Rotation Sampling in Time Series

Hong Nai Park*

1. Introduction

In sampling surveys carried out on a series of successive occasions in time, there are two problems to be considered: One is the scheme for partial retention of units in the sample after each occasion and the other is an estimation problem. Many studies concerned with these problems have been done. Most of the studies on multi-stage sampling on successive occasions are devoted to a composite estimator or a regression type estimator based on a specific rotation scheme such as U. S. C. P. S. redesign.

In this paper, the following new design is proposed and the properties of a linear estimator based on this proposed design will be investigated:

(1) The first-stage units are selected with probability proportional to size,

(2) The same P_n first-stage units from the first occasion are retained in the sample for every occasion; the remaining Q_n $(P+Q=1)$ units in the sample are replaced on every occasion, where the sample size n is assumed to be constant on all occasions,

(3) The second-stage units of size m_i are drawn from the ith first-stage units with equal probability and without replacement at each draw,

(4) The first ordered um_i $(0 \leq u \leq 1)$ second-stage units within the matched first-stage unit are discarded on each occasion, and the remaining $(1-u)m_i$ units which are matched with the previous occasion are supplemented by the next new units of um_i.

2. A Linear Estimator and Its Properties

2-1. Derivation of the Optimum Estimator

According to the sampling scheme as in 1, the linear estimator of the population

* Professor of Statistics, College of Agriculture, Seoul National University.

mean $\overline{\overline{Y}}_t$ on occasion t, denoted by $\overline{Y}_{t\cdot}$, is defined as a linear combination of sample means which are based on disjoint sets on each occasion $(t-k)$.

$$\overline{Y}^{t_0}=\sum_{k=0}^{t-1}a_k\overline{Y}_{(t-k)1}+\sum b_k\overline{Y}_{(t-k)2}, \tag{1}$$

where

$$\overline{Y}_{(t-k)1}=\frac{1}{M}\ \frac{1}{P_n}\sum_{i=1}^{P_n}\frac{M_i}{Z_i}\sum_{j=1}^{m_i}\frac{y_{(t-k)ij}}{m_i}$$

=sample mean per subunit $(t-k)$ of the matched portion.

$$\overline{Y}_{(t-k)2}=\frac{1}{M}\ \frac{1}{Q_n}\sum\frac{M_i}{Z_i}\sum\frac{y_{(t-k)ij}}{m_i}$$

=sample mean per subunit on occasion $(t-k)$ of the unmatched portion.

$a_0+b_0=1 \quad a_k+b_k=0$ for all $k=1, 2, \cdots\cdots, t-1$.

$M=\sum M_i$

Hence, its variance is

$$V(\overline{Y}_t)=\frac{V_t}{Q_n}+\sum_{k=0}^{t-1}a^2{}_k\frac{V_{t-k}}{nPQ}-2a_0\frac{V_{t+2}}{Q_n}\sum_{k(<l)=0}^{t-2}a_ka_l\frac{V_{(t-k)(t-1)}}{P_n} \tag{2}$$

where

$$V_{(t-k)(t-1)}=\sum_{i=1}^{N}\frac{Z_i}{M^2}\left(\frac{Y_{(t-k)i}}{Z_i}-Y_{(t-k)}\right)\left(\frac{Y_{(t-1)i}}{Z_i}-Y_{(t-1)}\right)$$
$$+\frac{1}{M^2}\sum\frac{M_i^2}{Z_i}\times\frac{S_{12i}}{m_i}(1-f_i-u_{(t-k)(t-1)})$$

$u_{(t-k)(t-1)}$=unmatched fraction of the second-stage unit on occasion $(t-k)$ and $(t-1)$.

The optimum value of a_k for given value of Q and r is found as follows;

$$a_{kw}=(-1)^k\frac{1}{\sqrt{V_{t-k}V_t}}\quad\frac{R_{t(t-k)}}{R_t}PV_t \text{ for all } k=0, \cdots\cdots, t-1.$$

and
$$r_{(t-k)(t-1)}=\frac{V_{(t-k)(t-1)}}{\sqrt{V_{t-k}V_{t-1}}}$$

where R_t is the correlation matrix of A_t

$$A_t=\begin{pmatrix} V_1 & QV_{12}\cdots\cdots\cdots\cdots\cdots\cdots QV_{1t} \\ QV_{12} & V_2\cdots\cdots\cdots\cdots\cdots\cdots\cdots QV_{2t} \\ \vdots & \vdots \qquad\qquad\qquad\qquad \vdots \\ QV_{1t} & QV_{2t}\cdots\cdots\cdots\cdots\cdots\cdots V_t \end{pmatrix}$$

$R_{t(t-k)}$ is the minor determinant of t-th row and $(t-k)$-th column element in R_t.

The variance of the optimum estimator $\overline{Y}_{t.w}$ of $\overline{\overline{Y}}_t$ is

$$V(\overline{Y}_{t.w})=\frac{V_t}{Q_n}-\frac{V_t}{nPQ}P^2C_t \tag{3}$$

$$C_t=\frac{1}{|R|^2_t}[R^2_{tt}-\sum R^2_{t(t-k)}+(-1)^{k+1-1}2\sum\sum Qr_{(t-k)(t-1)}R_{t(t-k)}R_{t(t-1)}].$$

2-2. Relationship of Percent Loss of Precision to the Number of Previous Occasions Included in Estimator

Since the estimator $\bar{Y}_{(t-k)}$ makes use of the information from k previous occasions, the sequence of variances of $\bar{Y}_{(t-k)}$ is asymptotically decreasing as k increases. Hence it is desired to determine the relationship of the loss of precision to the number of previous occasions included in the estimator.

Under the assumptions $r_{(t-k)(t-k+1)}=r$, $r_{(t-k)(t-k+2)}=r^2, \cdots\cdots$, for all k, Table 1 provides the percent loss of an estimator $\bar{Y}_{t,\infty}$ when k previous occasions are used in estimating $\bar{\bar{Y}}_t$.

Table 1 **Percent Loss of Precision of $\bar{Y}_t$**

k \ r	Q=.2: .5	.6	.7	.8	.9	Q=.6: .5	.6	.7	.8	.9
1	2.73	4.73	8.18	14.56	27.75	2.63	3.93	5.92	9.29	18.51
2	.13	.44	1.31	3.58	9.65	.01	.07	.27	1.06	5.05
3	.01	.08	.32	1.10	3.58	.00	.00	.02	.18	1.22
	Q=.3					**Q=.7**				
1	3.27	5.55	9.48	16.90	33.45	2.26	3.46	5.38	9.24	20.69
2	.10	.37	1.14	3.35	9.09	.00	.02	.10	.48	2.89
3	.01	.05	.24	.91	3.43	.00	.00	.00	.05	.57
	Q=.4					**Q=.8**				
1	3.41	5.67	9.53	16.98	34.82	1.57	2.31	3.42	5.59	12.59
2	.07	.25	.84	2.66	9.08	.00	.00	.02	.13	1.13
3	.00	.03	.14	.63	2.81	.00	.00	.00	.00	.16
	Q=.5					**Q=.9**				
1	3.24	5.26	8.68	15.41	32.65	0.81	1.13	1.57	2.30	4.74
2	.03	.15	.52	1.83	7.24	.00	.00	.00	.01	.15
3	.00	.01	.07	.36	2.01	.00	.00	.00	.00	.01

Percent loss of precision decreases rapidly from $k=1$ to $k=2$ and it is less than 5% when the number of previous occasions is more than two. Therefore, it is concluded that the preferred number of previous occasions is two over all values of Q.

Percent gain of $\overline{Y}_{t_1}$ and $\overline{Y}_{t_2}$ over $\overline{Y}_t$, the sample mean on the single occasion t, is presented in Table 2. the percent gain increases as r increases over all Q.

Table 2. Percent Gain of $\overline{Y}_{t_1}$, $\overline{Y}_{t_2}$ over $\overline{Y}_t$

r \ Q	.2	.3	.4	.5	.6	.7	.8	.9
	$\overline{Y}_{t_1}$ over $\overline{Y}_t$							
.5	4.21	5.67	6.66	7.14	7.05	6.36	4.99	2.90
.6	6.20	8.47	10.09	10.97	11.03	10.10	8.08	4.79
.7	8.69	12.06	14.53	16.22	16.65	15.66	12.89	7.88
.8	11.74	16.63	20.64	23.52	25.93	24.34	20.98	13.58
.9	15.46	22.47	28.75	34.03	37.82	39.28	36.81	26.89
	$\overline{Y}_{t_2}$ over $\overline{Y}_t$							
.5	4.93	6.45	7.36	7.69	7.42	6.36	5.07	2.91
.6	7.79	10.24	11.75	12.32	11.96	10.64	8.30	4.82
.7	11.88	15.77	18.27	19.36	18.99	17.08	13.51	8.00
.8	17.77	24.07	28.45	30.76	30.80	28.31	22.93	14.01
.9	26.49	37.20	45.73	51.59	54.09	52.31	44.84	29.39

2-3. Determination of the Optimum Discarded Fraction u for the Second-stage Units

In order to see the effect of P and r in reduction of the variance of $\overline{Y}_{t.w}$, the variance formula can be rewritten as

$$V(\overline{Y}_{t.w})=\frac{V_t}{n}\left(\frac{P-P^2C_t}{PQ}\right)$$

where the correction term $\frac{P-P^2C_t}{PQ}$ is a function of P and r. Table 3 provides the value of $\frac{P-P^2C_t}{PQ}$ for $Q\geq.2$ and $r\geq.5$ under the same assumption on r as in 2.2.

From this table, the precision of the estimator increases as r increases for all P and K. However, as defined in 2.1, r is directly related to u, that is, r increases as u approaches zero and r decreases as u becomes larger.

Hence, the optimum value of n is zero. This means the optimum partial retention scheme for second-stage units is to retain the originally selected second-stage units within the matched first-stage units in the sample on every occasion. However, this is difficult or impossible in many practical situations.

Table 3. $\frac{P-P^2C_t}{PQ}$ in $V_{t\ w}$

r \ k	.5	.6	.7	.8	.9	.5	.6	.7	.8	.9
	Q=.2					Q=.6				
1	.95	.94	.92	.89	.86	.93	.90	.85	.80	.72
2	.95	.92	.89	.84	.79	.93	.89	.84	.76	.64
3	.95	.92	.88	.82	.74	.93	.89	.83	.75	.62
	Q=.3					Q=.7				
1	.94	.92	.89	.85	.81	.94	.90	.86	.80	.71
2	.93	.90	.86	.80	.72	.93	.90	.85	.77	.65
3	.93	.90	.85	.78	.68	.93	.90	.85	.77	.64
	Q=.4					Q=.8				
1	.93	.90	.87	.82	.77	.95	.92	.88	.82	.73
2	.93	.89	.84	.77	.68	.95	.92	.88	.81	.69
3	.93	.89	.83	.76	.64	.95	.92	.88	.81	.68
	Q=.5					Q=.9				
1	.93	.90	.86	.80	.74	.97	.95	.92	.88	.78
2	.92	.89	.83	.76	.65	.97	.95	.92	.87	.77
3	.92	.88	.83	.75	.62	.97	.95	.92	.87	.77

Therefore, it is concluded that the practical solution is to choose as small a value of u as possible.

Finally, with regard to the determination of u, the following scheme for second-stage units is suggested to be worthy for further investigation;

If there is no change of a second-stage unit selected on the previous occasion, retain the same units in the sample for the succeeding occasion. If a part of the second-stage units on the previous occasion have changed, then supplement the remaining units by new units on the next occasion. In fact, this is a kind of compromise between $u=u_0$ $(0<u_0\leq 1)$ and $u=0$.

The comparison of the efficiency of the compromise scheme with those in which $u=u_0$ and $u=0$ will be demonstrated by an example. Let us consider the following four types of retention schemes, distinguished by the discarded fraction for second-stage units, u;

Type 1 is the proposed scheme where $u=0$

Type 2 is the compromise where $u_{12}=u_0$ but $u_{23}=0$

Type 3 is the compromise scheme where $u_{12}=0$ but $u_{23}=u_0$

Type 4 is the proposed scheme where $u_{12}=u_{23}=u_0$

and we assume

$$r_{12}=r_{23}=r, \quad r_{12}=r^2 \quad \text{when} \quad u=0$$

$$r_{12}=r_{23}=r', \quad r_{13}=r'^2 \quad \text{when} \quad u_{12}=u_{23}=u_0$$

$$r_{13}=r_{13}' \quad \text{when} \quad u_{13}=u_0$$

Table 4. C_3 in $V(\bar{Y}_{3,w})$

Type \ Q	$r=.9$ $r'=.81$ $r_{13}'=.73$ .3	.4	.5	.6	$r=.9$ $r'=.72$ $r'_{13}=.65$ .3	.4	.5	.6
1	1.11	1.21	1.34	1.53	1.11	1.21	1.34	1.53
2	1.11	1.20	1.32	1.50	1.11	1.19	1.30	1.47
3	1.09	1.15	1.26	1.39	1.07	1.13	1.19	1.28
4	1.09	1.15	1.24	1.37	1.06	1.11	1.17	1.25
	$r=.8$ $r'=.72$ $r_{13}'=.58$				$r=.8$ $r'=.64$ $r_{13}'=.51$			
1	1.08	1.15	1.24	1.35	1.08	1.15	1.24	1.35
2	1.08	1.14	1.22	1.34	1.07	1.13	1.22	1.33
3	1.07	1.12	1.18	1.27	1.05	1.09	1.14	1.21
4	1.06	1.11	1.17	1.26	1.05	1.08	1.13	1.19

It can be seen that Types 2 and 3, belonging to the suggested scheme, have less precision than the proposed scheme in which $u=0$, but higher precision than proposed scheme in which $u=0_0$.

REFERENCES

1. Cochran, W. G. [1963], *Sampling Techniques*, Second Edition, John Wiley and Sons, New York.
2. Des, Raj, [1965], "On Sampling over two Occasions with Proportionate to Size," *Annals of Mathematical Statistics*, 36: 327-330.
3. Dckler, A. R. [1955], "Rotation Sampling," *Annals of Mathematical Statistics*, 36: 664-685.
4. Hansen, N. W., Hurwitz, H. Nisselson and J. Steinberg. [1966], "The Redesign of

the Census Current Population Survey," *Journal of the American Statistical Association*, 50: 701-719.

5. Jessen, R. J. [1942], "Statistical Investigation of a Sample Survey for Obtaining Farm Facts", *Iowa Agricultural Experiment Station Research Bulletin*, No. 304, Ames, Iowa.
6. Onate, B. T. [1960], *Development of Multi-stage Designs for Statistical Serveys in the Phillipines*, Mimeo-Multith Series, No. 3, Statistical Laboratory, Iowa State University, Ames, Iowa.
7. Park, H. N. [1967], *Comparison of Estimators for Two Multistage Designs When Sampling on Successive Occasions*, Unpublished Ph. D. thesis, Department of Experimental Statistics, North Carolina State University, North Carolina.
8. Patterson, H. D. [1950], "Sampling on Successive Occasions with Partial Replacement of Units," *Journal of the Royal Statistical Society*, Series B. 12: 241-255.
9. Rao, J. N. K. and J. E. Graham. [1964], "Rotation Designs for Sampling on Repeated Occasions," *Journal of the American Statistical Association*, 59: 492-509
10. Tikkiwal, B. D. [1955], *Multiphase Sampling on Successive Occasions*, Unpublished Ph. D. thesis, Department of Experimental Statistics, North Carolina Stage University, North Carolina.
11. Yates, A. [1960], *Sampling Method for Census and Surveys*, Charles Griffin and Co., London.