Journal of Official Statistics
Vol. 7. No. 1, 1991, pp. 77–91

Design for Composite Estimation With Changing Survey Frames

Philip J. Hughes[1]

Abstract: We consider the design for composite estimation in rotating surveys with changing frames. These designs require not only the specification of the rotation rate but also the sampling rate for births. We report on and extend previously unpublished results on the optimum design for change in mean and change in total.

We show that designs which sample births optimally may be considerably more efficient than the common design in which births and continuing units are sampled with equal rates. We then consider constrained optimum designs in which the sample rate for births is optimised given a fixed sample rotation rate. The treatment of selected births over many periods is also examined.

Key words: Rotating surveys; composite estimation; global optimum design; constrained optimum design; changing survey frames.

1. Introduction

The frames of many rotating surveys change regularly with the addition of "births" and the deletion of "deaths." However, the standard theory of composite estimation, which was laid down by Patterson (1950) following the initial developments of Jessen (1942), assumes a fixed frame. Fixed frame results may be applied to the changing frame case by forming continuing death and birth sub-populations and applying fixed frame results to the continuing population. Mean estimates may be derived for the death and birth sub-populations.

Konijn (1973), Forsman and Garås (1982), and Babiker (1984) showed this approach required straightforward adjustments to the fixed frame estimation results. For survey design an additional parameter is required. As for the fixed frame case, the proportion of remaining, previously selected units, to be rotated out of sample (the rotation rate) needs specifying. An additional design issue for the changing frame case is the allocation of the new sample between the continuing and birth sub-populations.

Forsman and Garås (1982) considered such design issues for estimating mean and total, and change in mean and total. Births were assumed identifiable prior to selection. They reduced the design problem to the optimisation of the rotation rate by assuming a predetermined sampling rate for births.

Babiker (1984) examined the cases in which births are identifiable either before or after selection. He considered the global optimum design in which the rotation rate

[1] Australian Bureau of Statistics and University of Southampton.
Currently Research Manager with AGB: McNair, 168 Walker St., North Sydney, NSW 2060, Australia.

and the sample rate for the births are jointly optimised for mean and change in mean. To date Babiker's results are unpublished. They are summarised in Appendices A and B.

Babiker did not derive the global optimum design for change in mean under composite estimation but conjectured that it required the maximum possible sample overlap given the birth sample. This is not always true. We extend Babiker's results, deriving sufficient conditions under which the optimum designs for change in mean and change in total have this property.

In rotating surveys with changing frames, births and continuing units are often sampled at the same rate. We call this the *basic design*. We show that with composite estimation the basic design may be significantly less efficient for estimating change than designs in which births are sampled optimally. We also consider *constrained optimum designs* in which the sample fraction for births is optimised given a fixed rotation rate.

Finally we show that in order to avoid estimation problems in later periods, selected births should be sub-sampled prior to their amalgamation with the continuing units.

Many rotating surveys with changing frames are rotating business surveys and we pay particular attention to the application of the results to such surveys. We assume equi-probability sampling with the only supplementary variable being the previous value of the current variable.

2. Notation

Assume the population, P_1, on the first occasion consists of N_1 units. Before the second occasion the death population, P_{1D}, of size N_{1D} units, is deleted and the birth population, P_{2N}, of size N_{2N} units, is added to the population to make P_2 the population, of size N_2, on the second occasion. Let P_{1C} be the population, of size $N_{1C} = N_1 - N_{1D}$, of units in the population on both occasions. The same population, of size $N_{2C} = N_{1C}$, may be labelled P_{2C} on the second occasion, then $N_2 = N_{2C} + N_{2N}$.

We assume the sample size, n, is constant. Let s_1 be the sample on the first occasion. Let s_{1D} be the sub-sample, of size n_{1D}, of units in P_{1D} which are removed from the sample for the second occasion. Let s_{1C} be the sub-sample, of size $n_{1C} = n - n_{1D}$, of units in s_1 not in P_{1D}.

A sub-sample of s_{1C}, s_{1CM}, of size n_{1CM} ($n_{1CM} \leqslant n_{1C}$), is retained for the second occasion. s_{1CU} is the sub-sample, of size $n_{1CU} = n_{1C} - n_{1CM}$, of units in s_{1C} not in s_{1CM}.

The sample s_2 on the second occasion is made up from s_{2CM}, of size $n_{2CM} = n_{1CM}$, (the sub-sample labelled s_{1CM} on the previous occasion), as well as a sub-sample s_{2N}, of size n_{2N}, of units selected from P_{2N} from the births and a sample s_{2CU}, of size n_{2CU}, selected from the units in P_{1C} not in s_{1C}. Denote $s_{2C} = s_{2CU} \cup s_{2CM}$ the sub-sample, of size $n_{2C} = n_{2CU} + n_{2CM}$, of units selected for the second occasion from P_{1C}.

We assume that within the sub-populations samples are selected with equi-probability.

Let $\bar{Y}$, $\bar{y}$ and σ^2 denote the population and sample mean and the population variance of the survey variable, respectively. Subscripts are used to identify sub-groups.

Thus, for example, $\bar{y}_{2CM}$ denotes the sample mean on the second occasion of the units retained from the sample of continuing units on the second occasion. Define

$$k = \sigma_1/\sigma_2; \quad k_C = \sigma_{1C}/\sigma_{2C};$$

$$k_{2N} = \sigma_{2N}/\sigma_{2C}; \quad k_{1D} = \sigma_{1D}/\sigma_{1C};$$

$$W_{1D} = N_{1D}/N_1;$$

$$W_{1C} = N_{1C}/N_1 = 1 - W_{1D};$$

$$W_{2N} = N_{2N}/N_2;$$

$$W_{2C} = N_{2C}/N_2 = 1 - W_{2N};$$
$$g = W_{1C}\sigma_{1C}/W_{2C}\sigma_{2C} = k_C(N_2/N_1);$$
$$\tau_{1C} = n_{1C}/n; \quad \tau_{2C} = n_{2C}/n;$$
$$\tau_{1CM} = n_{1CM}/n; \quad \tau_{1CU} = n_{1CU}/n;$$
$$\tau_{2N} = n_{2N}/n;$$

and $r = n_{1CU}/n_{1C}$ the rotation rate applied to s_{1C}.

Let ρ_C be the correlation between the survey variables on occasions 1 and 2 for units in the common population P_{1C}.

Let $\hat{\mu}_{2,*}$ and $\hat{T}_{2,*}$ be estimators of $\bar{Y}_2$ and Y_2, the population mean and total, respectively, on the second occasion and let $\hat{\delta}_{2,*}$ and $\hat{D}_{2,*}$ be estimators of $\bar{Y}_2 - \bar{Y}_1$ and $Y_2 - Y_1$, the change in population mean and total, respectively, on the two occasions, where $* =$ M or C according to whether the estimator is a mean or composite estimator, respectively.

3. Global Optimum Designs for Change

The global optimum designs require τ_{1CM} and τ_{2N} values which minimise the relevant variance. Appendix A, based on Babiker (1984) and Hughes (1988), sets out the composite estimators and their variances for the estimation of mean, change in mean, total, and change in total.

The global optimum design for mean was derived by Babiker (1984). This design, together with that for the global optimum design for total, is summarised in Appendix B.

We consider the optimum design for change in mean and change in total denoting the optimum values by $\tau_{1CM(OPT)}$ and $\tau_{2N(OPT)}$.

Optimum values must satisfy the constraints $0 < \tau_{1CM} \leqslant \tau_{1C}$ and $\tau_{1CM} + \tau_{2N} \leqslant 1$.

For change in mean the optimum design, ignoring the above constraints, are solutions to

$$\partial V(\hat{\delta}_{2,C})/\partial\tau_{1CM} = 0; \text{ and } \partial V(\hat{\delta}_{2,C})/\partial\tau_{2N} = 0.$$

From (A6) in Appendix A these reduce, respectively, to

$$\tau_{1CM}^2\rho_C^2[\rho_C + \rho_C g^2 - 2g] + 2\tau_{1CM}\rho_C\alpha_C^2[\tau_{1C} + g^2(1 - \tau_{2N})] - \alpha_C^2\{\rho_C\tau_{1C}^2 + \rho_C g^2(1 - \tau_{2N})^2 + 2\tau_{1C}g(1 - \tau_{2N})\} = 0 \qquad (1)$$

and

$$W_{2C}\{\alpha_C - \rho_C^2(1 - r)(g - \rho_C)\}/\pi - W_{2N}k_{2N}/\tau_{2N} = 0 \qquad (2)$$

where $\alpha_C = \sqrt{(1 - \rho_C^2)}$ and $\pi = (1 - \tau_{2N}) \times (1 - \rho_C^2 r) + \tau_{1C}\rho_C^2 r(1 - r)$.

No analytic solution to these equations was presented by Babiker (1984). Using numerical methods Babiker conjectured that the optimum solution exists on the boundary of the solution set – that is $\tau_{1CM(OPT)} = \min(\tau_{1C}, 1 - \tau_{2N(OPT)})$. This solution corresponds to one of maximum sample overlap given the sample allocated to births. We extend Babiker's results by stating and proving a theorem on a sufficient condition for the existence of optimal solutions on the boundary. If these conditions are not met then the optimum solution may exist inside the boundary of the solution set, thus disproving Babiker's conjecture.

Theorem 1

A sufficient condition for the optimal design for composite estimation of change in mean to be on the boundary of the solution set, $\tau_{1CM(OPT)} = \min(\tau_{1C}, 1 - \tau_{2N(OPT)})$ is

$$k_1 \leqslant g \leqslant k_2$$

where $\alpha_1 = 1 - \alpha_C$; $\alpha_2 = 1 + \alpha_C$; $k_1 = \alpha_1/\rho_C$; $k_2 = \alpha_2/\rho_C$; and $g = W_{1C}\sigma_{1C}/W_{2C}\sigma_{2C}$.

Proof: Theorem 1 is proved by showing that for $k_1 \leqslant g \leqslant k_2$ and any feasible value of τ_{2N}, the optimum value of τ_{1CM} exceeds $\min(\tau_{1C}, 1 - \tau_{2N})$. This, in combination with the constraint that $\tau_{1CM} \leqslant \min(\tau_{1C}, 1 - \tau_{2N})$, gives the stronger result that $\tau_{1CM(OPT)} = \min(\tau_{1C}, 1 - \tau_{2N})$ for any feasible τ_{2N} value. The result follows by considering the particular case of $\tau_{2N} = \tau_{2N(OPT)}$.

Case i. $k_1 < g < k_2$.

For this case expression (1) may be treated as a quadratic in τ_{1CM} with fixed τ_{2N}. Solving for τ_{1CM} gives

$$\tau_{1CM} = \{-\alpha_C^2(\tau_{1C} + g^2(1 - \tau_{2N})) \pm |\varepsilon|\alpha_C\}/\{\rho_C^2 g^2 - 2\rho_C g + \rho_C^2\}$$

where $\varepsilon = \tau_{1C}(1 - \rho_C g) - (1 - \tau_{2N}) \times (g^2 - \rho_C g)$.

For $k_1 < g < k_2$ these roots are real and positive and the smaller root corresponds to a minimum variance point. Thus we need only consider the smaller root which, since $\rho_C^2 g^2 - 2\rho_C g + \rho_C^2 < 0$, is given by

$$\tau_{1CM}^{-} = \{-\alpha_C^2(\tau_{1C} + g^2(1 - \tau_{2N})) + |\varepsilon|\alpha_C\}/\{\rho_C^2 g^2 - 2\rho_C g + \rho_C^2\}.$$

This root depends on the sign of ε. We consider the cases of ε positive, negative or zero.

Case i.a $k_1 < g < k_2$ and $\varepsilon > 0$.

$$\tau_{1CM}^{-} = (k_2 - \rho_C)\{k_1\tau_{1C} + (1 - \tau_{2N})g\}/\{\rho_C(k_2 - g)\}$$

and

$$\begin{aligned} g > k_1 \Rightarrow & (k_2 - \rho_C)\{k_1 + g\}/\{\rho_C(k_2 - g)\} > 1 \\ \Rightarrow & (k_2 - \rho_C)\{k_1\tau_{1C} + (1 - \tau_{2N})g\}/\{\rho_C(k_2 - g)\} \\ = & \tau_{1CM}^{-} > \min(\tau_{1C}, 1 - \tau_{2N}) \end{aligned}$$

Case i.b $k_1 < g < k_2$ and $\varepsilon < 0$

$$\tau_{1CM}^{-} = (\rho_C - k_1)\{k_2\tau_{1C} + (1 - \tau_{2N})g\}/\{\rho_C(g - k_1)\}$$

$$\begin{aligned} g < k_2 \Rightarrow & (\rho_C - k_1)\{k_2 + g\}/\{\rho_C(g - k_1)\} > 1 \\ \Rightarrow & (\rho_C - k_1)\{k_2\tau_{1C} + (1 - \tau_{2N})g\}/\{(\rho_C(g - k_1)\} \\ = & \tau_{1CM}^{-} > \min(\tau_{1C}, 1 - \tau_{2N}) \end{aligned}$$

Case i.c $k_1 < g < k_2$ and $\varepsilon = 0$

$$\tau_{1CM}^{-} = \{\alpha_C^2(\tau_{1C} + g^2(1 - \tau_{2N}))\}/\{\rho_C^2 g^2 - 2\rho_C g + \rho_C^2\}$$

and

$$\alpha_C^2(1 + g^2) = \rho_C^2 g^2 - 2\rho_C g + \rho_C^2$$

$$\Rightarrow \tau_{1CM}^{-} \quad \min(\tau_{1C}, 1 - \tau_{2N}).$$

Case ii. $g = k_1$ or $g = k_2$.

For this case, expression (1) is linear in τ_{1CM} with the following solution

$$\begin{aligned} \tau_{1CM}^{-} = & 0.5\{\tau_{1C} + (1 - \tau_{2N})\} \\ & \geqslant \min(\tau_{1C}, 1 - \tau_{2N}). \end{aligned}$$

For all cases $\partial V(\hat{\delta}_{2,C})/\partial\tau_{1CM} < 0$ as τ_{1CM} approaches the optimum value from below. Thus the boundary value $\min(\tau_{1C}, 1 - \tau_{2N})$ is the constrained optimum solution for $k_1 \leqslant g \leqslant k_2$ and for any τ_{2N} since it is the extreme of the feasible range closest to the unconstrained optimum.

The theorem is completed by taking $\tau_{2N} = \tau_{2N(OPT)}$ which gives $\tau_{1CM(OPT)} = \min(\tau_{1C}, 1 - \tau_{2N(OPT)})$.

If $g < k_1$ or $g > k_2$, the existence of a non-boundary optimum solution depends on expression (2). The possible existence of a non-boundary solution for the global optimum design for composite estimation of change in mean then depends on g and ρ_C (since ρ_C is a function of k_1 and k_2). Table 1

Table 1. k_1 and k_2 values for given ρ_C

ρ_C	k_1	k_2
0.50	0.27	3.73
0.70	0.41	2.45
0.80	0.50	2.00
0.90	0.63	1.60
0.95	0.72	1.38
0.98	0.82	1.22
0.99	0.87	1.15

sets out the values of k_1 and k_2 for a range of ρ_C values.

Non-boundary solutions may exist if g lies outside the range (k_1, k_2). The value of g is generally close to unity and thus, from Table 1, non-boundary solutions would be most unusual unless ρ_C is very high, of the order of 0.98 or greater.

Babiker (1984) showed that if the optimum solution is on the boundary then it corresponds to the following design: Define

$$\Gamma = \rho_C W_{2C}^2 \{\rho_C g^2 - 2g + \rho_C\};$$

and

$$\beta = 2W_{2N}k_{2N}W_{2C}\{|1 - \rho_C g| - \alpha_C\},$$

then

i.
If $\tau_{1C} < \min(\Gamma/(\Gamma + \beta),\ W_{2C}|1 - \rho_C g| / \{W_{2N}k_{2N} + W_{2C}|1 - \rho_C g|\})$

$$\tau_{1CM(OPT)} = \tau_{1C}$$

$$\tau_{2N(OPT)} = W_{2N}k_{2N} / \{W_{2N}k_{2N} + W_{2C}|1 - \rho_C g|\}$$

hence

$$\tau_{2CU(OPT)} > 0.$$

ii.
If $\max(\Gamma/(\Gamma + \beta),\ W_{2C}\alpha_C/\{W_{2N}k_{2N} + W_{2C}\alpha_C\}) < \tau_{1C}$

$$\tau_{1CM(OPT)} = W_{2C}\alpha_C/\{W_{2N}k_{2N} + W_{2C}\alpha_C\}$$

$$\tau_{2N(OPT)} = W_{2N}k_{2N}/\{W_{2N}k_{2N} + W_{2C}\alpha_C\}$$

hence

$$\tau_{2CU(OPT)} = 0.$$

iii.
Otherwise

$$\tau_{1CM(OPT)} = \tau_{1C}$$

$$\tau_{2N(OPT)} = 1 - \tau_{1C}$$

hence

$$\tau_{2CU(OPT)} = 0. \tag{3}$$

4. Optimum Design for Composite Estimation of Change in Total

Appendix A shows an equivalence between the variance forms for change in mean (see (A6)) and for change in total (see (A8)) with the terms W_{1C}, W_{2C}, W_{1D}, and W_{2N} in (A6) being replaced by N_{1C}, N_{2C}, N_{1D}, and N_{2N}, respectively, in (A8). In particular the term g in the variance of change in mean is replaced by the term k_C for change in total. Replacing these terms leads to the expression $\Gamma/(\Gamma + \beta)$ in the above conditions being replaced by $N_{1C}(1 - \rho_C)/\{N_{2N}k_{2N} + N_{1C}(1 - \rho_C)\}$. This gives Theorem 2 on the optimum design for the composite estimation of change in total.

Theorem 2

A sufficient condition for the optimal design for composite estimation of change in total to be on the boundary of the solution set is $k_1 \leqslant k_C \leqslant k_2$.

This result follows from Theorem 1 substituting k_C for g.

The same substitution in combination with (3) shows that if the optimum solution is on the boundary then it is given by

i.
If $\tau_{1C} < N_{1C}(1 - \rho_C)/\{N_{2N}k_{2N} + N_{1C}(1 - \rho_C)\}$

$$\tau_{1CM(OPT)} = \tau_{1C}$$

$$\tau_{2N(OPT)} = N_{2N}k_{2N} / \{N_{2N}k_{2N} + N_{1C}(1 - \rho_C)\}$$

hence

$$\tau_{2CU(OPT)} > 0.$$

ii.
If $N_{1C}(1 - \rho_C)/\{N_{2N}k_{2N} + N_{1C}(1 - \rho_C)\} < \tau_{1C} < N_{1C}\alpha_C/\{N_{2N}k_{2N} + N_{1C}\alpha_C\}$

$$\tau_{1CM(OPT)} = \tau_{1C}$$

$$\tau_{2N(OPT)} = 1 - \tau_{1C}$$

hence

$$\tau_{2CU(OPT)} = 0.$$

iii.
If $\tau_{1C} > N_{1C}\alpha_C/\{N_{2N}k_{2N} + N_{1C}\alpha_C\}$

$$\tau_{1CM(OPT)} = N_{1C}\alpha_C/\{N_{2N}k_{2N} + N_{1C}\alpha_C\}$$

$$\tau_{2N(OPT)} = N_{2N}k_{2N}/\{N_{2N}k_{2N} + N_{1C}\alpha_C\}$$

hence

$$\tau_{2CU(OPT)} = 0. \tag{4}$$

Thus the sufficient conditions for a boundary solution for change in total are $k_1 \leqslant (\sigma_{1C}/\sigma_{2C}) \leqslant k_2$, whilst the corresponding sufficient condition for change in mean is given by $k_1 \leqslant (W_{1C}\sigma_{1C}/W_{2C}\sigma_{2C}) \leqslant k_2$. The sufficient conditions for boundary solutions for change in mean and change in total coincide if $N_1 = N_2$.

Hughes (1988) also shows for composite estimation of change in fixed frame surveys that a necessary and sufficient condition for the optimum rotation rate to be zero is $k_1 \leqslant \sigma_{1C}/\sigma_{2C} \leqslant k_2$.

It is also readily shown that if $\sigma_{1C} = \sigma_{2C}$ the optimal design for estimating change in total satisfies $\tau_{1CM(OPT)} = \min(\tau_{1C}, (1 - \tau_{2N(OPT)}))$ and is as given in Theorem 2.

5. Constrained Optimum Designs

The optimum rotation rate arising from the global optimum design may not always be desirable. A number of factors other than variance minimisation may need to be taken into account. Firstly, high rotation rates may be undesirable if the average sampling cost of newly selected units is much greater than that of previously selected units. On the other hand, low rotation rates increase the respondent burden of units and may also be undesirable. This suggests a more general analysis which accounts for these types of costs. The nature of these costs, however, make them difficult to quantify. In practice the rotation rate may be pre-determined subjectively, taking into account the issues of sampling efficiency, cost, and respondent burden.

We thus extend the above analysis and consider constrained optimum designs in which the sample fraction for the births is optimised given the rotation rate.

6. Constrained Optimum Design for Composite Estimation of Mean and Total

Values of τ_{2N} are required which minimise the appropriate variance given τ_{1CM}. For mean, $V(\hat{\mu}_{2,C})$, as given in (A2), is minimised for fixed τ_{1CM}, as follows:

$$\tau_{2N(OPT)} = \min\{1 - \tau_{1CM}, W_{2N}k_{2N}[\tau_{1C}\rho_C^2 r(1 - r) + 1 - \rho_C^2 r]/[(1 - \rho_C^2 r)(W_{2N}k_{2N} + W_{2C})]\}. \tag{5}$$

Replacing W_{1C}, W_{2C}, W_{1D}, and W_{2N} by N_{1C}, N_{2C}, N_{1D}, and N_{2N}, respectively, in (5) gives the result that the constrained optimum design for total and mean are identical.

7. Constrained Optimum Design for Composite Estimation of Change in Mean

Values of τ_{2N} are required which minimise $V(\hat{\delta}_{2,C})$, given τ_{1CM}.

The expression for $\partial V(\hat{\delta}_{2,C})/\partial\tau_{2N} = 0$, obtained from (2) above, reduces to

$$\tau_{2N}^{*} = \frac{W_{2N}k_{2N}\{(1-\rho_C^2 r)+\tau_{1C}(1-r)r\rho_C^2\}}{W_{2N}k_{2N}(1-\rho_C^2 r)+W_{2C}[\rho_C^2(1-g)(1-r)+g(1-\rho_C)(1+r\rho_C)]}. \tag{6}$$

The optimum value of τ_{2N} is constrained to be less than or equal to $(1-\tau_{1CM})$. Thus the constrained design is

$$\tau_{2N(OPT)}^{*} = \min(\tau_{2N}^{*}, 1-\tau_{1CM}). \tag{7}$$

8. Constrained Optimum Design for Composite Estimation of Change in Total

Values of τ_{2N} are required which minimise $V(\hat{D}_{2,C})$, given τ_{1CM}. This follows directly from (6) substituting k_C for g

$$\tau_{2N}^{**} = \frac{W_{2N}k_{2N}\{(1-\rho_C^2 r)+\tau_{1C}(1-r)r\rho_C^2\}}{W_{2N}k_{2N}(1-\rho_C^2 r)+W_{2C}[\rho_C^2(1-k_C)(1-r)+k_C(1-\rho_C)(1+r\rho_C)]} \tag{8}$$

and $\tau_{2N(OPT)}$ is thus given by

$$\tau_{2N(OPT)}^{**} = \min(\tau_{2N}^{**}, 1-\tau_{1CM}). \tag{9}$$

9. Empirical Analysis

We compare the basic and the constrained optimum designs under composite estimation for estimating total and change in total. The large number of possible parameters, estimators, and designs hinders interpretation. To simplify we consider typical values for a rotating business survey: $W_{2N} = W_{1D} = 0.05$, $k_C = 1.0$, $\rho_C = 0.95$, $k_{1D} = 1.0$, and $k_{2N} = 1.5$.

We consider the precision of the mean and composite estimators, relative to the mean estimator at $r = 0$, for estimating total and change in total under the basic design and the constrained optimum design. We consider a range of r values including the global optimum rotation rates for total ($r = 0.76$) and change in total ($r = 0.16$). We assume $\sigma_{1C}/\sigma_1 = \sigma_{2C}/\sigma_2 = 0.8$.

The subscripts BASIC and C_OPT denote the basic and constrained optimum design, respectively. The relative precision values for the mean estimator under the basic design and the constrained optimum design (BD and CD respectively in Table 2), relative to the mean estimator for $r = 0$, are defined

$$V(\hat{T}_{2,M}; r=0)_{BASIC}/V(\hat{T}_{2,M})_{BASIC}$$

and

$$V(\hat{D}_{2,M}; r=0)_{BASIC}/V(\hat{D}_{2,M})_{BASIC}$$

for level and change, respectively.

For the composite estimator relative precision values are $V(\hat{T}_{2,M}; r=0)_{BASIC}/V(\hat{T}_{2,C})_{*}$ and $V(\hat{D}_{2,M}; r=0)_{BASIC}/V(\hat{D}_{2,C})_{*}$ for total and change, respectively, where for BD, $* =$ BASIC and for CD, $* =$ C_OPT.

The relative design precision (DP) values for the composite estimator, relative to the basic design, in Table 2 are defined as

$$\mathrm{DP} = V(\hat{T}_{2,C})_{BASIC}/V(\hat{T}_{2,C})_{C\_OPT}$$

for total and

$$\mathrm{DP} = V(\hat{D}_{2,C})_{BASIC}/V(\hat{D}_{2,C})_{C\_OPT}$$

for change.

Table 2 provides some understanding of the gains in precision from composite estimation and the additional gains available through the respective constrained optimum designs. For all r values the most significant gains, relative to the mean estimators, are achieved through introduction of composite estimation. Further, but less significant, gains are available in moving from the basic to the constrained optimum design. We examine this table in more detail below.

Table 2. Relative precision of estimators of total (T) and change in total (D)

r	Estimating	Mean Estimator	Composite estimator		
		Basic design	Basic design	CD[1]	DP
0.00	Total	1.00	1.47	1.47	1.00
	Change	1.00	5.13	5.13	1.00
0.05	Total	1.00	1.53	1.55	1.01
	Change	0.94	5.03	6.36	1.26
0.10	Total	1.00	1.59	1.61	1.01
	Change	0.88	4.94	6.80	1.38
0.15	Total	1.00	1.64	1.68	1.02
	Change	0.83	4.83	6.93	1.43
0.16[2]	Total	1.00	1.65	1.69	1.02
	Change	0.82	4.81	6.93	1.44
0.20	Total	1.00	1.70	1.74	1.02
	Change	0.79	4.72	6.89	1.46
0.30	Total	1.00	1.80	1.85	1.03
	Change	0.71	4.47	6.55	1.47
0.50	Total	1.00	1.99	2.07	1.04
	Change	0.59	3.84	5.32	1.38
0.70	Total	1.00	2.11	2.21	1.05
	Change	0.51	2.96	3.69	1.25
0.76[3]	Total	1.00	2.12	2.23	1.05
	Change	0.49	2.61	3.12	1.20
0.90	Total	1.00	2.01	2.10	1.04
	Change	0.45	1.61	1.75	1.09

[1] CD = Constrained Optimum Design Given r.
[2] Global optimum rotation rate for composite estimation of Change.
[3] Global optimum rotation rate for composite estimation of total.
Parameters assumed: $W_{2N} = W_{1D} = 0.05$, $k_C = 1.0$, $\rho_C = 0.95$, $k_{1D} = 1.0$, $k_{2N} = 1.5$, and $\sigma_{1C}/\sigma_1 = \sigma_{2C}/\sigma_2 = 0.80$.

Estimation of level – The relative precision of composite estimates of level increase slowly with the rotation rate up to the point of optimum rotation. The constrained optimum design provides almost no additional precision.

Estimation of change – More significant gains in precision are available for change estimation than are available for level estimation. The choice of rotation rate with composite estimation is more sensitive to change than level estimation hence a compromise design would lean towards the optimum rotation rate for change. For this example this optimum rate of 16% is similar to the rotation rates used in many rotating business surveys.

Constrained and Global Optimum Designs – The optimum rate of rotation for composite estimation of change is 16% compared to 0% under a fixed frame design. For composite estimation of total change the optimum rate of rotation is 76% for either case. Thus the optimum design for change is more affected by the changing population structure than the optimum design for level.

This table indicates that the loss of precision in constraining rotation rates to be other than the optimum rate may be appreciable for level or change estimation.

10. Design Precision of the Constrained Optimum Design Relative to the Basic Design for Composite Estimation

We compare the constrained optimum and basic designs for the composite estimation of change in total. A broader range of parameters than considered in the previous section is used. We restrict attention to parameters which are generally observed in rotating business surveys in which the changes to the survey frame are limited, r is not large, and ρ_C is large. We consider a broad range of values of k_{2N}, as its likely value is uncertain, and assume $k_{1D} = k_C = 1$.

Under the basic design if $N_2 > N_1$ some units in s_{1C} will need rotating to maintain a constant sample size. The minimum r value is $(N_2 - N_1)/N_2$ and a greater rate will be required if units are to be rotated out of sample after a fixed number of periods. We restrict consideration of r to $r \geqslant (N_2 - N_1)/N_2$.

The relative design precision values are presented in Table 3. The values of the design precision for the constrained optimum design relative to the basic design for composite estimator presented in Table 3 are given by

$$V(\hat{D}_{2,C})_{BASIC}/V(\hat{D}_{2,C}))_{C\_OPT}.$$

Table 3 indicates that the relative design precision of the constrained optimum design, relative to the basic design, may be significant for the composite estimation of change. This is largely dependent on k_{2N}. For $k_{2N} = 0.5$ the increased precision is only marginal, whilst for $k_{2N} = 1.5$ it may be considerable.

This is because the constrained optimum design is a form of optimum allocation to the continuing and birth sub-populations and produces greater gains as sub-populations are increasingly heterogeneous.

We conclude that if the births have a higher variance than the continuing sub-population (i.e., $k_{2N} > 1$) then constrained optimum designs may produce significant precision gains for the composite estimation of change.

It is difficult to speculate on k_{2N} as it depends on the nature of the births and their stratification. If there is no size stratification, then births may tend to be smaller than the continuing units thus causing k_{2N} to be less than unity. If, however, the effect of size is controlled through size stratification then k_{2N} may be larger than unity since the births are likely to have a more variable performance than units which have been established for some time. One would expect the births to exhibit greater extremes of growth (positive or negative growth) than continuing units. Thus there may be significant gains in constrained optimum designs with size stratification.

The same table may be produced for the estimation of total. For such a table with the same parameters the design precision varies from 1.00 to 1.07. We conclude that constrained optimum designs produce very little gains in precision for estimating level, relative to the basic design.

11. Treatment of Births Over Many Periods

The preceding analysis assumed two survey periods. Over the longer term, it is necessary to consider the treatment of the birth sub-populations for each period. Considerable difficulties may arise if separate birth sub-populations are retained for each period as the total sample would be spread increasingly

Table 3. Design precision of constrained optimum design relative to basic design for composite estimation of change

r	W_{2N}	ρ_C	$V(\hat{D}_{2,C})_{BASIC}/V(\hat{D}_{2,C}))_{C\_OPT}$											
			$k_{2N} = 0.50$				$k_{2N} = 1.00$				$k_{2N} = 1.50$			
			W_{1D}				W_{1D}				W_{1D}			
			0.00	0.05	0.10	0.20	0.00	0.05	0.10	0.20	0.00	0.05	0.10	0.20
0.05	0.05	0.98		1.06	1.05	1.03		1.21	1.19	1.12		1.37	1.38	1.25
		0.95		1.04	1.04	1.02		1.14	1.14	1.09		1.26	1.29	1.21
		0.90		1.02	1.02	1.01		1.09	1.09	1.07		1.18	1.20	1.16
	0.10	0.98			1.05	1.05			1.15	1.17			1.24	1.34
		0.95			1.04	1.04			1.12	1.14			1.20	1.29
		0.90			1.02	1.03			1.09	1.11			1.16	1.23
	0.20	0.98				1.03				1.08				1.12
		0.95				1.03				1.07				1.11
		0.90				1.02				1.06				1.10
0.10	0.05	0.98	1.13	1.08	1.06	1.03	1.38	1.30	1.22	1.12	1.58	1.56	1.45	1.26
		0.95	1.06	1.05	1.04	1.02	1.20	1.19	1.15	1.10	1.36	1.38	1.33	1.21
		0.90	1.03	1.03	1.02	1.01	1.11	1.11	1.10	1.07	1.22	1.24	1.22	1.16
	0.10	0.98		1.08	1.08	1.05		1.20	1.25	1.19		1.38	1.41	1.38
		0.95		1.05	1.05	1.04		1.15	1.19	1.16		1.29	1.33	1.32
		0.90		1.03	1.03	1.03		1.10	1.13	1.12		1.20	1.25	1.25
	0.20	0.98				1.05				1.15				1.23
		0.95				1.04				1.13				1.21
		0.90				1.03				1.11				1.18

Table 3. (*Continued*)

r	W_{2N}	ρ_C	$V(\hat{D}_{2,C})_{BASIC}/V(\hat{D}_{2,C}))_{C\_OPT}$											
			$k_{2N} = 0.50$				$k_{2N} = 1.00$				$k_{2N} = 1.50$			
			W_{1D}				W_{1D}				W_{1D}			
			0.00	0.05	0.10	0.20	0.00	0.05	0.10	0.20	0.00	0.05	0.10	0.20
0.20	0.05	0.98	1.19	1.10	1.06	1.03	1.65	1.37	1.24	1.12	2.15	1.74	1.51	1.27
		0.95	1.08	1.05	1.04	1.02	1.30	1.22	1.16	1.10	1.60	1.46	1.36	1.22
		0.90	1.03	1.02	1.02	1.01	1.15	1.12	1.10	1.06	1.33	1.28	1.23	1.16
	0.10	0.98	1.23	1.15	1.10	1.06	1.55	1.45	1.36	1.22	1.74	1.72	1.64	1.44
		0.95	1.11	1.09	1.07	1.04	1.33	1.30	1.26	1.17	1.53	1.53	1.49	1.36
		0.90	1.05	1.05	1.04	1.03	1.19	1.19	1.17	1.13	1.35	1.37	1.35	1.28
	0.20	0.98	1.00	1.08	1.09	1.08	1.00	1.14	1.22	1.25	1.00	1.17	1.29	1.41
		0.95	1.00	1.05	1.07	1.07	1.00	1.12	1.18	1.21	1.00	1.15	1.26	1.36
		0.90	1.00	1.03	1.04	1.05	1.00	1.09	1.14	1.17	1.00	1.13	1.22	1.30

thinly over these sub-populations and the continuing sub-population. Eventually amalgamation of births and the continuing sub-population would be necessary. It would be possible to leave a number of birth strata for several periods. The judgement over the delay before births are absorbed into the continuing population depends upon how quickly they assume the pattern of other units. We assume that the births for the current period are amalgamated with the continuing population from the previous period to form the continuing population for the current period.

Upon amalgamation, units from the continuing sub-population may have been selected with unequal probabilities. This will necessitate some form of post-stratification or unequal sample weighting to form estimates for this sub-population. As the number of frame updates increases this may become unacceptably complex. The increase in efficiency resulting from departures from the basic design may thus have an off-setting cost in increased estimation complexity. This may be avoided if the births are sampled at the optimum rate when belonging to the birth sub-population and then sub-sampled prior to their amalgamation so that all units in the continuing population have equal selection probability.

Hughes (1988) raises the possibility of a composite estimator which takes account of births being sub-sampled and uses the past responses of sub-sampled births. This requires partitioning the continuing sub-population into a "previously continuing" component of units also continuing on the previous occasion and a "new continuing" component of units which were births on the previous occasion. Separate composite estimators could then be applied to these two components. The increased efficiency of this estimator is unlikely to off-set the increased estimation complexity unless the components of the partitioned continuing sub-population have significantly higher inter-period correlations than the combined continuing sub-population.

12. Sample Selection in Rotating Surveys With Changing Frames

The previous section raises the possibility of the over-sampling of and then subsequent sub-sampling of births. This is usually straightforward although some account must be taken of the particular selection methods used in rotating business surveys. Collocated sampling (Brewer, Early, and Joyce 1972) was developed by the Australian Bureau of Statistics to meet the particular requirements of rotating business surveys. Other similar methods have also been developed. These methods may be readily extended to permit the over-sampling of births and their later sub-sampling whilst not departing from the other requirements for which the methods were originally developed. Hughes (1988) outlines these extensions.

13. Conclusions

We have derived sufficient conditions for the global optimum design for composite estimates of change to be one of maximum rotation given the birth sample. Often the rotation rate will be fixed and we have derived the constrained optimum design.

We have shown that the constrained optimum designs may be significantly more efficient than the basic design. This depends greatly on the relative magnitudes of the population variances of the birth and continuing sub-populations. If the births exhibit a greater variance then the use of the constrained optimum design will provide significant precision gains in many situations with the gains increasing with the sample rotation rate and the inter-period correlation.

As well as providing more efficient estimates of change, designs which over-sample births also permit enhanced analysis of the dynamics of the changes in the population structure by providing more precise estimates from the births.

Appendix A

Composite Estimation With a Changing Frame

Appendix A summarises the optimum estimators and minimum variances derived by Babiker (1984) for the estimation of mean and change in mean. Also included are the extensions of Hughes (1988) to the estimation of total and change in total.

The two-period composite estimator of mean

The optimum two-period composite estimator for mean is given by

$$\hat{\mu}_{2,C} = W_{2N}\bar{y}_{2N} + W_{2C}\{a\bar{y}_{1CU} - a\bar{y}_{1CM} + b\bar{y}_{2CU} + (1 - b)\bar{y}_{2CM}\}$$

where

$$a = \tau_{1C}r(1 - r)\rho_C/\pi;$$

$$b = (1 - \rho_C^2 r)(1 - \tau_{2N} - \tau_{1C}(1 - r))/\pi$$

and

$$\pi = (1 - \tau_{2N})(1 - \rho_C^2 r) + \tau_{1C}\rho_C^2 r(1 - r) \quad \text{(A1)}$$

with variance given by

$$V(\hat{\mu}_{2,C}) = \{[(1 - \rho_C^2 r)W_{2C}^2\sigma_{2C}^2/\pi] + [W_{2N}^2\sigma_{2N}^2/\tau_{2N}]\}/n. \quad \text{(A2)}$$

The two-period composite estimation of total

The estimator for total is given by

$$\hat{T}_{2,C} = N_{2N}\bar{y}_{2N} + N_{2C}\{a\bar{y}_{1CU} - a\bar{y}_{1CM} + b\bar{y}_{2CU} + (1 - b)\bar{y}_{2CM}\} \quad \text{(A3)}$$

where a, b and π are as given in (A1) and variance given by

$$V(\hat{T}_{2,C}) = \{[(1 - \rho_C^2 r)N_{2C}^2\sigma_{2C}^2/\pi] + [N_{2N}^2\sigma_{2N}^2/\tau_{2N}]\}/n. \quad \text{(A4)}$$

The two-period estimator of change in mean

For change in mean the optimum two-period composite estimator is given by

$$\hat{\delta}_{2,C} = c\bar{y}_{2CM} + (W_{2C} - c)\bar{y}_{2CU} + d\bar{y}_{1CM} - (W_{1C} + d)\bar{y}_{1CU} + W_{2N}\bar{y}_{2N} - W_{1D}\bar{y}_{1D}$$

where

$$c = (1 - r)\{\tau_{1C}W_{2C} + W_{1C}\rho_C(1 - \tau_{2N}) - \tau_{1C}W_{1C}(1 - r)\rho_C\}/\pi$$

and

$$d = -(1 - r)\{W_{1C}(1 - \tau_{2N}) + r\rho_C\tau_{1C}W_{2C}\}/\pi \quad \text{(A5)}$$

and variance

$$V(\hat{\delta}_{2,C}) = W_{2C}^2\sigma_{2C}^2\{\tau_{1CM}\rho_C^2 + \tau_{1C}\alpha_C^2 + g^2[\alpha_C^2(1-\tau_{2N}) + \rho_C^2\tau_{1CM}] - 2g\rho_C\tau_{1CM}\}/n\tau_{1C}\pi + [W_{2N}^2\sigma_{2N}^2/n\tau_{2N}] + [W_{1D}^2\sigma_{1D}^2/n\tau_{1D}]. \quad \text{(A6)}$$

The two-period composite estimator of change in total

For change in total the optimum two-period composite estimator is given by

$$\hat{D}_{2,C} = e\bar{y}_{2CM} + (N_{2C} - e)\bar{y}_{2CU} + f\bar{y}_{1CM} - (N_{1C} + f)\bar{y}_{1CU} + N_{2N}\bar{y}_{2N} - N_{1D}\bar{y}_{1D}$$

where

$$e = (1-r)\{\tau_{1C}N_{2C} + N_{1C}\rho_C(1-\tau_{2N}) - \tau_{1C}N_{1C}(1-r)\rho_C\}/\pi$$

and

$$f = -(1-r)\{N_{1C}(1-\tau_{2N}) + r\rho_C\tau_{1C}N_{2C}\}/\pi \quad \text{(A7)}$$

with variance

$$V(\hat{D}_{2,C}) = N_{2C}^2\sigma_{2C}^2\{\tau_{1CM}\rho_C^2 + \tau_{1C}\alpha_C^2 + k_C^2[\alpha_C^2(1-\tau_{2N}) + \rho_C^2\tau_{1CM}] - 2k_C\rho_C\tau_{1CM}\}/n\tau_{1C}\pi + [N_{2N}^2\sigma_{2N}^2/n\tau_{2N}] + [N_{1D}^2\sigma_{1D}^2/n\tau_{1D}]. \quad \text{(A8)}$$

Appendix B

Design for Level

Appendix B summarises the results of Babiker (1984) for the optimum design for the composite estimation of mean. Included for completeness is the extension to the estimation of total.

Optimum design for composite estimation of mean

The global optimum design for the composite estimation of mean requires τ_{1CM} and τ_{2N} which minimise the variance in (A2) of Appendix A and which satisfy $0 < \tau_{1CM} \leqslant \tau_{1C}$, and $\tau_{1CM} + \tau_{2N} \leqslant 1$. The optimum values, $\tau_{1CM(OPT)}$ and $\tau_{2N(OPT)}$, are given by

i, if $\tau_{1C} \leqslant W_{2C}\alpha_2/(W_{2C}\alpha_C + W_{2N}k_{2N})$

$$\tau_{1CM(OPT)} = \tau_{1C}\alpha_C/\alpha_2$$

$$\tau_{2N(OPT)} = W_{2N}k_{2N}(\alpha_2 + \tau_{1C}\alpha_1)/\{\alpha_2(W_{2C} + W_{2N}k_{2N})\}$$

or ii, if $\tau_{1C} > W_{2C}\alpha_2/(W_{2C}\alpha_C + W_{2N}k_{2N})$

$$\tau_{1CM(OPT)} = W_{2C}\alpha_C/(W_{2C}\alpha_C + W_{2N}k_{2N})$$

$$\tau_{2N(OPT)} = W_{2N}k_{2N}/(W_{2C}\alpha_C + W_{2N}k_{2N}) \quad \text{(B1)}$$

where $\alpha_C = \sqrt{(1-\rho_C^2)}$; $\alpha_1 = 1 - \alpha_C$; and $\alpha_2 = 1 + \alpha_C$.

Optimum design for composite estimation of total

The variance expressions for the composite estimators of total and mean have the same form. From A2 and A4 the terms W_{1C}, W_{2C}, W_{1D}, and W_{2N} in the variance of the mean estimator are replaced by N_{1C}, N_{2C}, N_{1D}, and N_{2N}, respectively, in the variance of the total estimator. From this it is readily shown that the optimum designs for total and mean are identical.

14. References

Babiker, A.I. (1984). Sampling on Successive Occasions from a Changing Population. Ph.D Thesis, University of Southampton, U.K.

Brewer, K.R.W., Early, L.J., and Joyce, S.F. (1972). Selecting Several Samples from a Single Population. Australian Journal of Statistics, 14, 231–239.

Forsman, G. and Garås, T. (1982). Optimal Estimation of Changes in Sample Surveys. Proceedings of the Survey Research Methods Section, American Statistical Association, 417–420.

Hughes, P.J. (1988). Design and Estimation Issues for Rotating Business Surveys. Ph.D Thesis, University of Southampton, U.K.

Jessen, R.J. (1942). Statistical Investigation of a Farm Survey for Obtaining Farm Facts. Iowa Agricultural Experimental Station Research Bulletin, 403, 54–59.

Konijn, H.S. (1973). Statistical Theory of Sample Survey Design and Analysis. London: North-Holland.

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

Received July 1988
Revised October 1990