Journal of Official Statistics
Vol. 8. No. 4, 1992, pp. 405–416

The Use of Composite Estimators with Two Stage Repeated Sample Designs

D. Holt[1] *and T. Farver*[2]

Abstract: The use of composite estimation for inference from repeated survey designs has a long history. Whilst algebraic extensions have been made to take account of two stage sampling designs, there does not appear to have been a systematic attempt to understand how the clustered structure of the population affects the potential gains from composite estimation. What are the population structure characteristics, for example, that would imply greater or less potential benefits from composite estimation than in the simple random sampling case? A simple superpopulation model is used to characterise a clustered population structure over time and this is used to explore the effect of the population structure on the potential gains in efficiency for composite estimation.

Key words: Composite estimation; repeated surveys; two stage design; random effects; superpopulation model.

1. Introduction

National statistical agencies conduct many surveys at regular intervals to provide estimates of the population mean or total on each occasion (i.e., estimates of "level"). The difference between population means on successive occasions is also of interest (i.e., estimates of "change"). More rarely estimates of level for successive periods are sometimes averaged or totalled to provide an estimate for a longer period (e.g., four quarterly estimates averaged to provide an annual estimate). Whilst this latter case is comparatively unusual for social surveys, it is adopted for some epidemiological or veterinary applications when estimates such as the incidence of new cases of a disease or condition are required. In this case regular surveys throughout the period of interest are used to control for seasonality effects. We refer to this case as an estimate of "average level".

Many repeated surveys have rotating designs in which units selected into the surveys are retained for a fixed number of occasions and are then discarded from the sample. At each survey period the sample will comprise a set of usually equal rotation groups with some sample members included for the first time, some for the second time and so on.

The fact that observations on the same unit across time are positively correlated can lead to efficiencies for estimation and the theoretical foundations for the design and estimation of such surveys were developed by Jessen (1942) and Patterson (1950). The

[1] Department of Social Statistics, University of Southampton, Southampton SO9 5NH, U.K.
[2] Department of Epidemiology and Preventive Medicine, University of California, Davies, CA 95616-8735, U.S.A.

original theory was developed for simple random sampling although extensions to two stage sampling were provided by Singh (1968) for estimates of level, change and average level for the special case when whole PSU's are either retained or renewed. Other authors, notably Abraham, Khosta and Kathuria (1969), Singh and Kathuria (1969), Kathuria (1975) and Okafor (1987) have extended the results to more general two stage sample rotation patterns, although the main emphasis has been on estimating the level on each occasion. All of this work follows the usual sample survey theoretical framework in which estimators and properties of estimators are based on the randomization distribution determined by the sampling scheme.

The approach adopted in this paper is to propose a simple, but plausible, superpopulation model with parameters which are readily interpretable as features of the population structure. Within this framework it is relatively simple to derive minimum variance unbiased estimators for level, change and average level and express these in algebraic form. Under simplifying assumptions and without the complexity of finite population corrections, variances of estimators may also be expressed in relatively simple form.

Our purpose is to provide a framework in which the relative efficiencies of alternative estimators are more easily understood and in particular to provide an insight into those situations where the population structure may or may not have an effect on the gains in efficiency obtained from optimal estimation.

2. Model Framework

We assume, for simplicity, that the population consists of N clusters each containing M units although the formulation may be extended to take account of unequal cluster sizes. We assume that on occasion t the jth member of the ith cluster takes the value y_{tij}

$$y_{tij} = \mu_t + \alpha_{ti} + \varepsilon_{tij} \qquad (1)$$

and that

$$\begin{aligned}
V(\varepsilon_{tij}) &= \sigma_\varepsilon^2 && \text{for all } t, i, j; \\
V(\alpha_{ti}) &= \sigma_\alpha^2 && \text{for all } t, i \\
\text{Cov}(\varepsilon_{tij'}, \varepsilon_{t+1,i'j'}) &= \rho_2\sigma_\varepsilon^2 && (i,j) = (i',j') \\
&= 0 && \text{else} \\
\text{Cov}(\alpha_{ti}, \alpha_{t+1,i'}) &= \rho_1\sigma_\alpha^2 && i = i' \\
&= 0 && i \neq i' \\
\text{Cov}(\varepsilon_{tij}, \alpha_{t'i'}) &= 0 && \text{for all } t, t', i, i' \text{ and } j
\end{aligned}$$

This is a simple, but natural extension to the usual components of variance superpopulation model for one point in time t where α and ε are random components for the cluster and individual effects respectively and the usual intra-cluster correlation ρ is given by $\rho = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_\varepsilon^2)$. For this extension the variances of the cluster and individual level effects are each assumed to be constant over time so that the intra-cluster correlation ρ is the same for all time points t. However, the cluster effects are correlated over time with correlation coefficient ρ_1 and the individual level effects are correlated over time with correlation coefficient ρ_2.

We note that observations on the same unit are correlated over time since

$$\text{Cov}(y_{tij}, y_{t+1,ij}) = \rho_2\sigma_\varepsilon^2 + \rho_1\sigma_\alpha^2 \qquad (2)$$

and hence $\text{Corr}(y_{tij}, y_{t+1,ij}) = \rho\rho_1 + (1-\rho)\rho_2$.

Furthermore observations on different units within the same cluster are also correlated over time since

$$\text{Cov}(y_{tij}, y_{t+1,i,j'}) = \rho_1\sigma_\alpha^2 \qquad j \neq j'. \quad (3)$$

For the case of two surveys at times t and

$t + 1$ we define the population parameters for level, change and average level to be μ_{t+1}, $\Delta = \mu_{t+1} - \mu_t$ and $\xi = \frac{1}{2}(\mu_t + \mu_{t+1})$ respectively. In the superpopulation framework these quantities are model parameters and each may be identified with natural finite population values.

3. Survey Design and Sample Properties

We consider a survey design in which n PSU's are sampled on each occasion and from each sampled PSU a subsample of m units is selected. From the first occasion (t) to the next ($t + 1$), np_1 sampled PSU's are retained (matched PSU's) and the residual nq_1 are dropped from the sample after the first occasion and replaced at time $t + 1$ by a fresh sample of PSU's ($p_1 + q_1 = 1$). Within each PSU which is retained for both occasions there is a similar rotation pattern of second stage units with mp_2 being retained for the second occasion and mq_2 replaced by a new sample from within the same PSU ($p_2 + q_2 = 1$).

The sample data from occasion t may be separated into three subgroups as follows:

$\bar{y}_{tuu}$ is the mean of unmatched units from unmatched PSU's at time t

$\bar{y}_{tmu}$ is the mean of unmatched units from matched PSU's at time t

$\bar{y}_{tmm}$ is the mean of matched units from matched PSU's at time t.

The sample data from occasion $t + 1$ may be treated similarly.

It may be shown that the variance covariance matrix of these six sample means is given by

$$\Sigma = \begin{pmatrix} A+E & 0 & 0 & 0 & 0 & 0 \\ & B+D & D & 0 & \rho_1 D & \rho_1 D \\ & & C+D & 0 & \rho_1 D & \rho_1 D + \rho_2 C \\ & & & A+E & 0 & 0 \\ & & & & B+D & D \\ & & & & & C+D \end{pmatrix}$$

where the first three rows and columns relate to the three sample means for time t (in order uu, mu, mm) and the second three rows and columns relate to time $t + 1$ and,

$$A = \frac{\sigma_\varepsilon^2}{mnq_1}, B = \frac{\sigma_\varepsilon^2}{mnp_1 q_2}, C = \frac{\sigma_\varepsilon^2}{mnp_1 p_2},$$

$$D = \frac{\sigma_\alpha^2}{np_1}, \quad \text{and} \quad E = \frac{\sigma_\alpha^2}{nq_1}. \tag{5}$$

Note that this pattern can be extended to rotation patterns that span more periods but the proportion of overlap between PSU's and individuals will change for each off-diagonal block of the enlarged matrix. A second complication would be introduced if the correlation between times t and t' diminished as $|t - t'|$ became larger which is frequently observed in many situations.

In general we consider an estimator for some parameter θ which is made up of a linear combination of the six sample means

$$\hat{\theta} = \psi_1 \bar{y}_{tuu} + \psi_2 \bar{y}_{tmu} + \psi_3 \bar{y}_{tmm} + \phi_1 \bar{y}_{t+1,uu} + \phi_2 y_{t+1,mu} + \phi_3 \bar{y}_{t+1,mm} \tag{6}$$

and hence

$$\begin{aligned} V(\hat{\theta}) = & (\phi_1^2 + \psi_1^2)(A + E) \\ & + (\phi_2^2 + \psi_2^2)(B + D) \\ & + (\phi_3^2 + \psi_3^2)(C + D) \\ & + 2(\phi_2\phi_3 + \psi_2\psi_3)D \\ & + 2(\phi_2 + \phi_3)(\psi_2 + \psi_3)\rho_1 D \\ & + 2\phi_3\psi_3\rho_2 C. \end{aligned}$$

In the following sections we consider the cases of level, change and average level.

4. Estimating Level

The simplest estimator of μ_{t+1} is the sample mean of all data collected at time $t + 1$. In equation (6) this corresponds to the case $\psi = \{\psi_1, \psi_2, \psi_3\} = \{0, 0, 0\}$ and $\phi = \{\phi_1,$

$\phi_2, \phi_3\} = \{q_1, p_1q_2, p_1p_2\}$ yielding the well-known variance expression

$$V(\bar{y}_{t+1}) = \frac{\sigma_T^2}{nm}\{1 + (m - 1)\rho\} \tag{8}$$

where $\sigma_T^2 = \sigma_\varepsilon^2 + \sigma_\alpha^2$.

The MVUE of μ_{t+1}, designated $\hat{\mu}_{t+1}$, is obtained for given design parameters (m, n, p_1 and p_2) and population structure (ρ, ρ_1 and ρ_2) by minimizing (7) subject to the constraints $\Sigma\phi_i = 1$; $\Sigma\psi_i = 0$.

It may be shown that the optimum values of $\{\phi_i\}$ and $\{\psi_i\}$ are given by

$$\hat{\phi}_3 = BKX, \hat{\psi}_3 = -BKY,$$
$$\hat{\phi}_2 = CK(X - \rho_2 Y), \hat{\psi}_2 = CK(\rho_2 X - Y),$$
$$\hat{\phi}_1 = 1 - \hat{\phi}_2 - \hat{\phi}_3, \psi_1 = -(\hat{\psi}_2 + \hat{\psi}_3), \tag{9}$$

where

$$X = A(B + C) + B(C + D + E) + CD(1 + \rho_1\rho_2) + CE,$$
$$Y = \rho_1 D(B + C) + \rho_2 C(A + B + D + E)$$
$$K = (A + E)/(X - Y)(X + Y)$$

and $V(\hat{\mu}_{t+1})$ may be obtained by substituting the optimum values of $\{\phi_i\}$ and $\{\psi_i\}$ into equation (7).

The relative efficiency of composite estimation $\{V(\hat{\mu}_{t+1})/(V(\bar{y}_{t+1})\}$ 100% may be investigated numerically for a variety of design parameters (m, p_1 and p_2) and population structure parameters (ρ, ρ_1 and ρ_2). We shall restrict the choice of design parameters to those that are plausible in practice rather than those appropriate for an optimal design. Practical and cost considerations lead to designs which retain a substantial proportion of the sample from one occasion to the next whereas optimal design would be achieved by a much smaller retention.

Nevertheless, the investigation of the relative efficiency of $\hat{\mu}_{t+1}$ is a six dimensional non-linear function of the design and population structure parameters. Numerical investigations show that of these six parameters, the least important in general are the cluster sample size m and the intra-cluster correlation coefficient ρ. Thus for simplicity, Table 1 contains the relative efficiency of $\hat{\mu}_{t+1}$ for a choice of values of p_1, p_2, ρ_1 and ρ_2 but for the case when $\rho = 0.05$ and $m = 12$.

In broad terms, Table 1 shows what we would expect: that the greatest gains in efficiency occur when both correlations ρ_1 and ρ_2 are high and the proportion of retained clusters and individual units within clusters are low. However these gains in efficiency do depend on the population structure (ρ_1 and ρ_2) and design (p_1 and p_2). In general rotation of the clusters leads to greater gains in efficiency than rotation of the units within cluster. Thus $p_1 = 0.95$, $p_2 = 0.67$ and $p_1 = 0.67$, $p_2 = 0.95$ each result in the same proportion of units being retained from one occasion to the next but the former pattern will lead to lower gains in efficiency than the latter. This is to be expected since the effective matched sample size is increased by retaining the same PSU's taking it further away from the optimal matched proportion. In practice of course, this must be offset by the fact that rotation of PSU's may be more expensive than rotation of units within PSU's and thus for less extreme alternatives such as $p_1 = 0.9$, $p_2 = 0.8$ compared to $p_1 = 0.8$, $p_2 = 0.9$ the question of alternative costs for the two designs would dominate the negligible differences in efficiency gains.

The effects of the population structure (ρ_1 and ρ_2) indicates that high values of ρ_2 have a greater effect on the efficiency gain that ρ_1. However it is notable that high values of ρ_1 (the correlation between the cluster effects

Table 1. Relative efficiency (%) of $\hat{\mu}_{t+1}$ compared to $\bar{y}_{t+1}$ for values of p_1, p_2, ρ_1 and ρ_2 when m = 12 and ρ = 0.05

p_2	.95				.90				.80				.67			
ρ_2	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9
ρ_1																
	p_1 = 0.95															
.4	99	98	96	96	98	97	95	94	98	96	92	90	97	94	90	87
.6	99	97	96	95	98	97	94	93	98	95	92	90	97	94	89	87
.8	98	97	95	94	98	96	94	93	97	95	91	89	97	94	89	86
.9	98	97	95	94	98	96	94	92	97	95	91	89	97	93	89	86
	p_1 = 0.90															
.4	98	97	95	94	98	96	94	92	97	95	91	89	97	94	89	86
.6	98	96	94	93	97	95	93	91	97	94	90	88	97	93	88	85
.8	97	95	93	92	97	94	92	90	96	93	89	87	96	92	87	84
.9	97	95	92	91	96	94	91	89	96	93	89	87	96	92	87	84
	p_1 = 0.80															
.4	97	95	92	90	97	95	91	89	97	94	90	87	87	93	88	85
.6	96	94	90	89	96	93	90	87	96	92	88	85	96	92	87	83
.8	95	92	89	86	95	92	88	85	95	91	86	83	95	91	85	81
.9	94	91	87	85	94	91	87	84	94	90	85	82	94	90	84	80
	p_1 = 0.67															
.4	96	93	90	87	96	93	89	87	96	93	88	85	96	93	87	83
.6	95	91	87	85	95	91	87	84	95	91	86	82	95	91	85	81
.8	93	89	84	81	93	89	84	81	93	89	83	79	93	89	82	78
.9	92	88	83	80	92	88	82	79	92	88	81	78	92	88	81	76

over time) do result in a gain in efficiency of up to 8% even when the unit level correlation ρ_2 is modest.

Table 2 contains the relative efficiency of $\hat{\mu}_{t+1}$ when $p_1 = p_2 = 0.8$ and $m = 12$ and examines the effect of different values of the intra-cluster correlation ρ.

Table 2 shows that the effect of ρ is small except when ρ_1 is small (e.g., 0.4) and ρ_2 is large (e.g., 0.9).

Table 3 contains the relative efficiency of $\hat{\mu}_{t+1}$ when $p_1 = p_2 = 0.8$, $\rho = 0.05$ for various values of ρ_1, ρ_2 and m. It may be seen that the effect of m is modest and once again the largest effect is when ρ_1 is small (e.g., 0.4) and ρ_2 is large (e.g., 0.9).

The pattern of the results in Table 1 conforms in a general way to the corresponding results for simple random sampling which in modelling terms is equivalent to a

Table 2. Relative efficiency (%) of $\hat{\mu}_{t+1}$ for values of ρ, ρ_1 and ρ_2 when m = 12 and p_1 = p_2 = 0.8

ρ	0.01				0.05				0.10				0.20			
ρ_1 ρ_2	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9
.4	96	92	86	81	97	94	90	87	97	95	92	90	97	96	94	93
.6	96	92	85	81	96	92	88	85	95	93	90	88	95	93	92	91
.8	96	91	85	80	95	91	86	83	93	91	87	85	92	90	88	87
.9	96	91	84	80	94	90	85	82	92	89	86	84	90	89	86	85

Table 3. Relative efficiency (%) of $\hat{\mu}_{t+1}$ for values of m, ρ_1 and ρ_2 when $\rho = 0.05$ and $p_1 = p_2 = 0.8$

m	6				12				24				36			
ρ_1 ρ_2	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9
.4	97	93	87	84	97	94	90	87	98	95	92	90	97	95	93	92
.6	96	92	86	83	96	92	88	85	95	93	90	88	95	93	91	89
.8	95	91	85	82	95	91	86	83	93	91	87	85	93	90	88	86
.9	95	91	85	81	94	90	85	82	92	89	86	83	91	89	86	84

simplified model with no components of variance for cluster effects.

The relative efficiency of the usual composite estimator for simple random sampling is given (e.g., Cochran 1977) by

$$re = \frac{V(\hat{\mu}_{t+1})}{V(\bar{y}_{t+1})} = \frac{1 - q\rho^2}{1 - q^2\rho^2} \qquad (10)$$

where q is the proportion of unmatched units and ρ is the correlation between observations on the same unit on the two occasions. If $\rho = 0.8$, for example, the optimum value for q is 0.62 ($p = 0.38$) yielding a relative efficiency of 80% but with a low proportion of retained units. For a common design with $q = 0.2$ ($p = 0.8$) and $\rho = 0.8$ the relative efficiency is 89.5%. The greater the retention between surveys, the greater the difference between the actual and optimal designs and the more modest the efficiency gain for the composite estimator. A corresponding case in Table 1 is given by $p_1 = p_2 = 0.9$ (matched proportion = 0.81) and $\rho = 0.05$, $\rho_1 = \rho_2 = 0.8$ yielding an overall correlation between observations on the same unit of $\rho\rho_1 + (1 - \rho)\rho_2 = 0.8$. The corresponding efficiency gain given in Table 1 is 92% compared to the simple random sampling efficiency gain of 89.5%. The numerical results can be extended to include lower proportions of retained units such as $p_1 = p_2 = 0.4$. For brevity these results are not presented in the tables but the relative efficiency of $\hat{\mu}_{t+1}$ is slightly higher than for simple random sampling.

In general the relative efficiency for $\hat{\mu}_{t+1}$ with two stage sampling is higher than for simple random sampling although there are exceptions for some combinations of parameters. In the main the ratio of the two relative efficiencies is about 1.05 showing that a smaller gain in efficiency is obtained for two stage sampling. However the largest differences occur when p_2 is small (i.e., 0.67) and ρ_2 is large (i.e., 0.8 or 0.9) when $\rho = 0.05$ and $m = 12$ leading to a ratio as high as 1.1. In the most adverse cases when $\rho = 0.1$ or 0.2 and for large cluster sizes $m = 24$ or 36 the largest observed ratio was 1.2.

Thus as a rough guide to the relative efficiency for composite estimation used in conjunction with a two stage sample, the corresponding expression for simple random sampling could be used. A simple multiplicative factor of 1.05 would yield a reasonable approximation to the expected relative efficiency. This approximation will be too low if ρ_2 is high and a large rotation rate is used within each PSU and will be at its worst if in addition ρ and/or m are particularly large. In general efficiency gains for composite estimation for level are modest for both simple random sampling and two stage sampling. The best situation for simple random sampling is when the retained proportion is small and the correlation for observations on the same unit is high. These are the very conditions when the two stage population structure ameliorates the potential efficiency gain and so prevents a substantial gain being achieved.

5. Estimating Change

The simplest estimator of Δ is $\hat{\Delta}_s$ the difference between the sample means: $\bar{y}_{t+1} - \bar{y}_t$. This will have smaller variance than would be achieved by two independent samples at times t and $t + 1$ because of the effect of the overlap across the two samples. In equation (6) $\hat{\Delta}_s$ corresponds to the case $\phi = -\psi = \{q_1, p_1 q_2, p_1 p_2\}$ and it may be shown that

$$V(\hat{\Delta}_s) = \frac{2\sigma_T^2}{mn}(1 - p_1 p_2 \rho_2) \times \left[1 + \left\{\frac{m(1 - p_1 \rho_1)}{1 - p_1 p_2 \rho_2} - 1\right\}\rho\right]. \quad (11)$$

The MVUE of Δ, designated $\hat{\Delta}$ is obtained by minimizing (7) subject to the constraints $\Sigma\phi_i = 1, \Sigma\psi_i = -1$.

It may be shown that the optimum values of $\{\phi_i\}$ and $\{\psi_i\}$ are given by

$$\hat{\phi}_3 = -\hat{\psi}_3 = \frac{B(A + E)}{X - Y},$$

$$\hat{\phi}_2 = -\hat{\psi}_2 = \frac{C(1 - \rho_2)(A + E)}{X - Y}, \quad (12)$$

and

$$\hat{\phi}_1 = -\hat{\psi}_1 = 1 - \frac{(A + E)}{X - Y}\{B + C(1 - \rho_2)\}.$$

The relative efficiency of composite estimation compared to $\hat{\Delta}_s$ may be investigated numerically for a variety of design parameters (m, p_1 and p_2) and population structure parameters (ρ, ρ_1 and ρ_2). We restrict the choice of design parameters to the same values as were used in the previous section which are considered to be plausible in practice.

Table 4 contains the relative efficiency of $\hat{\Delta}$ compared to $\hat{\Delta}_s$ for values of p_1, p_2, ρ_1 and ρ_2 when $m = 12$ and $\rho = 0.05$

$$re = \frac{V(\hat{\Delta})}{V(\hat{\Delta}_s)} 100\%. \quad (13)$$

We note that this represents the additional gain in efficiency for composite estimation after taking into account the lower variance of $\hat{\Delta}_s$ associated with the overlap between the sample from the two periods. The consistency with analogous numerical results for simple random sampling is maintained in that efficiency gains are much larger for estimating change than for level. The general pattern and conclusions are

a. The greatest gains in efficiency are associated with high values of ρ_1 and ρ_2.

b. For any overall proportion of overlap, $p_1 p_2$, the greater gains in efficiency occur when p_2 is higher and p_1 lower. This is consistent with the pattern for estimating level.

c. For given values of p_1 and p_2, ρ_2 has a stronger effect on the gain in efficiency, although when ρ_2 is high, high values of ρ_1 can have a substantial additional effect.

Table 5 contains the relative efficiency of $\hat{\Delta}$ compared to $\hat{\Delta}_s$ for values of ρ, ρ_1 and ρ_2 when $m = 12$ and $p_1 = p_2 = 0.8$. The effect of ρ is small except when ρ_1 is small and ρ_2 is large. In this situation the effect of ρ is significant. This patterns is precisely the same as for estimating level in Table 2.

Table 6 contains the relative efficiency of $\hat{\Delta}$ compared to $\hat{\Delta}_s$ for values of ρ_1, ρ_2 and m when $p_1 = p_2 = 0.8$ and $\rho = 0.05$. When ρ_1 and ρ_2 are unequal, the effect of m is much stronger than for estimating level (Table 3). The strongest effects occur when ρ_1 is small and ρ_2 is large.

The pattern of results in Table 4 conforms in a general way with the corresponding results for simple random sampling. In this case the relative efficiency for composite

Table 4. Relative efficiency (%) of $\hat{\Delta}$ for values of p_1, p_2, ρ_1 and ρ_2 when $m = 12$ and $\rho = 0.05$

p_2	.95				.90				.80				.67			
ρ_2	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9
ρ_1																
	$p_1 = 0.95$															
.4	98	96	91	87	98	94	87	81	97	91	81	73	96	89	76	66
.6	97	94	87	81	97	92	83	75	96	89	77	66	95	87	71	59
.8	96	91	81	71	95	89	77	65	94	86	70	55	93	84	64	48
.9	95	90	77	63	95	88	72	56	93	84	65	48	93	82	60	41
	$p_1 = 0.90$															
.4	97	94	87	82	97	92	84	78	96	90	79	71	95	88	75	65
.6	96	91	82	74	95	90	79	70	94	87	74	63	94	86	70	57
.8	94	87	74	62	93	86	71	58	92	83	66	51	92	82	62	46
.9	92	84	68	53	92	83	65	49	91	81	60	43	91	79	57	39
	$p_1 = 0.80$															
.4	96	91	82	76	95	90	80	73	95	88	77	68	95	87	74	64
.6	93	86	75	66	93	86	73	63	93	84	70	59	93	84	67	55
.8	90	81	64	51	90	80	63	49	90	79	60	46	90	79	59	43
.9	88	77	57	42	88	76	56	40	88	76	54	37	88	76	53	36
	$p_1 = 0.67$															
.4	94	88	78	71	94	88	77	69	94	87	75	66	94	87	74	64
.6	91	83	70	60	91	83	69	58	91	82	67	56	92	82	66	54
.8	87	76	58	45	87	76	57	44	87	76	57	43	88	77	57	42
.9	84	71	51	35	84	72	50	35	85	72	50	34	86	73	51	34

estimation is given by

$$re = \frac{V(\hat{\Delta})}{V(\hat{\Delta}_s)} = \frac{1-\rho}{(1-\rho q)(1-\rho p)}. \qquad (14)$$

If we consider the case when $p_1 = p_2 = 0.8$, $\rho = 0.05$ and $m = 12$, Table 7 contains the ratio of the relative efficiency under simple random sampling compared to the relative efficiency under the two stage model.

We note that when $\rho_2 = 0.9$ the relative efficiency of composite estimation under two stage sampling does not yield as much of an efficiency gain as in the case of simple random sampling (e.g., $\rho_1 = 0.4$, $\rho_2 = 0.9$, ratio $= 1.64$). However when ρ_1 is high (0.9)

Table 5. Relative efficiency (%) of $\hat{\Delta}$ for values of ρ, ρ_1 and ρ_2 when $m = 12$ and $p_1 = p_2 = 0.8$

ρ	0.01				0.05				0.1				0.2			
ρ_2	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9
ρ_1																
.4	94	84	64	46	95	88	77	68	95	91	84	79	95	93	89	87
.6	94	83	61	42	93	84	70	59	91	85	75	68	90	86	81	77
.8	93	82	58	38	90	79	60	46	85	76	62	52	80	73	63	57
.9	93	81	57	36	88	76	54	37	81	70	52	39	71	62	49	41

Table 6. Relative efficiency (%) of $\hat{\Delta}$ for values of m, ρ_1 and ρ_2 when $\rho = 0.05$ and $p_1 = p_2 = 0.8$

m	6				12				24				36			
ρ_2	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9
ρ_1																
.4	95	86	70	58	95	88	77	68	95	91	83	78	95	92	86	83
.6	93	84	65	50	93	84	70	59	92	85	75	68	91	86	78	73
.8	92	81	59	42	90	79	60	46	86	77	62	51	83	75	63	54
.9	91	79	56	36	88	76	54	37	82	70	52	39	77	67	51	40

and ρ_2 is low (0.4, 0.6) the composite estimator yields a greater efficiency gain when used in conjunction with two stage sampling as compared to simple random sampling.

The pattern becomes more distinct as ρ and m increase. In extreme cases such as $\rho = 0.2$ and $m = 36$ the ratio of the two relative efficiencies can be as low as 0.5 when ρ_1 is high (0.9) and ρ_2 is low (0.2). This shows that composite estimation yields substantially greater gains in efficiency in this case for two stage models as compared to simple random sampling.

However when the situation is reversed and ρ_1 is small (0.2) and ρ_2 large (0.9) the ratio of relative efficiences can be as high as 1.7 showing that composite estimation does not lead to as much of an efficiency gain for two stage models compared to simple random sampling.

In general when ρ_1 and ρ_2 are both large (0.9) the composite estimator for two stage sampling does not quite achieve as high an efficiency gain as in the case of simple random sampling.

Table 7. Ratio (× 100) of relative efficiency of composite estimator under a two stage model to that under single random sampling

ρ_1 ρ_2	.4	.6	.8	.9
.4	101	106	126	164
.6	99	102	118	151
.8	96	96	105	127
.9	94	93	96	107

6. Estimating Average Level

The simplest estimator of $\xi = \frac{1}{2}\{\mu_t + \mu_{t+1}\}$ is $\hat{\xi}_s$ the average of the sample means: $\frac{1}{2}\{\bar{y}_{t+1} + \bar{y}_t\}$. In equation (6) this corresponds to the case $\phi = \psi = \frac{1}{2}\{q_1, p_1 q_2; p_1 p_2\}$ and it may be shown that

$$V(\hat{\xi}_s) = \frac{\sigma_T^2}{2mn}(1 + p_1 p_2 \rho_2)$$
$$\times\left[1 + \left\{\frac{m(1 + p_1\rho_1)}{1 + p_1 p_2 \rho_2} - 1\right\}\rho\right]. \quad (15)$$

The MVU of ξ, designated $\hat{\xi}$ is obtained by minimizing (7) subject to the constraints $\Sigma\phi_i = \Sigma\psi_i = \frac{1}{2}$.

It may be shown that the optimal values of $\{\phi_i\}$ and $\{\psi_i\}$ are given by

$$\hat{\psi}_3 = \hat{\phi}_3 = \frac{B(A + E)}{2(X + Y)}$$
$$\hat{\psi}_2 = \hat{\phi}_2 = \frac{C(1 + p_2)(A + E)}{2(X + Y)} \quad (16)$$
$$\hat{\psi}_1 = \hat{\phi}_1 = \tfrac{1}{2} - \hat{\phi}_2 - \hat{\phi}_3.$$

The relative efficiency of composite estimation compared to $\hat{\xi}_s$ may be investigated numerically for the same design parameters (m, p_1 and p_2) and population structure parameters (ρ, ρ_1 and ρ_2) as used in previous sections. Table 8 contains the relative efficiency for values of p_1, p_2, ρ_1 and ρ_2 when $m = 12$ and $\rho = 0.05$. The overall

Table 8. Relative efficiency (%) of $\hat{\xi}$ for values of p_1, p_2, ρ_1 and ρ_2 when m = 12 and ρ = 0.05

p_2	.95				.90				.80				.67			
ρ_2	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9
ρ_1																
	p_1 = 0.95															
.4	99	99	98	97	99	98	97	97	99	97	96	95	98	97	95	94
.6	99	98	98	97	99	98	97	96	98	97	96	95	98	97	95	94
.8	99	98	98	97	99	98	97	96	98	97	96	95	98	97	95	94
.9	99	98	97	97	99	98	97	96	98	97	96	95	98	97	95	94
	p_1 = 0.90															
.4	99	98	97	96	99	98	96	96	98	97	95	95	98	96	95	94
.6	98	98	97	96	98	97	96	95	98	97	95	94	98	96	95	94
.8	98	97	96	96	98	97	96	95	98	96	95	94	98	96	94	93
.9	98	97	96	96	98	97	95	95	98	96	95	94	97	96	94	93
	p_1 = 0.80															
.4	98	97	96	95	98	97	95	94	98	96	95	94	98	96	94	93
.6	97	96	95	94	98	97	95	94	98	96	94	93	97	96	94	93
.8	97	96	94	93	97	95	94	93	97	95	93	93	97	95	93	92
.9	96	95	94	93	96	95	94	93	96	95	93	92	96	95	93	92
	p_1 = 0.67															
.4	97	95	94	94	97	96	94	93	97	96	94	93	97	96	94	93
.6	97	95	93	93	97	95	93	92	97	95	93	92	97	95	93	92
.8	94	94	93	92	96	94	92	92	96	94	92	91	96	94	93	92
.9	95	94	92	91	95	94	92	91	95	94	92	91	96	94	92	91

conclusion is that composite estimation offers little improvement over the simple estimator. The most favourable situation is when p_1 is small (.67) and ρ_1 and ρ_2 are high but even in this case the efficiency gains are extremely modest.

Table 9 contains the relative efficiencies for values of ρ, ρ_1 and ρ_2 when m = 12 and $p_1 = p_2 = 0.8$ and it is clear that ρ has no effect on the gain in efficiency. Table 10 contains the corresponding results for values of m, ρ_1, ρ_2 when $\rho = 0.05$ and $p_1 = p_2 = 0.8$. The conclusion is the same: that gains in efficiency are extremely modest.

These results are analogous to the corresponding results for efficiency gains using composite estimators in the case of simple random sampling. In general there is even

Table 9. Relative efficiency (%) of $\hat{\xi}$ for values of ρ, ρ_1 and ρ_2 when m = 12 and $p_1 = p_2 = 0.8$

ρ	0.01				0.05				0.1				0.2			
ρ_2	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9
ρ_1																
.4	98	95	93	92	98	96	95	94	98	97	96	95	98	97	97	96
.6	97	95	93	92	97	96	94	93	97	96	95	94	97	96	95	95
.8	97	95	93	91	97	95	93	93	96	95	94	93	96	95	94	94
.9	97	95	93	91	95	95	93	92	96	95	93	93	95	94	93	93

Table 10. Relative efficiency (%) of ξ for values of m, ρ_1 and ρ_2 when ρ = 0.05 and $p_1 = p_2 = 0.8$

m	6				12				24				36			
ρ_2	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9	.4	.6	.8	.9
ρ_1																
.4	98	96	94	93	98	96	95	94	98	97	96	95	98	97	96	96
.6	97	95	93	92	97	96	94	93	97	96	95	94	97	96	95	95
.8	97	95	93	92	97	95	93	93	96	95	94	93	96	95	94	93
.9	97	95	93	92	96	95	93	92	96	95	93	93	95	94	93	93

less benefit than for the corresponding results for estimating μ_{t+1}.

7. Discussion

In the case of simple random sampling the efficiency gain to be achieved from composite estimation when estimating μ_{t+1} for designs that are generally used in practice is modest unless the period to period correlation ρ is extremely high. For two stage sampling this general pattern is confirmed and for most parameter combinations the efficiency gain is smaller than in the simple random sampling case. Nonetheless, the role of the population structure is interesting and the results reveal the separate effects of the PSU level correlation ρ_1 and the individual level correlation ρ_2. Most designs used in practice will employ a higher matched sample than is optimal. One interpretation of the results is that a positive PSU level correlation, ρ_1, will increase the effectively matched proportion and so move the design further from the optimal matched proportion. Thus the achieved efficiency gain in two stage sampling is not as much as for simple random sampling.

The reverse is true for estimating change where the efficiency gains are much greater for both simple random sampling and two stage sampling. However the two stage design can lead to even greater efficiency gains because of the increase in the effectively matched sample proportion. The separate effects of ρ_1 and ρ_2 are of considerable interest and we see that high values of ρ_1 can lead to substantial efficiency gains even when ρ_2 is low. The values of ρ and m have little effect on this situation unless ρ_2 is high and ρ_1 is low.

As in the case of simple random sampling the value of composite estimators when estimating the average level is modest for designs used in practice.

Perhaps the most interesting observation is the modest effect that the cluster sample size and the intra-cluster correlation have on the relative efficiency of optimal estimation, particularly for estimates of level.

8. References

Abraham, T.P., Khosta, R.K., and Kathuria, O.P. (1969). Some Investigations of the Use of Successive Sampling in Pest and Disease Surveys. Journal of the Indian Society of Agricultural Statistics, 21, 43–57.

Jessen, R.J. (1942). Statistical Investigations of a Sample Survey for Obtaining Farm Facts. IOWA Agricultural Experimental Statistics Research Bulletin, 304, 54–59.

Cochran, W.G. (1977). Sampling Techniques (3rd Edition), p. 428. New York: John Wiley.

Kathuria, O.P. (1975). Some Estimators in Two-Stage Sampling in Successive Occasions with Partial Matching at Both Stages. Sankhya, Ser. C, 37, 147–162.

Okafor, F.C. (1987). Comparison of Estimators of Population Total in Two-Stage

Successive Sampling Using Auxiliary Information. Survey Methodology, 13, 109–121.

Patterson, H.D. (1950). Sampling on Successive Occasions with Partial Replacement of Units. Journal of the Royal Statistical Society, Ser. B., 12, 241–255.

Singh, D. (1968). Estimates in Successive Sampling Using a Multistage Design. Journal of the American Statistical Association, 63, 99–112.

Singh, D. and Kathuria, O.P. (1969). On Two-Stage Successive Sampling. The Australian Journal of Statistics, 11, 59–66.

Received January 1992
Revised October 1992