2002 Proceedings of the Spring Conference, Korean Statistical Society

Three-Way Balanced Multi-level Semi Rotation Sampling Designs

YouSung Park, JaiWon Choi and KeeWhan Kim*

Abstract

The two-way balanced one-level rotation design has been discussed (Park, Kim and Choi, 2001), where the two-way balancing is done on interview time in monthly sample and rotation group. We extend it to three-way balanced multi-level design under the most general rotation system. The three-way balancing is accomplished on interview time not only in monthly sample and rotation group but also in recall time. We present the necessary condition and rotation algorithm which guarantee the three-way balancing. We propose multi-level composite estimators (MCE) from this design and derive their variances and mean squared errors (MSE), assuming the correlation from the measurements of the same sample unit and three types of biases in monthly sample.

Keywords : Interview time; Rotation group; Recall time; Necessary condition; Multi-level composite estimator; Variance; Mean squared error; Design efficiency.

1 Introduction

Rotation sampling designs may be classified into two categories by the frequency that a selected sample unit appears in the sample. The first type uses the same sample unit for the entire life of the survey. The typical examples of such design are the U.S. Monthly Retail Trade Survey (MRTS) and the U.S. Monthly Wholesale Trade Survey (these two surveys recently changed their designs from rotating panels to a single fixed panel). The second type uses the sample unit only for a fixed number of times, and we call it semi rotation sampling design. The U.S. Current Population Survey (CPS), the Canadian Labor Force Survey, and

*YouSung Park is an associate professor, Department of Statistics, Korea University, 5-1 Anam-Dong, Sungbuk-gu, KOREA; Jai Won Choi is a mathematical statistician, NCHS, CDC, 6525 Belcrest RD, Hyattsville, MD, U.S.A., and KeeWhan KIM is a Ph.D., Department of Statistics, Korea University, 5-1 Anam-Dong, Sungbuk-gu, KOREA.

the U.S. Consumer Expenditure Survey (CES) belong to this category. In both types of design the entire sample or population is partitioned into a finite number of rotation groups and each group includes a number of sample units. In the first type, group is rotated in and out for monthly sample. When a group is rotated in, all the sample units in the group are interviewed, but in the second type only certain sample units in the group are rotated for interview. Thus, one advantage of the second type is that the respondents have less response burden than those in the first type.

In multi-level rotation design, there are three types of possible biases in monthly sample (Bailar 1975, Cantwell and Caldwell 1998). These biases arise from different interview times of monthly sample (i.e., rotation group bias), different rotation groups in monthly sample (i.e., panel imbalance), and different recall times in monthly sample (i.e., recall time bias). Because the data for a given month t are collected during the current month t as well as the $l-1$ proceeding months in l-level rotation design, we do not have complete information for month t until the month $t+l-1$. Until then we release only the preliminary estimate of month t for characteristics such as monthly level and monthly change. A month later we obtain a second-preliminary estimate, adding one month recall from month $t+1$ to the data of the previous month t. This is repeated until the final estimate for month t includes all $l-1$ recalls from month $t+1$ to month $t+l-1$. The difference between the two successive estimates for month t is called the revision to the estimate. It is desirable that this revision is as small as possible to have a stable estimate. Cantwell and Caldwell (1998) indicated that this revision was accounted mainly by the panel imbalance.

As we want to reduce the three types of biases from monthly sample and revision as much as possible, we propose a class of multi-level semi rotation designs which is a general version of the previous one-level semi rotation designs (Park, Kim and Choi 2001, Cantwell 1990). Such multi-level semi rotation designs are balanced on interview times in monthly sample, rotation group as well as recall time. By this three-way balancing, all rotation groups have equal opportunity to be represented in the sample for every recall time, and some sample units are replaced by similar ones from the same rotation groups. Thus the three-way balanced design eliminates bias arising from taking certain rotation groups more often than others.

The three-way balancing also provides the rotation pattern depending only on interview time, but not on other factors such as survey month, rotation group, and recall time. This implies that the correlation between measurements at any two different survey months is determined only by their interview times. This property of three-way balanced designs enables us to obtain some MCEs and their revisions free from the rotation group bias, panel imbalance, and recall time bias. For example, the MCEs for characteristics such as monthly and yearly changes are completely free from all these three biases.

We present the three-way balanced design by imposing a condition on rotation scheme and use an algorithm to satisfy this condition in constructing such design. When this condition is applied to one-level semi rotation design, we can create more general class of two-way balanced designs than Park, Kim and Choi (2001) could. Therefore, this three-way balancing includes the previous concept of two-way balancing as a special case.

2 Three-Way Balanced Multi-Level Rotation Designs

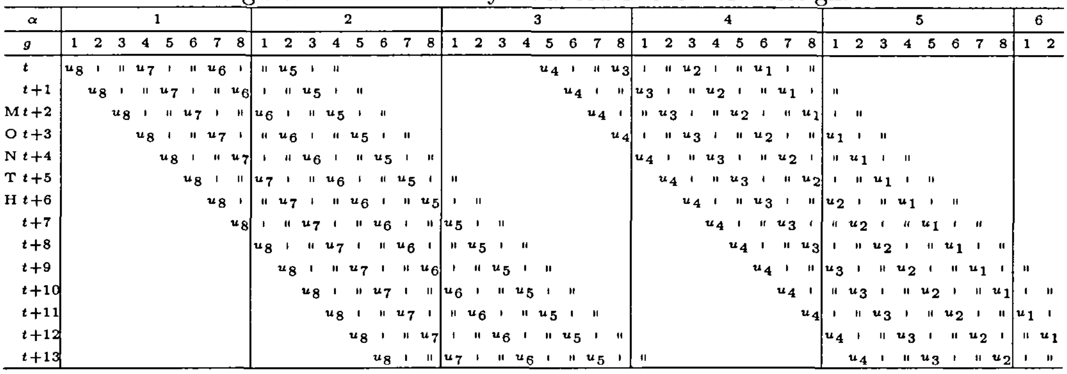
To understand the basic concept of the three-way balanced design, we illustrate the multi-level rotation design with the "3-level" 4-8-4 design in Figure 1: one-level version of this 3-level 4-8-4 design is currently used in the CPS. The "3-level" means that each sample unit reports the information of the interview month as well as of the two previous months. In the 3-level 4-8-4 rotation design, a sample unit is in the sample every third month for 4 times, out of the sample for the next 10 months, and finally returns to the sample every third month for another 4 times. The notation (α, g) in Figure 1 is the index for the αth sample unit in the gth rotation group, and the u_i indicates the corresponding unit α interviewed for the ith time $(i = 1, 2, \ldots, 8)$ in any given month.

The symbols "I" and "II" above the sample unit u_i means that the same sample unit u_i provides the information of the 2 previous months. The recall time of the unit u_i is 0 at the very survey month, "I" right above u_i means one month recall time (recall time 1) from the survey month and "II" means 2 months recall time (recall time 2) from the survey month. For example, the sample unit indexed by $(\alpha = 2, g = 3)$ is interviewed for the 6th time with recall times 0 at month $t+4$, and recall times 1 and 2 at months $t+3$ and $t+2$, respectively. That is, this sample unit provides the information of month $t+3$ by recalling one month at month $t+4$ and the information of month $t+2$ by recalling two months at month $t+4$. This same sample unit was previously interviewed for the 5th time with recall times 0, 1 and 2 at the respective months of $t+1$, t, and $t-1$. In Figure 1, all 8 rotation groups are included in the sample for any survey month. Each group is represented three times by 3 different sample units, each with respective recall times of 0, 1, and 2.

Figures 1-(i), (ii) and (iii) show the two-way balancing on interview times at the respective recall times of 0, 1, and 2. These three pictures are obtained from Figure 1, ignoring α. In each picture, all 8 interview times from 1 to 8 are included in any survey month. This is the horizontal balancing on interview time. This balancing introduces one new sample unit in each month and eliminates the possible bias arising from unbalanced interview times.

Considering any span of 8 months in each of Figures 1-(i), (ii) and (iii), each group contains all 8 interview times. This is the perpendicular balancing on interview times. These horizontal and perpendicular balancing is called as the two-way balancing on interview time in monthly sample and in rotation group in one-level design (Park, Kim and Choi 2001). One more balancing of recall time is added to the two-way balancing in multi-level design.

Figure 1: The three-way balanced 3-level 4-8-4 design.



(i) recall time=0

Month	Rotation Groups							
	1	2	3	4	5	6	7	8
t	u_8	u_5	u_2	u_7	u_4	u_1	u_6	u_3
$t+1$	u_3	u_8	u_5	u_2	u_7	u_4	u_1	u_6
$t+2$	u_6	u_3	u_8	u_5	u_2	u_7	u_4	u_1
$t+3$	u_1	u_6	u_3	u_8	u_5	u_2	u_7	u_4
$t+4$	u_4	u_1	u_6	u_3	u_8	u_5	u_2	u_7
$t+5$	u_7	u_4	u_1	u_6	u_3	u_8	u_5	u_2
$t+6$	u_2	u_7	u_4	u_1	u_6	u_3	u_8	u_5
$t+7$	u_5	u_2	u_7	u_4	u_1	u_6	u_3	u_8
$t+8$	u_8	u_5	u_2	u_7	u_4	u_1	u_6	u_3
$t+9$	u_3	u_8	u_5	u_2	u_7	u_4	u_1	u_6
$t+10$	u_6	u_3	u_8	u_5	u_2	u_7	u_4	u_1
$t+11$	u_1	u_6	u_3	u_8	u_5	u_2	u_7	u_4
$t+12$	u_4	u_1	u_6	u_3	u_8	u_5	u_2	u_7
$t+13$	u_7	u_4	u_1	u_6	u_3	u_8	u_5	u_2

(ii) recall time=1

Month	Rotation Groups							
	1	2	3	4	5	6	7	8
t	u_3	u_8	u_5	u_2	u_7	u_4	u_1	u_6
$t+1$	u_6	u_3	u_8	u_5	u_2	u_7	u_4	u_1
$t+2$	u_1	u_6	u_3	u_8	u_5	u_2	u_7	u_4
$t+3$	u_4	u_1	u_6	u_3	u_8	u_5	u_2	u_7
$t+4$	u_7	u_4	u_1	u_6	u_3	u_8	u_5	u_2
$t+5$	u_2	u_7	u_4	u_1	u_6	u_3	u_8	u_5
$t+6$	u_5	u_2	u_7	u_4	u_1	u_6	u_3	u_8
$t+7$	u_8	u_5	u_2	u_7	u_4	u_1	u_6	u_3
$t+8$	u_3	u_8	u_5	u_2	u_7	u_4	u_1	u_6
$t+9$	u_6	u_3	u_8	u_5	u_2	u_7	u_4	u_1
$t+10$	u_1	u_6	u_3	u_8	u_5	u_2	u_7	u_4
$t+11$	u_4	u_1	u_6	u_3	u_8	u_5	u_2	u_7
$t+12$	u_7	u_4	u_1	u_6	u_3	u_8	u_5	u_2
$t+13$	u_2	u_7	u_4	u_1	u_6	u_3	u_8	u_5

(iii) recall time=2

Month	Rotation Groups							
	1	2	3	4	5	6	7	8
t	u_6	u_3	u_8	u_5	u_2	u_7	u_4	u_1
$t+1$	u_1	u_6	u_3	u_8	u_5	u_2	u_7	u_4
$t+2$	u_4	u_1	u_6	u_3	u_8	u_5	u_2	u_7
$t+3$	u_7	u_4	u_1	u_6	u_3	u_8	u_5	u_2
$t+4$	u_2	u_7	u_4	u_1	u_6	u_3	u_8	u_5
$t+5$	u_5	u_2	u_7	u_4	u_1	u_6	u_3	u_8
$t+6$	u_8	u_5	u_2	u_7	u_4	u_1	u_6	u_3
$t+7$	u_3	u_8	u_5	u_2	u_7	u_4	u_1	u_6
$t+8$	u_6	u_3	u_8	u_5	u_2	u_7	u_4	u_1
$t+9$	u_1	u_6	u_3	u_8	u_5	u_2	u_7	u_4
$t+10$	u_4	u_1	u_6	u_3	u_8	u_5	u_2	u_7
$t+11$	u_7	u_4	u_1	u_6	u_3	u_8	u_5	u_2
$t+12$	u_2	u_7	u_4	u_1	u_6	u_3	u_8	u_5
$t+13$	u_5	u_2	u_7	u_4	u_1	u_6	u_3	u_8

3 Multi-level Composite Estimator

For the interview time $i = 1, 2, \ldots, G$ and for the recall time $j = 0, 1, \ldots, l-1$, let $x_{ti}^{(j)}$ be a measurement of month t from the sample unit interviewed for the ith time at survey month $t+j$ by recalling j previous months. In order to use this $x_{ti}^{(j)}$ in estimation of a characteristic for month t, we need to wait j more months from month t. Since the three-way balanced l-level rotation design has $G = \sum_{i=1}^{m} r_{1i}$ interview times for each recall time in any survey month, denote the average of measurements over the interview times by $\bar{x}_t^{(j)} = 1/G \sum_{i=1}^{G} x_{ti}^{(j)}$ for recall time j and month t.

Hereafter we assume the level $l \geq 2$. Define the following $l-1$ preliminary estimators $y_t^{(j)}$ for $j = 0, 1, \ldots, l-2$, and the final estimator $y_t^{(l-1)}$ for the month t. we call these estimators as the multi-level composite estimators (MCE), and the monthly level MCE $y_t^{(j)}$ of the jth level for month t is defined as

$$y_t^{(j)} = \begin{cases} \bar{x}_t^{(0)} - \omega \bar{x}_{t-1}^{(1)} + \omega y_{t-1}^{(0)} & \text{for } j = 0 \\ (1-\beta_j)\bar{x}_t^{(j)} + \beta_j y_t^{(j-1)} - \omega\big((1-\beta_j)\bar{x}_{t-1}^{(j+1)} + \beta_j y_{t-1}^{(j-1)}\big) + \omega y_{t-1}^{(j)} & \text{for } 1 \le j \le l-2 \\ (1-\beta_{l-1})\bar{x}_t^{(l-1)} + \beta_{l-1} y_t^{(l-2)} & \text{for } j = l-1 \end{cases} \tag{1}$$

where $0 \le \omega, \beta_1, \ldots, \beta_{l-1} \le 1$, and β_j concerns only recall time j. The MCE given in (1) is a specific form of the generalized composite estimator used in one-level rotation design when $j = 0$ (Breau and Ernst 1983). In particular, the equation (1) is reduced to the 2-level Wolter's composite estimator (Wolter 1979) when $l = 2$.

3.1 Bias

Denote the rotation group bias of interview time i, the panel imbalance of rotation group g, and the recall time bias at the recall time j by τ_i, γ_g and η_j, respectively. To reflect these biases into the expectation, we assume that $E(x_{ti}^{(j)}) = \mu_t + \tau_i + \gamma_g + \eta_j$ where μ_t is the monthly effect of $x_{ti}^{(j)}$. Here, γ_g is actually $\gamma_{g(i)}$ to stress the dependency of i. The repeated interviews of the same sample unit are more likely to be correlated. Denote this correlation between months t and $t+t'$ by $\rho_{t'}$. The interview and recall times of a sample unit may also have some impact on its variance. Hence we allow that the variance does not remain the same, but rather varies over the course of interview time i and recall time j of sample unit: $Var(x_{t,i}^{(j)}) = \sigma_{ij}^2$ for all t.

For $t' \ge 0$, interview time $i, i' = 1, 2, \ldots, G$, rotation group $g = 1, 2, \ldots, G$, and recall time $j, j' = 0, 1, \ldots, l-1$ for $l \ge 2$, we summarize above discussion as

$$E(x_{ti}^{(j)}) = \mu_t + \tau_i + \gamma_g + \eta_j \tag{2}$$

$$Cov(x_{ti}^{(j)}, x_{t+t',i'}^{(j')}) = \begin{cases} \rho_{t'}\sigma_{ij}\sigma_{i'j'} & \text{if both are from the same sample unit} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

where $\rho_0 = 1$.

Lemma 3.1. *Suppose that the l-level $r_{11} - r_{21} - \cdots - r_{2,m-1} - r_{1m}$ design is balanced in 3-ways. Under (2), the expectation of the MCE is given by*

$$E(y_t^{(j)}) = \mu_t + \bar{\tau} + \bar{\gamma} + \frac{1}{1-\omega}(\mathbf{a}_j' - \omega \mathbf{b}_j')\boldsymbol{\eta}_j \quad for \quad j = 0, 1, \ldots, l-1$$

where $\bar{\tau} = \sum_{i=1}^{G} \tau_i / G$, $\bar{\gamma} = \sum_{g=1}^{G} \gamma_g / G$, and $\boldsymbol{\eta}_j' = (\eta_0, \eta_1, \cdots, \eta_{j+1})$ for $0 \le j \le l-2$ and $\boldsymbol{\eta}_{l-1}' = \boldsymbol{\eta}_{l-2}'$ for $j = l-1$, $l \ge 2$.

3.2 Variance

By the l-level $r_{11} - r_{21} - \cdots - r_{2,m-1} - r_{1m}$ rotation system, a sample unit, which is interviewed for the first time at month t, is interviewed again for the ith time at month $t + t_i$ where $i = 1, 2, \ldots, G$ and $t_i = (i-1)l + \sum_{j=1}^{m-1} r_{2j} I_{[i > \sum_{\xi=1}^{j} r_{1\xi}]}$. This implies that when we have two sample units interviewed for the i_1th and i_2th times at the respective survey months t and $t + t'$ for $t' \geq 0$, these two sample units are the same only if $t' = t_{i_2} - t_{i_1}$. This relationship can be expressed by the $G \times G$ matrix $L_{t'}$ in which the (i,j)th element is 1 if $t_i = t_j - t'$ for $j \geq i$ and is 0 otherwise. Then, the $L_{t'}$ indicates that two sample units interviewed for the ith time at month t and the jth time at month $t + t'$ are the same only if $(L_{t'})_{ij} = 1$.

Lemma 3.2. *Suppose that a multi-level rotation design is balanced in 3-ways. Then under the covariance structure given in (3) and* $\bar{x}_t^{(j)} = 1/G \sum_{i=1}^{G} x_{ti}^{(j)}$,

$$Cov(\bar{x}_t^{(j_1)},\ \bar{x}_{t+t'}^{(j_2)}) = \frac{\rho_{t'}}{G^2} \mathbf{1}' \Lambda_{j_1} L_{|t'+j_2-j_1|} \Lambda_{j_2} \mathbf{1}, \quad for\ \ j_1, j_2 = 0, 1, \ldots, l-1 \ \ and\ \ t' = 0, 1, \ldots$$

where $\mathbf{1}$ *is the* $G \times 1$ *unit vector and* $\Lambda_j = diag(\sigma_{1j},\ \sigma_{2j},\ \cdots,\ \sigma_{Gj})$ *for* $j = 0, 1, \ldots, l-1$, $l \geq 2$.

Theorem 3.3. *Under the same assumptions given in Lemma 3.2, the variance of* $y_t^{(j)}$ *for* $j = 0, 1, \ldots, l-1$, $l \geq 2$ *can be expressed by*

$$\begin{aligned}(1-\omega^2)Var(y_t^{(j)}) &= \mathbf{a}_j'(V_{0,j} + 2\omega B_{1,0}^{(j)})\mathbf{a}_j + \omega^2 \mathbf{b}_j'(V_{0,j} + 2\omega B_{1,0}^{(j)})\mathbf{b}_j \\ &- 2\omega \mathbf{b}_j'(B_{1,0}^{(j)} + B_{1,1}^{'(j)})\mathbf{a}_j.\end{aligned}$$

References

Bailar, B. (1975). The effects of rotation group bias on estimates from panel survey, *Journal of the American Statistical Association.* 70, 23-30.

Cantwell, P.J. and Caldwell, C.V. (1998). Examining the revisions in Monthly Retail and Wholesale Trade Surveys under a rotating panel design. *Journal of Official Statistics.* Vol. 14, No. 1, 47-59.

Lent, J., Miller, S.M., Cantwell, P.J. and Duff, M. (1999). Effects of composite weights on some estimates from the Current Population Survey. *Journal of Official Statistics.* Vol. 15, No. 3, 431-448.

Park, Y.S., Kim, K.W. and Choi, J.W. (2001). One-level rotation design balanced on time in monthly sample and in rotation group. *Journal of American Statistical Association.* 96, 1483-1496.