Estimation of Parameters in a Time Series with Rotation Sampling

by

Ing-tzer Wey

I. Introduction

The technique of rotation sampling on successive occasions is widely employed in many scientific research works to estimate parameters of a population on consecutive occasions in order to measure time-trends as well as the current parameters in a time series. If there exists a relationship between the value of an element in the population at time t and the changed value of the same element at the succeeding time t′, then it is possible to use the information contained in earlier samples to improve the current estimate of the population parameters. In order to use the earlier sample information, one must carry out the sampling in such a way that the two samples drawn at successive times t and t′ have some elements in common. For example, on the h-th occasion we may have parts of the sample that are matched with the (h-1)th occasion, parts that are matched with both the (h-1)th and the (h-2)th occasions, and so on. The optimum replacement policy for two occasions has been investigated by Jessen (1942) and the general problem of replacement for more than two occasions has been examined by Yates (1960), Patterson (1950), Eckler (1955), and several others. However, the theory has been almost exclusively confined to infinite populations. A good summary of these papers is given by Cochran (1963). Rao and Graham (1964) have developed a unified finite population theory for composite estimators of both the current level and change in level between consecutive occasions when a rotation sample design is used.

Recently, Pathak and Rao (1967) have improved an estimator of the population mean in the following sampling scheme:

(1.a) Select a sample of n units on the first occasion by simple random sampling without replacement.

(1.b) On the second occasion, retain a sample of m units drawn with equal probability and without replacement from the sample obtained in (1.a) and select independently a sample of n - m units by simple random sampling without replacement from the whole population of size N.

For estimating the population mean on the second occasion, the customary estimator (Jessen and Cochran) is given by

$$\overline{T}\ell=\phi\bar{y}_{2u}+(1-\phi)\ \bar{y}\ell_{d}$$

where $\bar{y}\ell_{d}=\bar{y}_{2m}+b\ (\bar{y}_{1n}-\bar{y}_{1m})$ is the regression estimator on the matched portion.

$\bar{y}_{hu}$ is the sample mean of the unmatched portion on occasion h $(h=1,2)$.

$\bar{y}_{hm}$ is the sample mean of the matched portion on occasion h $(h=1,2)$.

$\bar{y}_{1n}$ is the sample mean of the whole sample on the first occasion.

ϕ is determined such that the variance of $\overline{T}\ell$ is minimized.

The improved estimator derived by Pathak and Rao is given by

$$\overline{T}\ell^{*}=\phi^{*}\left(\frac{m\bar{y}_{2m}+m'\bar{y}_{2m'}}{m+m'}\right)+(1-\phi^{*})\ \bar{y}\ell_{d} \tag{2}$$

where m' is the number of units in the independent sample of size n–m selected on the second occasion which are not present in the n units selected on the first occasion and ϕ^{*} is so chosen as to minimize the variance of $\overline{T}\ell^{*}$. They have shown that the customary estimator $\overline{T}\ell$ is always less efficient than the estimator $\overline{T}\ell$.

Des Raj (1965) has considered the other sampling scheme over two occasions where sampling with probability proportional to an estimate of size (ppes sampling) is used on both the occasions:

(2.a) Select a sample of n units on the first occasion by ppes sampling with replacement.

(2.b) On the second occasion, a simple random sample of size m is selected without replacement from the sample obtained in (2.a) and an independent sample of n–m units is selected by ppes sampling with replacement from the whole population.

As an estimator of the population total on the second occasion Des Raj has considered the following estimator:

$$\hat{Y}_{ppes}=w\ \hat{Y}_{2u}+(1-w)\ \hat{Y}_{2d} \tag{3}$$

where $\hat{Y}_{2u}=\frac{1}{n-m}\Sigma\frac{y_{2i}}{p_{i}}$

is the estimator of the population total based on the unmatched sample on the second occasion,

$$\hat{Y}_{2d} = (\hat{Y}_{2m} - \hat{Y}_{1m}) + \hat{Y}_{1n}$$

is the difference estimator of the population total on the second occasion based on the matched portion,

$$\hat{Y}_{hm} = \frac{1}{m} \Sigma \frac{y_{hi}}{p_i} \qquad h = 1, 2$$

is the estimator of the population total based on the matched sample on occasion h, and

$$\hat{Y}_{1n} = \frac{1}{n} \Sigma \frac{y_{1i}}{p_i}$$

is the estimator of the population total based on the whole sample, on the first occasion, and w is so chosen as to minimize the variance of $\hat{Y}_{pps}$.

The main object of this paper is to derive estimators of the population mean and total with the other sampling schemes which are confined to two occasions only. It is also proposed to work out the expected gain in efficiency.

II. Estimators for Equal Probability Sampling Scheme

Avadhani and Sukhatme (1969) have considered the following sampling scheme:

(3.a) Select a sample of n units on the first occasion by simple random sampling without replacement.

(3.b) On the second occasion, retain a sample of m units drawn with equal probability and without replacement from the sample obtained in (3.a) and take a fresh sample of m′ units selected independently from the remaining N–n units by simple random sampling without replacement so as to get a sample of predetermined size $n = m + m'$.

For estimating the mean of the population of size N on the second occasion, we construct the following estimator for the sampling scheme (3.a.b) which is similar to the estimator $\overline{T}\ell$ for the sampling scheme (1.a.b).

$$\hat{\bar{Y}}\ell = a\bar{y}_{2m}' + (1-a)\bar{y}\ell_d \tag{4}$$

where $\bar{y}\ell_d = \bar{y}_{2m} + b\ (\bar{y}_{1n} - \bar{y}_{1m})$

and the constant a is so determimined that the variance of $\hat{\bar{Y}}\ell$ is minimum.

Further, let

$$\begin{aligned} \hat{\bar{Y}}\ell^* &= a^* \left(\frac{m\bar{y}_{2m} + m'\bar{y}_{2m}'}{m+m'} \right) + (1-a^*)\ \bar{y}\ell_d \\ &= a^*\bar{y}_{2n} + (1-a^*)\ \bar{y}\ell_d \end{aligned} \tag{5}$$

where the constant a^* is so chosen as to minimize the variance of $\hat{\bar{Y}}_{\ell}^*$. It is clear that the estimator $\hat{\bar{Y}}_{\ell}^*$ corresponding to the sampling scheme (3.a.b) is similar to the estimator $\bar{T}_{\ell}^*$ corresponding to the sampling scheme (1.a.b).

For the comparison of the estimators $\hat{\bar{Y}}_{\ell}$ and $\hat{\bar{Y}}_{\ell}^*$, consider first the approximate variance of $\hat{\bar{Y}}_{\ell}^*$. Assume that the sample size is large enough to make the bias in $\bar{y}_{\ell d}$ negligible. For simplicity, the finite population correction terms will be ignored. Then it can be easily seen that the variance of $\hat{\bar{Y}}_{\ell}^*$ for optimal choice of a^* is given by

$$\mathrm{Var}\,(\hat{\bar{Y}}_{\ell}^*) = \frac{\mathrm{Var}(\bar{y}_{2n})\mathrm{Var}(\bar{y}_{\ell d})-[\mathrm{Cov}(\bar{y}_{2n}, \bar{y}_{\ell d})]^2}{\mathrm{Var}(\bar{y}_{2n})+\mathrm{Var}(\bar{y}_{\ell d})-2\mathrm{Cov}(\bar{y}_{2n}, \bar{y}_{\ell d})} \tag{6}$$

It is obvious that

$$\mathrm{Var}(\bar{y}_{2n}) = \frac{S_y^2}{n} \tag{7}$$

It can be shown (see Cochran, p.336) that

$$\mathrm{Var}(\bar{y}_{\ell d}) = \frac{1}{m} S_y^2(1-\rho^2)+\frac{1}{n}\rho^2 S_y^2 \tag{8}$$

We observe that

$$\mathrm{Cov}(\bar{y}_{2n}, \bar{y}_{\ell d}) = \frac{m}{n}\mathrm{Cov}(\bar{y}_{2m}, \bar{y}_{\ell d}) + \frac{m'}{n}\mathrm{Cov}(\bar{y}_{2m'}, \bar{y}_{\ell d}) \tag{9}$$

Now, it can be shown that

$$\begin{aligned}\mathrm{Cov}(\bar{y}_{2m}, \bar{y}_{\ell d}) &\doteq \mathrm{Var}(\bar{y}_{2m})+B\mathrm{Cov}(\bar{y}_{2m}, \bar{y}_{1n})-B\mathrm{Cov}(\bar{y}_{2m}, \bar{y}_{1m})\\ &= \frac{1}{m} S_y^2+\frac{1}{n}\rho^2 S_y^2-\frac{1}{m}\rho^2 S_y^2\\ &= \frac{1}{m} S_y^2(1-\rho^2)+\frac{1}{n}\rho^2 S_y^2 \qquad (10)\end{aligned}$$

and

$$\begin{aligned}\mathrm{Cov}(\bar{y}_{2m'}, \bar{y}_{\ell d}) &\doteq \mathrm{Cov}(\bar{y}_{2m'}, \bar{y}_{2m})+B\mathrm{Cov}(\bar{y}_{2m'}, \bar{y}_{1n})-B\mathrm{Cov}(\bar{y}_{2m'}, \bar{y}_{1m})\\ &= -\frac{1}{N} S_y^2-\frac{1}{N}\rho^2 S_y^2+\frac{1}{N}\rho^2 S_y^2\\ &= -\frac{1}{N} S_y^2 \text{ (ignored)} \qquad (11)\end{aligned}$$

Thus

$$\mathrm{Cov}(\bar{y}_{2n}, \bar{y}_{\ell d}) \doteq \frac{1}{n} S_y^2(1-\rho^2)+\frac{1}{n^2}\rho^2 S_y^2 \tag{12}$$

Putting $m=n(1-k)$, $m'=nk$ and substituting from (7), (8), and (12) into (6), we have after simplification

$$\mathrm{Var}(\hat{\bar{Y}}_{\ell}^{*}) \doteq \frac{S_y^2}{n}\frac{(1-k\rho^2)(1+\rho^2-k\rho^2)}{1+\rho^2-2k\rho^2} \tag{13}$$

Consider now the estimator $\hat{\bar{Y}}_{\ell}$. Then proceeding in similar fashion, it can be shown that the approximate variance of $\hat{\bar{Y}}_{\ell}$ for optimal choice of a is given by

$$\mathrm{Var}(\hat{\bar{Y}}_{\ell}) \doteq \frac{S_y^2}{n}\frac{1-k\rho^2}{1-k^2\rho^2} \tag{14}$$

The difference between the variances of $\hat{\bar{Y}}_{\ell}^{*}$ and $\hat{\bar{Y}}_{\ell}$ is

$$\mathrm{Var}(\hat{\bar{Y}}_{\ell}^{*})-\mathrm{Var}(\hat{\bar{Y}}_{\ell}) = \frac{S_y^2}{n}\frac{k\rho^2(1-k)(1-k\rho^2)^2}{(1-k^2\rho^2)(1+\rho^2-2k\rho^2)} \tag{15}$$

which is always positive unless $\rho=0$. We have thus proved the following result.

Theorem 1. In the sampling scheme (3.a.b), the estimator $\hat{\bar{Y}}_{\ell}$ which is similar to the estimator $\bar{T}_{\ell}$ in the sampling scheme (1.a.b) is always more efficient than the estimator $\hat{\bar{Y}}_{\ell}^{*}$ which is similar to the estimator $\bar{T}_{\ell}^{*}$ derived by Pathak and Rao in the sampling scheme (1.a.b), provided $0<k<1$ and $\rho\neq 0$.

Let us further consider the finite population model used by Avadhani and Sukhatme (1969) for comparing the two different estimators $\hat{\bar{Y}}_{\ell}$ and $\hat{\bar{Y}}_{\ell}^{*}$. The finite population model is defined as follows. Let

$$y_i = A+Bx_i+e_i, \quad i=1,2,\cdots\cdots,N \tag{16}$$

where $\sum_{i=1}^{N} e_i = 0 = \sum_{i=1}^{N} e_i x_i$

and $e_i^2 = cx_i^{\frac{1}{2}g}$ with $c>0$ and $0\leq g\leq 2$.

Since $e_i/x_i = \pm\sqrt{cx_i^{\frac{1}{2}g-1}}$, it is clear that for very small values of x_i, $|e_i/x_i|>1$ which is impossible in view of (16). They therefore assume that for very small values of x_i, e_i can assume only positive values or that x_i cannot be so small that $|e_i/x_i|>1$.

For comparing the two different estimators under this finite population model, it is shown that ignoring the finite population correction terms, the variances of $\hat{\bar{Y}}_{\ell}^{*}$ and $\hat{\bar{Y}}_{\ell}$ for optimal choice of a* and a under the finite population model are just the same as those obtained in (13) and (14) under the large sample theory.

The optimal value of the replacement fraction k that minimizes $\mathrm{Var}(\hat{\bar{Y}}_{\ell})$ can be shown to be given by

$$\text{Opt. } k = [1+\sqrt{1-\rho^2}]^{-1} \tag{17}$$

Then, the minimum variance of $\bar{Y}\ell$ is

$$\mathrm{Var}(\bar{Y}\ell)\min. = \frac{S_y^2}{n}\,\frac{1+\sqrt{1-\rho^2}}{2} \tag{18}$$

If $A=0$ in the finite population model (16), Avadhani and Sukhatme have proposed the following ratio estimator of the population mean with the sampling scheme (3.a.b):

$$\bar{T}_R = a\,\bar{y}_{2m} + (1-a)\,\bar{y}_{Rd}$$

where $\bar{y}_{Rd} = \hat{R}_{2m}\,\bar{y}_{1n}$, $\hat{R}_{2m} = \bar{y}_{2m}/\bar{y}_{1m}$

and they have obtained the minimum variance of $\bar{T}_R$ for optimal choice of a and k which is the same as the minimum variance of $\bar{Y}\ell$ given by (18). This shows that the regression estimator $\bar{Y}\ell$ and the ratio estimator $\bar{T}_R$ are equally efficient with $A=0$ in the finite population model.

Further, it should be noted that the estimator $\bar{Y}\ell$ with the sampling scheme (3.a.b) is similar to the estimator $\bar{T}\ell$ with the sampling scheme (1.a.b) and both estimators are equally efficient. However, in the sampling scheme (3.a.b), the estimator $\bar{Y}\ell$ is more efficient than the estimator $\bar{Y}\ell^*$ while in the sampling scheme (1.a.b), the estimator $\bar{T}\ell$ is less efficient than the estimator $\bar{T}\ell^*$ which is similar to the estimator $\bar{Y}\ell^*$.

III. Estimators for Unequal Probability Sampling Scheme

We now consider the sampling scheme (2.a;b) stated in section I. On the first occasion a sample of n units is selected with probabilities p_i, $\sum_{i=1}^{N} p_i = 1$ and with replacement. For estimating the population total Y_1 we have

$$\hat{Y}_{1n} = \frac{1}{n}\sum_{i=1}^{n} y_{1i}/p_i$$

$$\mathrm{Var}(\hat{Y}_{1n}) = \frac{1}{n}\sum_{i=1}^{N} p_i\left(\frac{y_{1i}}{p_i} - Y_1\right)^2 = \frac{V_1^2}{n} \tag{19}$$

where y_{1i} is the value of the i-th unit on the first occasion. On the second occasion a simple random sample of m units is selected without replacement from the sample obtained on the first occasion and an independent sample of $u=n-m$ units is selected by ppes sampling with replacement from the whole population. Based on the matched part an unbiased estimator of the population total Y_2 is given by

$$\hat{Y}_m = \hat{Y}_{2m} + b_o(\hat{Y}_{1n} - \hat{Y}_{1m}) \tag{20}$$

where $\hat{Y}_{hm} = \frac{1}{m}\sum_{i=1}^{m} y_{hi}/p_i$ is the estimator of the population total Y_h on the h-th occasion and b_o is a preassigned constant.

Using theorems on conditional expectations and variances, we have

$$\mathrm{Var}(\hat{Y}_m) = \frac{V_2^2}{n} + \left(\frac{1}{m} - \frac{1}{n}\right)(V_2^2 + b_0^2V_1^2 - 2b_0\delta V_1V_2) \tag{21}$$

where $V_h^2 = \sum_{i=1}^{N} p_i \left(\frac{y_{hi}}{p_i} - Y_h\right)^2, \; h=1,2$

and $\delta = \sum_{i=1}^{N} p_i \left(\frac{y_{1i}}{p_i} - Y_1\right)\left(\frac{y_{2i}}{p_i} - Y_2\right)/V_1V_2$

is the correlation coefficient between y_{1i}/p_i and y_{2i}/p_i.

It can be easily seen that the value of b_o which minimizes the variance of $\hat{Y}_m$ is

$$b_o = B = \sum_{i=1}^{N} p_i \left(\frac{y_{1i}}{p_i} - Y_1\right)\left(\frac{y_{2i}}{p_i} - Y_2\right)/V_1^2 \tag{23}$$

Then, the minimum variance of Y_m is given by

$$\mathrm{Var}(\hat{Y}_m)_{min.} = \frac{1}{m}(1-\delta^2)V_2^2 + \frac{1}{n}\delta^2V_2^2 \tag{24}$$

which is similar to the variance of $\bar{y}\ell_d$ given by (8) with equal probability sampling scheme.

If b_o must be computed from the sample, an effective estimator is likely to be of the form

$$\hat{B} = \frac{\sum_{i=1}^{m}\left(\frac{y_{1i}}{p_i} - \hat{Y}_{1m}\right)\left(\frac{y_{2i}}{p_i} - \hat{Y}_{2m}\right)}{\sum_{i=1}^{m}\left(\frac{y_{1i}}{p_i} - \hat{Y}_{1m}\right)^2} \tag{25}$$

Then, it can be shown that the bias in B is of order $n^{-\frac{1}{2}}$. Now, the population total Y_2 is estimated by

$$\hat{Y}_m{*} = \hat{Y}_{2m} + \hat{B}\,(\hat{Y}_{1n} - \hat{Y}_{1m}) \tag{26}$$

If the sample size is large enough to make the bias in $\hat{B}$ negligible, then the approximate variance of $\hat{Y}_m*$ can be shown to be the same as the minimum variance of $\hat{Y}_m$ given by (24).

Further, based on the unmatched part an unbiased estimator of the population total Y_2 is

$$\hat{Y}_{2u} = \frac{1}{n-m} \Sigma \frac{y_{2i}}{p_i}$$

$$Var(\hat{Y}_{2u}) = V_2^2/(n-m)$$

Using weights W_m and W_u as the inverses of the variances, the best combined estimator of Y_2 is found by weighting the two estimators $\hat{Y}_m^*$ and $\hat{Y}_{2u}$ as follows:

$$\hat{Y}^*_{ppes} = W_m \hat{Y}_m^* + W_u \hat{Y}_{2u} \tag{27}$$

Putting $m = n(1-k)$ and $u = nk$, the variance of $\hat{Y}^*_{ppes}$ is obtained after simplification as

$$Var(\hat{Y}^*_{ppes}) = \frac{V_2^2}{n} \frac{1-k\delta^2}{1-k^2\delta^2} \tag{28}$$

The optimal value of the replacement fraction k that minimizes can be shown to be given by

$$\text{Opt. } k = (1+\sqrt{1-\delta^2})^{-1} \tag{29}$$

Then, the minimum variance of $\hat{Y}^*_{ppes}$ is

$$Var(\hat{Y}^*_{ppes})_{min.} = \frac{V_2^2}{n} \frac{1+\sqrt{1-\delta^2}}{2} \tag{30}$$

which is similar to the minimum variance of $\hat{Y}\ell$ given by (18) with equal probability sampling scheme.

Under the assumption that $V_1^2 = V_2^2$, the minimum variance of $\hat{Y}_{ppes}$ derived by Des Raj (1965) is given by

$$Var(\hat{Y}_{ppes})_{min.} = \frac{V_2^2}{n} \frac{1+\sqrt{2(1-\delta)}}{2} \tag{31}$$

The difference between the variances of $\hat{Y}_{ppes}$ and $\hat{Y}^*_{ppes}$ is

$$Var(\hat{Y}_{ppes})_{min.} - Var(\hat{Y}^*_{ppes})_{min.} = \frac{V_2^2}{2n}(\sqrt{2(1-\delta)} - \sqrt{1-\delta^2}) \tag{32}$$

which is always positive unless $\delta = 1$ in this case where both estimators are equally efficient. We have thus proved the following theorem.

Theorem 2. Under the large sample theory and the assumption that the variance of y_i/p_i in ppes sampling (2.a.b) is the same on both occasions, the estimator $\hat{Y}^*_{ppes}$ proposed in this study is always more efficient than the estimator $\hat{Y}_{ppes}$ derived by Des Raj, provided $\delta \neq 1$.

IV. Summary

For estimating the mean of a population on the second of two occasions by equal probability sampling without replacement, the estimator proposed in this study is more efficient than the one similar to that of Pathak and Rao(1967). The variance of this estimator under the large sample theory is the same as it under the finite population model used by Avadhani and Sukhatme (1969), provided that the finite population correction terms are negligible.

For estimating the population total on the second of two successive occasions by sampling with unequal probability and with replacement, the estimator proposed in this study is more efficient than the estimator given by Des Raj (1965) under the large sample theory and the assumption that the variance of y_i/p_i in ppes sampling is the same on both occasions.

References

1. Avadhani, M. S. and Sukhatme, B. V. (1969): A comparison of two sampling procedures with an application to successive sampling. Journal of Royal Statistical Society, Series C, Vol. 19: 251-259.
2. Cochran, W. G. (1963): Sampling Techniques. Second Edition. John Wiley and Sons, Inc., New York. pp341-353.
3. Eckler, A.R. (1955): Rotation sampling. Annals of Mathematical Statistics, 26: 664-685.
4. Jessen, R. J. (1942): Statistical investigation of a sample survey for obtaining farm facts. Iowa Agricultural Experiment Station Research Bulletin No. 304.
5. Pathak, P. K. and Rao, T. J. (1967): Inadmissibility of customary estimators in sampling over two occasions. Sankhyā, Series A, Vol. 29, Part 1, pp49-54.
6. Patterson, H. D. (1950): Sampling on successive occasions with partial replacement of units. Journal of Royal Statistical Society, Series B, 12: 241-255.
7. Raj, Des (1965): On sampling over two occasions with probability proportionate to size. Annals of Mathematical Statistics, 36: 327-330.
8. Rao, J. N. K. and Graham, J. E. (1964): Rotation designs for sampling on repeated occasions. Journal of American Statistical Association, 59: 492-509.
9. Yates, F. (1960): Sampling Methods for Censuses and Surveys. Third Edition. Charles Griffin and Co., London.