.2
0.9	0.7	0.6	0.0	0.6	0.2	100.0	119.5
0.8	0.9	0.3	0.3	0.7	0.5	112.7	111.4
0.8	0.8	0.3	0.2	0.6	0.3	104.0	111.5
0.8	0.7	0.4	0.1	0.5	0.2	101.6	112.2
0.7	0.9	0.2	0.4	0.7	0.5	123.5	108.3
0.7	0.8	0.3	0.2	0.6	0.3	108.7	106.7
0.7	0.7	0.2	0.2	0.5	0.2	103.4	107.1

C. Discussion of the Numerical Results

The most important conclusion to be derived from the foregoing tables is the fact that moderate efficiency gains may be achieved through the use of a multi-component estimator rather than a simple composite estimator when $\rho_2 \geq \rho_1$ in a three visit one cycle rotation design. It is evident that $\bar{x}_0''$ is at least as efficient as $\bar{x}_0$, $\bar{x}_0'$ and $\bar{x}_0'''$ under an optimum choice of the Q weight coefficients. From Table 22 it is noted that $\bar{x}_0'''$ is more efficient than $\bar{x}_0'$ when $\rho_1 < \rho_2$ and less efficient when $\rho_2 \geq \rho_1$. The optimum Q values and the corresponding relative efficiencies are only approximate since the interval of tabulation was 0.1 which is somewhat large. Thus the relative efficiency of 100.0% quoted in the case $\rho_1 = 0.9$, $\rho_2 = 0.7$ would no doubt be slightly greater since this is the minimum attainable value. The optimum values quoted are, however, rather robust for no appreciable efficiency differences occur in the neighbourhood of these values. The efficiency falls off very rapidly as the deviation from optimum values becomes more pronounced. This gives greater confidence in the conclusions drawn from Tables 13 to 22. In practice the true values of the correlation coefficients would not be exactly known so that the robustness property of the optimum Q's is indeed desirable. As would be expected, the optimum value of Q_3 in the estimator $\bar{x}_0''$ decreases as ρ_2 decreases for a given ρ_1. Similarly the optimum value of Q_3 in the estimator $\bar{x}_0'''$ decreases as ρ_2 decreases, this estimator's variance being independent of ρ_1.

One would suspect that more spectacular efficiency gains might be scored with a multi-component estimator in a rotation design with more than three visits. Even though $\bar{x}_0''$ explicitly takes advantage of the fact that $\rho_2 \geq \rho_1$, there is only a one-third sample overlap between occasions 0 and -2 in the particular design situation investigated. The results of Chapter III have indicated that a fifty per cent overlap is close to optimum for $\bar{x}_0'$, and therefore a somewhat larger match between the more highly correlated occasions is indicated.

A multi-component estimator of the change of level between the previous and current occasions, given by

$$d_0'' = \bar{x}_0'' - \bar{x}_{-1}'' , \tag{6.48}$$

should also be a successful alternative to $d_0' = \bar{x}_0' - \bar{x}_{-1}'$ when $\rho_2 \geq \rho_1$. No attempt was made at deriving the variance of (6.48) however.

D. An Alternative Generalized Composite Estimator

By solving a second order difference equation it was possible, as shown in A of this chapter, to readily express the estimator $\bar{x}_0''$ in terms of the observations $x_{\alpha,k}$ and to therefore obtain $V(\bar{x}_0'')$. It was noted that with more general designs the solution of higher order difference equations with unspecified coefficients is required. The method proposed is therefore lacking somewhat in generality. The form of the estimator $\bar{x}_0''$ can be modified to obtain an alternative estimator $\hat{x}_0'$ which will still exploit the correlation structure between units separated by more than one time interval. The variance of $\hat{x}_0'$ is more easily obtained but the simplification is at the expense of some loss in precision over the use of $\bar{x}_0''$. The alternative generalized composite estimator $\hat{x}_0'$ is of the same structure as the improved C. P. S. - type generalized composite estimator (5. 8). It is the purpose of this section to both derive the variance of $\hat{x}_0'$ and to make some evaluation of the efficiency loss in using this estimator rather than $\bar{x}_0''$. This comparison has implications concerning the improved C. P. S. - type estimator as well, as already mentioned in Chapter V.

The alternative estimator is

$$\hat{x}_0' = \delta \bar{x}_0' + (1 - \delta)\bar{x}_0''' , \tag{6.49}$$

where
$$\bar{x}_0' = Q_1(\bar{x}_{-1}' + \bar{x}_{0,-1} - \bar{x}_{-1,0}) + (1 - Q_1)\bar{x}_0 , \tag{6.50}$$

$$\bar{x}_0''' = Q_2(\bar{x}_{-2}''' + \bar{x}_{0,-2} - \bar{x}_{-2,0}) + (1 - Q_2)\bar{x}_0 , \tag{6.51}$$

$$0 < Q_1, Q_2, \delta < 1 .$$

The sample design assumed is a three visit one cycle pattern with correlations ρ_1 and ρ_2 for units separated by one and two time intervals respectively. The derivation of $V(\hat{x}_0')$ is in the same spirit as that of (5. 8) and is therefore abbreviated here. Now

$$V(\hat{x}_0') = \delta^2 V(\bar{x}_0') + (1 - \delta)^2 V(\bar{x}_0''') + 2\,\delta(1 - \delta)\,\mathrm{Cov}(\bar{x}_0', \bar{x}_0''') . \tag{6.52}$$

By setting

$$\bar{x}_0' = \sum_{a=0}^{-\infty} \sum_{k=1}^{N} w_{a,k}\, x_{a,k} , \qquad \bar{x}_0''' = \sum_{a=0}^{-\infty} \sum_{k=1}^{N} v_{a,k}\, x_{a,k} ,$$

it will be seen that the weights $w_{a,k}$ and $v_{a,k}$ are as given in Table 23.

Table 23. Weights $w_{a,k}$ and $v_{a,k}$ for the estimators $\bar{x}'_0$ and $\bar{x}'''_0$.

Estimator	Occasion	Visit number	Weight
$\bar{x}'_0$	0	1	$(1 - Q_1)/n$
		2, 3	$(2 + Q_1)/2n$
	$a < 0$	1	$- Q_1^{-a}(1 + 2Q_1)/2n$
		2	$- Q_1^{-a}(1 - Q_1)/2n$
		3	$Q_1^{-a}(2 + Q_1)/2n$
$\bar{x}'''_0$	0	1, 2	$(1 - Q_2)/n$
		3	$(1 + 2Q_2)/n$
	$2a$ $(a = -1, -2, \ldots)$	1	$- Q_2^{-a}(2 + Q_2)/n$
		2	$Q_2^{-a}(1 - Q_2)/n$
		3	$Q_2^{-a}(1 + 2Q_2)/n$
	$2a + 1$ $(a = -1, -2, \ldots)$	1, 2, 3	0

It is then a straightforward exercise to derive

$$V(\bar{x}'_0) = (\frac{1}{n} - \frac{1}{N})S_0^2 + Q_1 S_0^2(3Q_1(2+Q_1) - 2\rho_1(2+2Q_1 - Q_1^2 - 3Q_1^3)$$

$$- \rho_2 Q_1(1+2Q_1)(2+Q_1))/6n(1 - Q_1^2), \qquad (6.53)$$

$$V(\bar{x}_0''') = (\frac{1}{n} - \frac{1}{N})S_0^2 + 2Q_2(2+Q_2)S_0^2(3Q_2 - \rho_2(1+2Q_2))/3n(1-Q_2^2). \qquad (6.54)$$

$Cov(\bar{x}_0', \bar{x}_0''')$ is evaluated with the aid of (5.17). The final result is

$$Cov(\bar{x}_0', \bar{x}_0''') = (\frac{1}{n} - \frac{1}{N})S_0^2 + S_0^2 \; [\, 3Q_1 Q_2(1+Q_1+2Q_1^2+2Q_1 Q_2) + \rho_1(-2Q_1 - Q_1^2 + 3Q_1 Q_2(1+Q_1) - Q_1 Q_2^2(1+2Q_1)) - \rho_2(Q_2(2+Q_1)(2+Q_2) + Q_1^2(1+2Q_1)(1+2Q_2))\,] \; /6n(1-Q_1^2 Q_2). \qquad (6.55)$$

The usual checks involving a direct evaluation when $x_{a,k} = x_{0,k}$ were made on (6.53), (6.54) and (6.55) but are not reproduced here. Hence, by virtue of (6.52),

$$V(\hat{x}_0') \doteq (\frac{1}{n} - \frac{1}{N})S_0^2 + \delta^2 Q_1 S_0^2(3Q_1(2+Q_1) - 2\rho_1(2+2Q_1 - Q_1^2 - 3Q_1^3) - \rho_2 Q_1(1+2Q_1)(2+Q_1))/6n(1-Q_1^2) + 2(1-\delta)^2 Q_2(2+Q_2)S_0^2(3Q_2 - \rho_2(1+2Q_2))/3n(1-Q_2^2) + \delta(1-\delta)S_0^2 \; [\, 3Q_1 Q_2(1+Q_1+2Q_1^2+2Q_1 Q_2) + \rho_1(-2Q_1 - Q_1^2 + 3Q_1 Q_2(1+Q_1) - Q_1 Q_2^2(1+2Q_1)) - \rho_2(Q_2(2+Q_1)(2+Q_2) + Q_1^2(1+2Q_1)(1+2Q_2))\,] \; /3n(1-Q_1^2 Q_2). \qquad (6.56)$$

In order to reduce the calculating burden with (6.56) the approximate optimum values of Q_1 and Q_2 which minimize $V(\bar{x}_0')$ and $V(\bar{x}_0''')$ were taken from Table 22. It appeared to be a more or less reasonable assumption that these would also be the approximate optimum values in (6.56) since (6.55) is considerably less than either (6.54) or (6.53). A numerical study was then undertaken to determine the approximate optimum δ values for the nine pairs of ρ_1 and ρ_2 values. The results of this study are summarized in Table 24, where the approximate optimum values $\hat{Q}_1$, $\hat{Q}_2$ and $\hat{\delta}$ are displayed together with the relative efficiency of $\hat{x}_0'$ with respect to $\bar{x}_0$ as an estimator of $\bar{X}_0$. The approximate optimum relative efficiency of $\bar{x}_0'''$ with respect to $\bar{x}_0$ is also recorded to facilitate a comparison.

It is evident that although $\hat{x}_0'$ is less efficient than $\bar{x}_0''$, the loss incurred is small. Hence $V(\hat{x}_0')$ would be expected to serve as a good conservative approximation to $V(\bar{x}_0'')$. That $\bar{x}_0''$ is recorded as being less efficient than $\hat{x}_0'$ when $\rho_1 = 0.9$ and $\rho_2 = 0.7$ is due to the fact that the optimum weights $\hat{Q}_1$, $\hat{Q}_2$ and $\hat{\delta}$ are only approximate. In this particular instance the optimum Q_3 was estimated to be $\hat{Q}_3 = 0$ for $\bar{x}_0''$ which undoubtedly yields an ultraconservative estimate of the relative efficiency of $\bar{x}_0''$. Although no indication is given in Table 24, $V(\hat{x}_0')$ was observed to be stable about the approximate optimum δ value.

The foregoing study led us to deduce that the generalized composite estimator (5.8) would generate estimates almost as precise as those of the multi-component estimator (5.23) under an optimum choice of parameter values in the C.P.S. design situation. In practice one would prefer (6.2) to (6.49) both because of the slight precision gain and because the former requires a choice of two parameter values versus the three selections for the latter. Whilst the variance function of (6.49) is known and that of (6.2) is not known in more complex rotation designs, the possibility of short-cut methods of variance estimation in multi-stage designs obviates this advantage of a generalized composite estimator over a multi-component estimator.

Table 24. Relative efficiency in per cent of $\hat{x}_0'$ and $\bar{x}_0''$ with respect to $\bar{x}_0$ at the approximate optimum Q_1, Q_2, δ values for selected (ρ_1, ρ_2) combinations

ρ_1	ρ_2	$\hat{Q}_1$	$\hat{Q}_2$	$\hat{\delta}$	Relative efficiency of $\hat{x}_0'$	Relative efficiency of $\bar{x}_0''$
0.9	0.9	0.7	0.5	0.7	173.7	179.2
0.9	0.8	0.6	0.3	0.9	145.9	147.5
0.9	0.7	0.6	0.2	0.9	135.8	135.3
0.8	0.9	0.7	0.5	0.5	163.5	166.9
0.8	0.8	0.6	0.3	0.6	136.5	137.9
0.8	0.7	0.5	0.2	0.8	125.6	126.9
0.7	0.9	0.7	0.5	0.3	158.2	162.3
0.7	0.8	0.6	0.3	0.5	131.4	133.0
0.7	0.7	0.5	0.2	0.6	120.7	121.2

VII. SUMMARY AND RECOMMENDATIONS

It is apparent that composite estimation techniques used in conjunction with rotation sample designs can be an extremely powerful method for extracting information over time from a dynamic population. Such considerations as the increased precision of estimates, reduction in costs, control of nonsampling errors, and administrative convenience among others are attractive to the sampling statistician, the economist, the social scientist or indeed anyone who has reason to make inferences concerning some changing population. Attention has primarily been devoted in this dissertation to the first of these factors.

The criterion for selecting the appropriate sample design and estimator for any one characteristic has been that combination which gives minimum variance in the estimate compared with selected alternative designs and estimators. In practice the optimae so decided upon would be modified in light of the other considerations mentioned above. Other circumstances such as the collection of information on a wide variety of characters will often conflict with one another. The statistician is then forced to make some compromise which will satisfactorily meet all needs.

A unified approach to the problem of sampling on successive occasions with a fixed rotation design in a finite population has been developed here. This was accomplished by considering the finite population to be comprised of the $N!$ possible rotation patterns that can be constructed in the population of size N. The population is assumed fixed from occasion to occasion with no units either immigrating or migrating. The sample then consists of one rotation pattern selected at random from this population. The infinite cycle rotation pattern is formulated by permitting a rotation group of n_2 units to remain in the sample for $r \geq 2$ occasions, to withdraw for the next m occasions, to return for another r consecutive occasions, and so on without limit. There is no difficulty if the rotation pattern is superimposed on a subpopulation of N^* units randomly selected without replacement from the N population units.

In Chapter III the simple composite estimator

$$\bar{x}'_0 = Q(\bar{x}'_{-1} + \bar{x}_{0,-1} - \bar{x}_{-1,0}) + (1 - Q)\bar{x}_0 \qquad (7.1)$$

of the current occasion population mean, $\bar{X}_0$, is considered. If sampling was instituted on occasion $-u < 0$, then $\bar{x}'_0$ may be written as

$$\bar{x}'_0 = \sum_{\alpha=0}^{-u} \sum_{k=1}^{N} w_{\alpha,k}\, x_{\alpha,k} . \qquad (7.2)$$

For purposes of variance evaluation it is assumed that u is large so that $\bar{x}'_0$ is essentially given by

$$\bar{x}'_0 = \sum_{\alpha=0}^{-\infty} \sum_{k=1}^{N} w_{\alpha,k}\, x_{\alpha,k} , \qquad (7.3)$$

where $x_{\alpha,k}$ is the value associated with the k-th unit on the α-th occasion and the $w_{\alpha,k}$ are rotation weight variables. The exact variance of (7.3) in the finite population of size N is then developed in general and under the assumption of a stationary covariance structure with an exponential correlation model. The simple composite estimator of the change in level between the previous and current occasions is

$$d_0' = \bar{x}_0' - \bar{x}_{-1}' . \qquad (7.4)$$

The variance of this estimator is derived as well. A generalized variance criterion is suggested as a possible solution to the problem of selecting a single Q which is best for (7.1) and (7.4) simultaneously. A brief section on variance estimation is presented as well.

The theory of sampling on successive occasions is extended in Chapter IV to include two-stage designs where either primary or secondary sampling units are rotated. Composite ratio estimators are introduced and their application in estimating the mean per secondary sampling unit in a two-stage sample design is illustrated.

In Chapter V mention is made of the ℓ cycle rotation design wherein a rotation group performs a total of ℓ cycles of r visits each before dropping out of the sample forever. The Current Population Survey conducted monthly by the U.S. Bureau of the Census employs a two cycle rotation design. A rotation group remains in the sample for four consecutive months, drops out for the next eight months, returns for another four months and then withdraws forever. Under the correlation model

$$\rho_{ij} = \rho_1^i \, \rho_2^j , \qquad (i = 1, 2, \ldots, 11; \; j = 1, 2, 3, \ldots,)$$

which exhibits possibly large year-to-year correlations, the variance of the composite estimator $\bar{x}_0'$ is derived. An improved "generalized composite estimator"

$$\overset{*}{x}{}'_0 = \delta \bar{x}'_0 + (1 - \delta)\bar{x}'''_0 , \tag{7.5}$$

where

$$\bar{x}'''_0 = Q_2(\bar{x}'''_{-12} + \bar{x}_{0,-12} - \bar{x}_{-12,0}) + (1 - Q_2)\bar{x}_0 , \tag{7.6}$$

which explicitly exploits any high year-to-year correlations, is introduced and its variance is given.

A "multi-component estimator" of the structure

$$\bar{x}''_0 = Q_1 \bar{x}_0 + Q_2(\bar{x}''_{-1} + \bar{x}_{0,-1} - \bar{x}_{-1,0}) + Q_3(\bar{x}''_{-2} + \bar{x}_{0,-2} - \bar{x}_{-2,0})$$

is applied in the special case of a three visit one cycle rotation design in Chapter VI. The correlations ρ_1 and ρ_2 between measurements on the same unit one and two occasions apart are unspecified. The solution of a second order difference equation is involved in the process of expressing $\bar{x}''_0$ in a form such as (7.3) in order to arrive finally at $V(\bar{x}''_0)$. The generalized composite estimator analogous to (7.5) is compared in efficiency with $\bar{x}'_0$. Inference is then made concerning the increase in efficiency that might be anticipated in employing a multi-component estimator rather than $\overset{*}{x}{}'_0$ in the important C.P.S. sampling situation.

Due to the complexity of the variance functions arising from the different designs and estimators, recourse was made to a numerical exploration of these relationships in order to determine approximate optimum values of the design and estimator parameters. Some of the more significant conclusions reached were:

(a) The optimum rotation pattern for the estimation of the current occasion mean with a simple composite estimator is a two visit infinite cycle design. As anticipated, the optimum value of Q in $\bar{x}'_0$ is proportional to the correlation ρ existing between consecutive observations on the same units. Moderate efficiency gains over $\bar{x}_0$ are resultant under a choice of Q which is at least near to the optimum value.

(b) The optimum design corresponding to the composite estimator of change, $\bar{x}'_0 - \bar{x}'_{-1}$, would be a fixed panel when an efficiency comparison with the difference of sample means obtained from independent sample draws on occasions 0 and -1 is made. But when $\bar{x}'_0 - \bar{x}'_{-1}$ is paralleled with the difference in arithmetic means as calculated from the same rotation design, a two visit design again becomes superior. Important efficiency gains are registered in either case when ρ is not small. The optimum values of the weight coefficients Q are larger for the composite estimator of change than for the composite estimator of level. The generalized variance criterion for simultaneous optimization in the latter comparison yields a value of Q midway between the individual optimae.

(c) A simple composite estimator employed in a monthly rotation design does not utilize to any appreciable extent the high year-to-year correlations that are frequently encountered in surveys of an economic nature. In such cases a generalized composite estimator and consequently a multi-component estimator could introduce important efficiency gains in resultant estimates.

The foregoing list is by no means exhaustive; a more extensive coverage has already been presented in the main text.

There is room for a considerable amount of additional research in the area of rotation sampling. The variance formulas appropriate to (a) the estimated average level over several occasions, $\sum_{\alpha=0}^{-a} x'_{\alpha, k}/(a+1)$ and to (b) the change in level $\bar{x}'_0 - \bar{x}'_{-a}$ between non-successive occasions, have not been developed. The methodology of this dissertation could be extended to cover these two cases although a great deal of algebraic simplification would be entailed in arriving at the final results. The behavior of the different variance functions under correlation models other than the exponential and arithmetic would be worthwhile looking into for the purpose of studying robustness properties. One such possibility would arise by considering a population composed of a mixture of two groups of units; the first group exhibiting a fast exponential decay and the second a slow exponential decay. A mixture of exponential models of the type $\rho_{\alpha, \alpha+t} = c_1\rho_1^t + c_2\rho_2^t$ might then be expected to hold in the population as a whole.

The generalized variance criterion has some undesirable properties and may well be felt to not be fully satisfactory as a solution to the problem of estimating both level and change or other parameters jointly. The utilization of non-linear programming techniques may yield more convincing results. The possibility of introducing improved composite estimators following the lead of Chapters V and VI should be pursued in earnest for the rewards can indeed be great as we have already seen. Nor have we examined the reduction in variance that might be effected through the introduction of either generalized composite estimators or multi-component estimators of population parameters other than level.

Multi-level rotation sampling was outside the scope of this study. It is possible that the rotation weight variable technique may be applied in this area as well; Eckler (1955) has already shown the relationship between one-level and two-level rotation sampling.

The important factor of cost was ignored in the treatment of rotation designs and composite estimators; it was thus implicitly assumed that no difference in cost was attached to the sampling of new versus matched units on any occasion. Reference has already been made to the fact that the cost of matched units in the sampling of human populations tends to be smaller than that of new units, Cochran (1963). The situation is reversed in, e.g., forest inventory surveys, Ware and Cunia (1962). One would suspect that the $r = 2$ optimum design would be altered with the introduction of cost functions exhibiting differential costs.

Little, if any, research has been conducted on the problem of sampling on successive occasions from a moving population. It was assumed that no units moved into or out of the population P although the value of a character associated with the units did vary over time. In many types of surveys, e.g., area samples, this is a realistic assumption. Some populations are characterized, however, by significant growth or decline in numbers. Examples are the number of homes in the fringe areas of a city versus the membership in the coal miners union over the past fifteen years. Two papers that could possibly be somewhat relevant are by Das (1951) and Seal (1962). Das describes a two-phase sampling procedure which can be applied to a population P having some subpopulation P* whose frame is unknown and for which an estimate of the P* mean is required. A sample of size n is selected from P and it is observed that n_1 of these belong to P*. In the second phase a sample of predetermined size r_1 is selected from the n_1 units if $r_1 < n_1$; if $r_1 \geq n_1$ then all n_1 units are investigated. Thus a sample of size $r = r_1$ or $r = n_1$ has been drawn from P*. Then X, the population total of P*, is unbiasedly estimated by

$\hat{X} = Nn_1 \sum_{i=1}^{r} z_i/rn$ where N is the size of the original population and the z_i are the observed sample values. Seal discusses the problem of stratified sampling from a moving population where the sample frame is not up-to-date. A sample drawn at time $t < 0$ would therefore produce biased estimates of a population total Y_0 at time $t = 0$. A simple birth and death process is assumed for the number $N(t)$ of population units actually in existence at time t. It is further assumed that an additional sample frame referring to an earlier (or later) time point is also available. On the basis of these two frames the unknown parameters of the stochastic process are estimated. Such a probabilistic approach might well lead to a satisfactory solution of the moving population dilemma. It remains, however, to be explored.

VIII. LITERATURE CITED

Ajgaonkar, S. G. B. and B. D. Tikkiwal. 1961. On the theory of univariate successive sampling. (Abstract) Annals of Mathematical Statistics, 32: 340.

Bershad, M. A. and H. Nisselson. 1962. Sampling for a weekly time series. Proceedings of the Social Statistics Section, American Statistical Association 1962: 207-213.

Bose, C. 1943. Note on the sampling error in the method of double sampling. Sankhyā, 6: 329-330.

Cochran, W. G. 1939. The use of the analysis of variance in enumeration by sampling. Journal of the American Statistical Association, 34: 492-510.

________. 1953. Sampling techniques. First edition. New York, New York, John Wiley and Sons, Incorporated.

________. 1963. Sampling techniques. Second edition. New York, New York, John Wiley and Sons, Incorporated.

Das, A. C. 1951. On two-phase and sampling with varying probabilities. International Statistical Conferences, 33: 105-112.

Eckler, A. R. 1955. Rotation sampling. Annals of Mathematical Statistics, 26: 664-685.

Gray, P. G. and T. Cortlett. 1950. Sampling for the Social Survey. Journal of the Royal Statistical Society, Series A, 113: 150-206.

Hansen, M. H., W. H. Hurwitz, and W. G. Madow. 1953. Sample survey methods and theory. Volumes 1 and 2. New York, New York, John Wiley and Sons, Incorporated.

________, ________, H. Nisselson and J. Steinberg. 1955. The redesign of the census current population survey. Journal of the American Statistical Association, 50: 701-719.

Hartley, H. O. 1959. Theory of advanced design of surveys. (Mimeographed lecture notes.) Ames, Iowa, Department of Statistics, Iowa State University of Science and Technology.

________. 1961. Analytic studies of sample surveys. (Mimeographed report.) Ames, Iowa, Department of Statistics, Iowa State University of Science and Technology.

Jessen, R. J. 1942. Statistical investigation of a sample survey for obtaining farm facts. Iowa Agricultural Experiment Station Research Bulletin 304.

Kailin, H. 1955. Description of the sample for the monthly retail trade report (Revised). Washington, D.C., United States Department of Commerce.

Kendall, M. G. 1948. The advanced theory of statistics. Second edition. Volume 2. London, England, Charles Griffin and Company, Incorporated.

Mahalanobis, P. C. 1952. Some aspects of the design of sample surveys. Sankhyā, 12: 1-7.

Maki, W. R. and N. V. Strand. 1961. Iowa livestock producers choice of markets. Iowa Agricultural and Home Economics Experiment Station Research Bulletin 492.

Narain, R. D. 1953. On the recurrence formula in sampling on successive occasions. Journal of the Indian Society of Agricultural Statistics, 5: 66-69.

________. 1954. The general theory of sampling on successive occasions. Bulletin de L'Institut International de Statistique, 34: 87-89.

Neter, J. and J. Waksburg. 1961. Measurement of nonsampling errors in a survey of homeowners expenditures for alterations and repairs. Proceedings of the Social Statistics Section, American Statistical Association 1961: 201-210.

Neyman, J. 1934. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. Journal of the Royal Statistical Society, 97: 558-625.

________. 1938. Contribution to the theory of sampling human populations. Journal of the American Statistical Association, 33: 101-116.

Oñate, B. T. 1960a. Development of multistage designs for statistical surveys in the Philippines. (Mimeo-Multilith Series, Number 3). Ames, Iowa, Statistical Laboratory, Iowa State University of Science and Technology.

________. 1960b. Development of multistage designs for statistical surveys in the Philippines. The Statistical Reporter, 4: 1-48.

________. 1960c. Development of multistage designs for statistical surveys in the Philippines. Unpublished Ph.D. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.

Patterson, H. D. 1950. Sampling on successive occasions with partial replacement of units. Journal of the Royal Statistical Society, Series B, 12: 241-255.

Rao, J. N. K. 1962a. On an unbiased method of variance estimation. (Unpublished paper. Manuscript). Ames, Iowa, Department of Statistics, Iowa State University of Science and Technology.

________. 1962b. Rotation sampling. (Unpublished paper. Manuscript). Ames, Iowa, Department of Statistics, Iowa State University of Science and Technology.

________ and J. E. Graham. 1962. On rotation sampling. (Abstract). Annals of Mathematical Statistics, 33: 1498.

Robson, D. S. 1952. Multiple sampling of attributes. Journal of the American Statistical Association, 47: 203-215.

________ and A. J. King. 1953. Doubling sampling and the Curtis impact survey. New York (Ithaca) Agriculture Experiment Station Memorandum 231.

Schach, C. F. 1962. Sampling problems involving rotational survey designs and unequal probability sampling. Unpublished M.S. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology.

Seal, K. C. 1951. On errors of estimates in various types of double sampling procedures. Sankhyā, 11: 125-144.

________. 1962. Use of out-dated frames in large scale sample surveys. Calcutta Statistical Association Bulletin, 11: 68-84.

Snedecor, G. W. and A. J. King. 1942. Recent developments in sampling for agricultural statistics. Journal of the American Statistical Association, 37: 95-102.

Sukhatme, P. V. 1954. Sampling theory of surveys with applications. Ames, Iowa, Iowa State College Press.

Tikkiwal, B. D. 1951. Theory of successive sampling. Unpublished Thesis for the Diploma in Agricultural Statistics. New Delhi, India, Library, Institute of Agricultural Research Statistics.

________. 1953. Optimum allocation in successive sampling. Journal of the Indian Society of Agricultural Statistics, 1: 100-102.

________. 1954. Theory of successive multiphase sampling. (Abstract) Annals of Mathematical Statistics 25: 174.

________. 1955a. Multiphase sampling on successive occasions. Unpublished Ph.D. thesis. Raleigh, North Carolina, Library, North Carolina State College.

________. 1955b. On the efficiency of estimates in successive multiphase sampling. (Abstract) Annals of Mathematical Statistics, 26: 158.

________. 1956a. An application of the theory of multiphase sampling on successive occasions to surveys of livestock marketing. Journal of the Karnatak University, 1: 120-130.

________. 1956b. A further contribution to the theory of univariate sampling on successive occasions. Journal of the Indian Society of Agricultural Statistics, 8: 84-90.

________. 1958a. An examination of the effect of matched sampling on the efficiency of estimators in the theory of successive sampling. Journal of the Indian Society of Agricultural Statistics, 10: 16-22.

________. 1958b. Theory of successive two-stage sampling. (Abstract) Annals of Mathematical Statistics, 29: 1291.

________. 1960. On the theory of classical regression and double sampling estimation. Journal of the Royal Statistical Society, Series B, 22: 131-138.

Ware, K. D. and T. Cunia. 1962. Continuous forest inventory with partial replacement of samples. Forest Science Monograph 3.

Woodruff, R. S. 1959. The use of rotating panels in the Census Bureau's monthly surveys. Proceedings of the Social Statistics Section, American Statistical Association 1959: 130-138.

Yates, F. 1949. Sampling methods for censuses and surveys. London, England, Charles Griffin and Company, Incorporated.

Zarkovich, S. S. 1961. Sampling methods and censuses. Volume 1. Collecting data and tabulation. Rome, Italy, Food and Agriculture Organization of the United Nations.

IX. ACKNOWLEDGMENTS

The author wishes to express his deep gratitude to Professor H. O. Hartley for his guidance throughout the preparation of this dissertation. Most sincere thanks are also extended to Professor J. N. K. Rao for his valuable suggestions made during the course of many informative discussions on the subject matter. The author is indebted to Mrs. June Smith and Mrs. Mary Anne Carney for their efforts in preparing the electronic computer programs.

Recognition must also be given to Professor Norman V. Strand for enlightening the author on many of the more practical aspects of sample survey methodology. Several past and present staff members and graduate students of Iowa State University have all assisted, in one way or another, to make the author's stay a happy and memorable one.

Acknowledgment is made for the support of this study by a basic research grant from the Army Research Office, United States Army.

Finally, the author wishes to express his heartfelt thanks to his wife, Audrey, for the encouragement given by her throughout the course of his graduate studies in statistics, and for the many hours spent in preparing this final typewritten form.

X. APPENDICES

A. Variance of the Composite Estimator of the Current Occasion Mean With a Geometric Trend in S_a

In the derivation of (3.16) it was assumed that $S_a = S_0$, $a = -1, -2, \ldots$. We now generalize this condition by assuming that there is a geometric trend in S_a,

$$S_a = S_0 k^a .$$

If the covariance structure is assumed stationary over time in the sense that

$$S_{a,\, a+t} = \rho^{-t} S_a S_{a+t} ,$$

it then follows that

$$S_{a,\, a+t} = (\rho/k)^{-t} k^{2a} S_0^2 .$$

The substitution of these values into (3.13) and an extremely lengthly simplification will finally lead to

$$V(\bar{x}_0') = (\tfrac{1}{n} - \tfrac{1}{N}) S_0^2 + B\Big[(r-1)(C + Q^r \rho^{r-1} E + Q^{m+r} \rho^{m+r} I)/k^3$$

$$+ (r-1)Q^m \rho^{m+1} G/k^2 + (Q^r \rho^{r-1} F + Q^m \rho^{m+1} H)/k^3$$

$$+ (D + Q^{m+r} \rho^{m+r} J)/k^2 \Big] ,$$

where

$$B = (2Q/k) / r^2 (r-1)^2 n_2 (1 - (Q/k)^2)(1 - (Q\rho/k)^{r+m})(1 - Q\rho/k)^2 ,$$

$$C = -(r-1)k^3 \rho + r(k^2+1)k^2 Q/2 - 2k^3 Q\rho + rk^2 Q\rho^2 + (k^2-1)rQ^3 \rho^2/2 ,$$

$$D = rkQ^2 - ((r^2 - r + 1)k^2 + r)Q^2\rho + rkQ^2\rho^2 ,$$

$$E = rk^2\rho - k\rho^2 + rQ^2\rho - kQ^2\rho^2 ,$$

$$F = -kr^2Q + (rk^2 + r)Q\rho - k(r^2 - 2r + 2)Q\rho^2 ,$$

$$G = (r-1)k^2 - (r-2)k^2Q - rkQ\rho - (k^2 + r)Q^2 + 2rkQ^2\rho ,$$

$$H = -rkQ^3 + r(k^2 + r)Q^3\rho - r^2kQ^3\rho^2 ,$$

$$I = -k^4r + k^3\rho + rk^2(k^2 + 1)Q/2 + k^3(r-2)Q\rho - rkQ^3\rho$$
$$+ (k^2 + 1)rQ^3\rho^2/2 ,$$

$$J = kr^2Q^2 - (r^2 + rk^2(r-1) + k^2)Q^2\rho + rkQ^2\rho^2 .$$

Upon setting $k = 1$ the foregoing expression reduced to (3.16) which is a check on the correctness of the lengthly calculations involved in the derivation.

B. Rotation Designs With $m < r$

The multi-cycle rotation designs developed in the text have all been constructed such that any rotation group drops out of the sample for $m \geq r$ consecutive occasions where r is the number of consecutive occasions it remains in the sample. There is no *a priori* reason that m should not be less than r however, as, for example, when rotating secondaries within small primary sampling units. The evaluation of the variance function of a composite estimator becomes exceedingly more complex when $m < r$ and for this reason discussion of such cases was avoided.

Reference here will be made only to an infinite cycle rotation design and the simple composite estimator of the current occasion mean. The general variance formula (3.12) is again applied to obtain the variance of $\bar{x}'_0$. The system of rotation weight variables specified in (3.5) and (3.6) are still applicable here, their value being independent of m. The main source of difficulty in the $m < r$ sampling situation arises in the evaluation of the component $\sum_{\substack{\alpha \neq \alpha' \\ =0}}^{-\infty} NE(w_{\alpha,k} w_{\alpha',k}) S_{\alpha,\alpha'}$ in (3.12).

It is not intended to give a term by term derivation of the general variance function, $V(\bar{x}'_0)$, here. Perhaps the most efficient procedure is to set up a rectangular array with the occasion serving as abscissa and visit number of the rotation group as ordinate, (see Figure 1). The various complexities introduced when $m < r$ will then become visually evident. Four different cases may be distinguished: (a) $r - m \geq 4$, (b) $r - m = 3$, (c) $r - m = 2$, (d) $r - m = 1$. Only the corresponding variance formulas are quoted here. They are derived by referring to (3.5) and (3.6) and the rectangular array which is not reproduced here.

For brevity let (1) $-\ell(r+m)$ be denoted by z, (2) $\sum_{\ell=0}^{\infty} Q^{\ell(r+m)}$ by Z_0, (3) $\sum_{\ell=1}^{\infty} Q^{\ell(r+m)}$ by Z_1, and (4) $\sum_{\alpha=-1}^{-\infty} Q^{-2\alpha}$ by Q_1. The finite population correction, $-S_0^2/N$, is deleted from the following equations.

(a) The case $r - m \geq 4$.

$$V(\bar{x}'_0) = S_0^2/n + Q^2 n_2 S_0^2/(nn_1) + Q_1 n_2 (Q^2 + 2n_2 Q/n_1 + 1) S_\alpha^2/(nn_1)$$

$$+ 2Z_1 n_2 Q(n_2 + Qn_1) S_{0,z}/(nn_1^2)$$

$$- 2Z_0 \sum_{t=1}^{m} Q^t n_2^2 (n_1 + n_2 Q)(tQ + r - t) S_{0,z-t}/(n^2 n_1^2)$$

$$+ 2Z_0 n_2^3 Q^{m+1} (-(r+m)Q^2$$

$$+ (-r^2 + r - rm + 2m)Q + m(r-1) S_{0,z-m-1}/(n^2 n_1^2)$$

$$+ 2Z_0 \sum_{t=m+2}^{r-1} n_2^3 Q^t (-(r+m)Q^2$$

$$+ (r(1-m) + 2m)Q + m(r-1)) S_{0,z-t}/(n^2 n_1^2)$$

$$+ 2Z_0 \sum_{t=r}^{m+r-1} n_2^3 Q^t ((t-r-m)Q^2 + ((t-m+2)r + 2(m-t))Q$$

$$+ r^2 + r(m-t-1) + t - m) S_{0,z-t}/(n^2 n_1^2)$$

$$- 2Z_0 Q_1 \sum_{t=1}^{m} Q^t n_2^3 (1-Q)^2 S_{a,a+z-t}/(n^2 n_1^2)$$

$$+ 2Z_0 Q_1 Q^{m+1} n_2^3 (-(r+m+1)Q^2$$

$$+ (-r^2 + 2r + 2m + 1)Q - (r+m)) S_{a,a+z-m-1}/(n^2 n_1^2)$$

$$- 2Z_0 Q_1 \sum_{t=m+2}^{r-2} Q^t (r+m) n_2^3 (1-Q)^2 S_{a,a+z-t}/(n^2 n_1^2)$$

$$+ 2Z_0 Q_1 Q^{r-1} n_2^3 (-Q^2 (r+m)$$

$$+ Q(2r + 2m - r^2) - (r+m)) S_{a,a+z-r+1}/(n^2 n_1^2)$$

$$- 2Z_0 Q_1 \sum_{t=r}^{r+m-1} Q^t n_2^3 (1 - Q)^2 (r + m - t) S_{a, a+z-t} / (n^2 n_1^2)$$

$$+ 2Z_1 Q_1 n_2 (Q^2 + 2n_2 Q/n_1 + 1) S_{a, a+z} / (nn_1) .$$

(b) The case $r - m = 3$.

By deleting the fourth from last term from the previous formula, the appropriate variance is given.

(c) The case $r - m = 2$.

$$V(\bar{x}_0') = S_0^2/n + Q^2 n_2 S_0^2/(nn_1) + Q_1 n_2 (Q^2 + 2n_2 Q/n_1 + 1) S_a^2/(nn_1)$$

$$+ 2Z_1 n_2 Q(n_2 + Qn_1) S_{0, z} / (nn_1^2)$$

$$- 2Z_0 \sum_{t=1}^{m} Q^t n_2^2 (n_1 + n_2 Q)(tQ + r - t) S_{0, z-t} / (n^2 n_1^2)$$

$$+ 2Z_0 n_2^3 Q^{m+1} (-(r + m)Q^2$$

$$+ (- r^2 + r - rm + 2m)Q + m(r - 1)) S_{0, z-m-1} / (n^2 n_1^2)$$

$$+ 2Z_0 \sum_{t=r}^{m+r-1} n_2^3 Q^t ((t - r - m)Q^2$$

$$+ ((t - m + 2)r + 2(m - t)) Q + r^2 + r(m - t - 1)$$

$$+ t - m) S_{0, z-t} / (n^2 n_1^2)$$

$$- 2Z_0 Q_1 \sum_{t=1}^{m} Q^t n_2^3 (1 - Q)^2 S_{a, a+z-t} / (n^2 n_1^2)$$

$$- 4Z_0 Q_1 Q^{m+1} n_2(n_2 Q + n_1)(n_2 + Qn_1) S_{a, a+z-m-1}/(n^2 n_1^2)$$

$$- 2Z_0 Q_1 \sum_{t=r}^{r+m-1} Q^t n_2^3 (1 - Q)^2 (r + m - t) S_{a, a+z-t}/(n^2 n_1^2)$$

$$+ 2Z_1 Q_1 n_2 (Q^2 + 2n_2 Q/n_1 + 1) S_{a, a+z}/(nn_1) .$$

(d) The case $r - m = 1$.

$$V(\bar{x}'_0) = S_0^2/n + Q^2 n_2 S_0^2/(nn_1) + Q_1 n_2 (Q^2 + 2n_2 Q/n_1 + 1) S_a^2/(nn_1)$$

$$+ 2Z_1 n_2 Q(n_2 + Qn_1) S_{0,z}/(nn_1^2) - 2Z_0 \sum_{t=1}^{m} Q^t n_2^2 (n_1 + n_2 Q)(tQ + r - t)$$

$$S_{0, z-t}/(n^2 n_1^2) + 2Z_0 n_2 Q^{m+1} (1 - Q)(n_2 Q + n_1) S_{0, z-m-1}/(n^2 n_1)$$

$$+ 2Z_0 \sum_{t=r+1}^{m+r-1} Q^t n_2^3 ((t - r - m)Q^2 + ((t - m + 2)r + 2(m - t))Q + r^2$$

$$+ r(m - t - 1) + t - m) S_{0, z-t}/(n^2 n_1^2) + 2Z_0 Q_1 \sum_{t=1}^{m-1} n_2^2 Q^t (1 - Q)^2 ((m + t)n_2$$

$$- n_1) S_{a, a+z-t}/(n^2 n_1^2) + 2Z_0 Q_1 Q^m n_2 (n_2 Q + n_1)(n_2 + Qn_1)$$

$$S_{a, a+z-t}/(n^2 n_1^2) - 2Z_0 n_2 (n_2 + Qn_1)(n_1 + n_2 Q) Q_1 Q^{m+1}$$

$$S_{a, a+z-m-1}/(n^2 n_1^2) - 2Z_0 Q_1 \sum_{t=r+1}^{r+m-1} Q^t n_2^3 (1 - Q)^2 (r + m - t)$$

$$S_{a, a+z-t}/(n^2 n_1^2) + 2Z_1 Q_1 n_2 (Q^2 + 2n_2 Q/n_1 + 1) S_{a, a+z}/(nn_1) .$$