Multi-Level Rotation Designs for Unbiased Generalized Composite Estimator

YouSung Park, JaiWon Choi and KeeWhan Kim*

Abstract

We define a broad class of rotation designs whose monthly sample is balanced in interview time, level of recall, and rotation group, and whose rotation scheme is time-invariant. The necessary and sufficient conditions are obtained for such designs. Using these conditions, we derive a minimum variance unbiased generalized composite estimator (MVUGCE). To examine the existence of time-in-sample bias and recall bias, we also propose unbiased estimators and their variances. Numerical examples investigate the impacts of design gap, non-sampling error sources, and two types of correlations on the variance of MVUGCE.

Keywords : Three-way balanced design; Time-invariant rotation scheme; Necessary and sufficient conditions; Generalized composite estimator; Unbiased estimator.

1 Introduction

There are two types of serial correlations arising from rotation schemes in rotation design. One is the first-order correlation between measurements of the same sample unit, and the other is the second-order correlation between different sample units from the same rotation group (Kumar and Lee 1983, Park, Kim and Choi 2001). This means that a time-invariant rotation scheme yields a time-invariant correlation structure between two monthly samples so that estimators used in a rotation design have time-invariant variances.

We characterize the design satisfying the two properties discussed above (i.e., the balanced monthly sample and time-invariant rotation scheme) as three-way balanced design by showing its necessary and sufficient conditions. The three-way balanced design includes previous balanced

*YouSung Park is 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 Ph.D., Department of Statistics, Korea University, 5-1 Anam-Dong, Sungbuk-gu, KOREA.

designs (Cantwell 1990, Park, Kim and Choi 2001) as special cases. Using the necessary and sufficient conditions, we derive the mean squared error of the generalized composite estimator (GCE) which is defined as a linear combination of the current and past information. We also provide a minimum variance unbiased GCE (MVUGCE) by choosing coefficients of the GCE.

The time-in-sample bias arises from different interview times and the recall bias arises from different levels of recall. Typical sources of those two biases are telescoping, panel conditioning, omission, and respondent burden. Thus, the existence of the time-in-sample bias or recall bias implies that those non-sampling error sources influence precision in estimation. We propose unbiased estimators and their variances to measure the existence of the two biases.

2 Three-Way Balanced Multi-Level Rotation Designs

We describe most general rotation system for a ℓ-level rotation design. When a sample unit is selected from each rotation group, this unit returns to the sample for every ℓth month until its r_{11}th interview and is out of the sample for the next $r_{21} + \ell - 1$ successive months. Then, the same sample unit is again interviewed for every ℓth month until its $(r_{11} + r_{12})$th interview and is out of the sample for the next $r_{22} + \ell - 1$ months. This procedure is repeated m cycles until this sample unit returns to the sample for its final $(\sum_{i=1}^{m} r_{1i})$th interview. We denote the ℓ-level rotation design with this rotation system as ℓ-level $r_{11} - \cdots - r_{2,m-1} - r_{1m}$ design. When $m = 1$, we call it the ℓ-level r_{11} in-then-out design.

Figure 1 illustrates the 3-level 2-4-2 design. In this design, there are 4 rotation group and a sample unit is in the sample every third month for 2 times, out of the sample for the next 6 months, and finally returns to the sample every third month for another 2 times. Then the unit retires from the sample completely.

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 $(i = 1, 2, \ldots, 4)$ in any given month. The symbols "ı" and "ıı" above the sample unit u_i means that the same sample unit u_i provides the information of the 2 previous months. The recall level of the unit u_i is 0 at the very survey month, "ı" right above u_i means one month recall (recall level 1) from the survey month and "ıı" means 2 months recall (recall level 2) from the survey month. For example, the sample unit indexed by $(\alpha = 4, g = 3)$ is interviewed for the 3rd time with recall level 0 at month $t + 4$, and recall levels 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 $t+4$.

Figure 1: The three-way balanced 3-level 2-4-2 design.

3-level $2-4-2$ design

α	1				2				3				4				5			
g	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
t	u_4	'	''	u_3	'	''					u_2	'	''	u_1	'	''				
$t+1$		u_4	'	''	u_3	'	''					u_2	'	''	u_1	'	''			
$t+2$			u_4	'	''	u_3	'	''					u_2	'	''	u_1	'	''		
$t+3$				u_4	'	''	u_3	'	''					u_2	'	''	u_1	'	''	
$t+4$					u_4	'	''	u_3	'	''					u_2	'	''	u_1	'	''
$t+5$						u_4	'	''	u_3	'	''					u_2	'	''	u_1	'

(i) recall time=0

month	g			
	1	2	3	4
t	u_4	u_1	u_2	u_3
$t+1$	u_3	u_4	u_1	u_2
$t+2$	u_2	u_3	u_4	u_1
$t+3$	u_1	u_2	u_3	u_4
$t+4$	u_4	u_1	u_2	u_3
$t+5$	u_3	u_4	u_1	u_2

(ii) recall time=1

month	g			
	1	2	3	4
t	u_3	u_4	u_1	u_2
$t+1$	u_2	u_3	u_4	u_1
$t+2$	u_1	u_2	u_3	u_4
$t+3$	u_4	u_1	u_2	u_3
$t+4$	u_3	u_4	u_1	u_2
$t+5$	u_2	u_3	u_4	u_1

(iii) recall time=2

month	g			
	1	2	3	4
t	u_2	u_3	u_4	u_1
$t+1$	u_1	u_2	u_3	u_4
$t+2$	u_4	u_1	u_2	u_3
$t+3$	u_3	u_4	u_1	u_2
$t+4$	u_2	u_3	u_4	u_1
$t+5$	u_1	u_2	u_3	u_4

Figure 2: The three-way balanced 3-level 4-0 design and CEX design

(a) 3-level $4-0$ design

α	1				2				3				4				5	
g	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2
t	u_4	'	''	u_3	'	''	u_2	'	''	u_1	'	''						
$t+1$		u_4	'	''	u_3	'	''	u_2	'	''	u_1	'	''					
$t+2$			u_4	'	''	u_3	'	''	u_2	'	''	u_1	'	''				
$t+3$				u_4	'	''	u_3	'	''	u_2	'	''	u_1	'	''			
$t+4$					u_4	'	''	u_3	'	''	u_2	'	''	u_1	'	''		
$t+5$						u_4	'	''	u_3	'	''	u_2	'	''	u_1	'	''	
$t+6$							u_4	'	''	u_3	'	''	u_2	'	''	u_1	'	''
$t+7$								u_4	'	''	u_3	'	''	u_2	'	''	u_1	'
$t+8$									u_4	'	''	u_3	'	''	u_2	'	''	u_1

(b) CEX design

α	1			2			3			4			5		
g	1	1	1	2	2	2	3	3	3	4	4	4	1	1	1
t	u_4	'	''	u_3	'	''	u_2	'	''	u_1	'	''			
$t+1$		u_4	'	''	u_3	'	''	u_2	'	''	u_1	'	''		
$t+2$			u_4	'	''	u_3	'	''	u_2	'	''	u_1	'	''	
$t+3$				u_4	'	''	u_3	'	''	u_2	'	''	u_1	'	''
$t+4$					u_4	'	''	u_3	'	''	u_2	'	''	u_1	'
$t+5$						u_4	'	''	u_3	'	''	u_2	'	''	u_1
$t+6$							u_4	'	''	u_3	'	''	u_2	'	''
$t+7$								u_4	'	''	u_3	'	''	u_2	'
$t+8$									u_4	'	''	u_3	'	''	u_2

Figures 1-(i), (ii) and (iii) show the two-way balancing on interview times at the respective recall levels of 0, 1, and 2. These three pictures are obtained from Figure 1, ignoring α. In each picture, all 4 interview times from 1 to 4 are included in any survey month. This is the horizontal balancing on interview time. Since the horizontal balancing gives the monthly sample balanced on interview times and levels of recall, any estimator can be unbiased for change as long as it consists of such a balanced monthly sample. The perpendicular balancing is a necessary condition for a time-invariant rotation scheme, which will be shown in the following subsection. Figure 2 shows this relation. The two designs in Figure 2 are the same 3-level 4 in-then-out designs. In particular, Figure 2-(b) is the rotation scheme used in the U.S. consumer expenditure survey (CEX). In the CEX, each sample unit is, in fact, interviewed for 5 times. However, since the information from the first interview is not used for estimation but only for bounding which is a technique to prevent misdating the occurrence of an event and is usually used to reduce the time-in-sample bias (Silberstein and Jacobs 1989), it is understood that we ignore the first interview whenever it is used only for bounding.

Based on observations in Figure 1 and Figure 2, we formally define the three-way balanced multi-level rotation design.

Definition 2.1. The ℓ-level $r_{11} - r_{21} - \cdots - r_{2,m-1} - r_{1m}$ design is balanced in three-ways if it has G rotation groups where $G = \sum_{i=1}^{m} r_{1i}$ and the following three properties are satisfied:

(a) For each survey month, all G rotation groups are represented in the sample and each rotation group is represented by its ℓ different sample units, one with recall time 0, another with recall time 1, ..., and the last with recall time $\ell - 1$.

(b) For each survey month and for each recall time j, $j = 0, 1, \ldots, \ell - 1$, the monthly sample is balanced in such a way that one of the G sample units is interviewed for the 1st time, one for the 2nd time, ..., and one for the Gth time.

(c) For every span of G survey months and for each recall time, each of the G rotation groups contributes its G sample units in which one sample unit is interviewed for the first time, ..., one for Gth time.

3 Minimum Variance Unbiased Generalized Composite Estimator

Some sample units are used repeatedly for a fixed number of months according to their rotation pattern. We can use such repeated measurements of the same sample unit to obtain an efficient estimator. Typical examples are the composite estimator by Rao and Graham (1964) and the generalized composite estimator (GCE) by Breau and Ernst (1983).

For the interview time $i = 1, 2, \ldots, G$ and for the recall time $j = 0, 1, \ldots, \ell - 1$, let x_{tij} 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. Then the composite estimator for level at month t is defined as

$$
\begin{aligned}
y_t &= \sum_{i=1}^{G} \sum_{j=0}^{\ell-1} a_{ij} x_{tij} - \omega \sum_{i=1}^{G} \sum_{j=0}^{\ell-1} b_{ij} x_{t-1,i,j} + \omega y_{t-1} \\
&= \mathbf{a}' \mathcal{X}_t - \omega \mathbf{b}' \mathcal{X}_{t-1} + \omega y_{t-1} \qquad (1)
\end{aligned}
$$

where $0 \le \omega < 1$, $\mathbf{a} = (a_{10}, \cdots, a_{G0}, \cdots, a_{G,\ell-1})'$ and $\mathbf{b} = (b_{10}, \cdots, b_{G0}, \cdots, b_{G,\ell-1})'$ with $\mathbf{a}'\mathbf{1} = \mathbf{b}'\mathbf{1} = 1$, and $\mathcal{X}_t = (x_{t10}, \cdots, x_{tG0}, \cdots, x_{t,G,\ell-1})'$.

When we let $\omega = 0$ in (1) (i.e., when we ignore the information from past months in estimation), y_t is reduced to $\mathbf{a}'\mathcal{X}_t$ which, in turn, becomes the simple estimator when all a_{ij} are the same (i.e.,

$a_{ij} = b_{ij} = ((\ell - 1)G)^{-1}$ for all i and j). We call $\mathbf{a}'\mathcal{X}_t$ simple weighted estimator. We denote the simple weighted estimator by $\tilde{y}_t$ and the simple estimator by $\bar{y}_t$ to distinguish them from the generalized composite estimator (GCE) y_t defined in (1).

3.1 Bias

Typical sources of non-sampling error in a rotation design are telescoping referring to misdating the occurrence of an event, panel conditioning referring to the extent that previous interviews influence the results of subsequent interviews, omission referring to failing to report events that actually occurred, and respondent burden from repeated interviews. These non-sampling error sources are realized by time-in-sample bias and recall bias.

In the three-way balanced design, the expectation $E(x_{tij})$ depends on the interview time i and recall level j. Denote the time-in-sample bias of interview time i and the recall bias of recall level j by η_i and ξ_j, respectively. To reflect these biases into the expectation, we assume that $E(x_{tij}) = \mu_t + \eta_i + \xi_j$ where μ_t is the true mean we want estimate. Let $\boldsymbol{\eta} = (\eta_1, \eta_2, \cdots, \eta_G)'$ and $\boldsymbol{\xi} = (\xi_0, \xi_1, \cdots, \xi_{\ell-1})'$. Then, it is easy to see that, for $k = 0, 1, \cdots,$

$$E(\mathcal{X}_{t-k}) = \mu_t \mathbf{1} + \mathbf{1}_\ell \otimes \boldsymbol{\eta} + \boldsymbol{\xi} \otimes \mathbf{1}_G.$$

where $\mathbf{1}_\ell$ and $\mathbf{1}_G$ are the $\ell \times 1$ and $G \times 1$ unit vectors, respectively. Since $E(y_t) = \mathbf{a}'E(\mathcal{X}_t) - \omega\mathbf{b}'E(\mathcal{X}_{t-1}) + \omega E(y_{t-1})$, by solving this equality recursively, we have the following result.

Lemma 3.1. *For the three-way balanced design,*

$$E(y_t) = \mu_t + \frac{1}{1-\omega}(\mathbf{a}' - \omega\mathbf{b}')[(\mathbf{1}_\ell \otimes \boldsymbol{\eta}) + (\boldsymbol{\xi} \otimes \mathbf{1}_G)]$$

where $\otimes$ denotes the Kronecker product.

Thus, the bias of y_t is invariant to survey time t. This implies that $y_t - y_{t'}$ is unbiased for $\mu_t - \mu_{t'}$ for $t \neq t'$ so that the GCE is unbiased for changes such as monthly and yearly changes.

The three-way balancing implies that each monthly sample contains G interview times from 1 to G for each fixed recall time j so that

$$E(\sum_{i=1}^{G} x_{tij}) = G\mu_t + \sum_{i=1}^{G} \eta_i + G\xi_j. \tag{2}$$

The three-way balancing also ensures that there are G measurements which are from G sample units interviewed for the ith time ($i = 1, \ldots, G$) for any consecutive G months so that

$$E(\sum_{t=1}^{G} x_{tij}) = \sum_{t=1}^{G} \mu_t + G\eta_i + G\xi_j. \tag{3}$$

Therefore, unbiased estimators of $\xi_j - \xi_0$ and $\eta_i - \eta_1$ are obtained from (2) and (3), respectively and they are given by

$$\frac{1}{TG^2}\sum_{t=1}^{TG}\sum_{i=1}^{G}(x_{tij} - x_{ti0}) \text{ for } \xi_j - \xi_0, \text{ and } \frac{1}{TG\ell}\sum_{t=1}^{TG}\sum_{j=0}^{\ell-1}(x_{tij} - x_{ti0}) \text{ for } \eta_i - \eta_1 \tag{4}$$

where we assume that the survey has been conducted at least TG months, $T \geq 1$.

When there is the internal-telescoping, it usually gives $\xi_j - \xi_0 < 0$ for all j and when there is omission it gives $\xi_{j_1} - \xi_{j_2} > 0$ for $j_1 < j_2$, while external-telescoping produces $\eta_i - \eta_1 < 0$ when the first interview is not bounded in a r_{11} in-then-out design. To examine these effects using the unbiased estimators given (4), we need to derive their variances.

Let $\mathbf{u}_k$ be an elementary vector whose kth element is 1 and the remaining elements are all zeros. Observe that the unbiased estimator of $\xi_j - \xi_0$ can written as

$$\frac{1}{TG^2}\sum_{t=1}^{TG}\sum_{i=1}^{G}(x_{tij} - x_{ti0}) = \frac{1}{TG^2}(\mathbf{u}_{j+1}\otimes\mathbf{1}_G - \mathbf{u}_1\otimes\mathbf{1}_G)'\sum_{t=1}^{TG}\mathcal{X}_t$$

where $\mathbf{u}_j$ is a $\ell \times 1$ elementary vector for $j = 1, \cdots, \ell - 1$.

Thus, Lemma ?? gives

$$\begin{aligned}
&Var\Big(\frac{1}{TG^2}\sum_{t=1}^{TG}\sum_{i=1}^{G}(x_{tij} - x_{ti0})\Big)\\
&= \frac{1}{T^2G^4}(\mathbf{u}_{j+1}\otimes\mathbf{1}_G - \mathbf{u}_1\otimes\mathbf{1}_G)'Var\Big(\sum_{t=1}^{TG}\mathcal{X}_t\Big)(\mathbf{u}_{j+1}\otimes\mathbf{1}_G - \mathbf{u}_1\otimes\mathbf{1}_G)\\
&= \frac{1}{T^2G^4}(\mathbf{u}_{j+1}\otimes\mathbf{1}_G - \mathbf{u}_1\otimes\mathbf{1}_G)'\sum_{i=0}^{TG-1}(TG-i)\mathbf{V}_i\ (\mathbf{u}_{j+1}\otimes\mathbf{1}_G - \mathbf{u}_1\otimes\mathbf{1}_G)
\end{aligned}$$

where $\mathbf{V}_i = Cov(\mathcal{X}_t, \mathcal{X}_{t-i})$.

Similarly, the unbiased estimator of $\eta_i - \eta_0$ can written as

$$\frac{1}{TG\ell}\sum_{t=1}^{TG}\sum_{j=1}^{\ell-1}(x_{tij} - x_{tij}) = \frac{1}{TG\ell}(\mathbf{1}_\ell\otimes\mathbf{u}_i - \mathbf{1}_\ell\otimes\mathbf{u}_1)'\sum_{t=1}^{TG}\mathcal{X}_t$$

where $\mathbf{u}_i$ is a $G \times 1$ elementary vector for $i = 1, \cdots, G$ to have

$$\begin{aligned}
&Var\Big(\frac{1}{TG\ell}\sum_{t=1}^{TG}\sum_{j=1}^{\ell-1}(x_{tij} - x_{tij})\Big)\\
&= \frac{1}{T^2G^2\ell^2}(\mathbf{1}_\ell\otimes\mathbf{u}_i - \mathbf{1}_\ell\otimes\mathbf{u}_1)'\sum_{i=0}^{TG-1}(TG-i)\mathbf{V}_i\ (\mathbf{1}_\ell\otimes\mathbf{u}_i - \mathbf{1}_\ell\otimes\mathbf{u}_1).
\end{aligned}$$

4 Conclusion Remarks

We introduced three-way balanced ℓ-level $r_{11} - r_{21} - \cdots - r_{2,m-1} - r_{1m}$ design as a multi-level extension of two-way balanced one-level rotation design. The conditions for two-way balancing by Park, Kim and Choi (2001) are special cases of our necessary and sufficient condition for three-way balancing. Using the properties of three-way balancing, we derived the variances and MSEs of EGCEs for various characteristics such as monthly level, monthly change, yearly level, and yearly change under two types of correlations and biases. We also proposed respective unbiased estimators for differences of time-in-sample biases and recall biases.

When constant variance and no bias are assumed for the measurement depending on interview time and recall time, the simple weighted estimator for characteristics of interest competes with EGCE in variance. However, when variance and bias vary with recall time, EGCE is much better than the simple estimator. In particular, MSE of EGCE decreases while that of simple estimator increases for increasing variance and recall bias depending on recall level. This means that the design with lager recall level can obtain more data with a fixed cost and more accurate information as well only when EGCE is used for estimation. The three-way balanced design with design gaps (months where the respondent is not in the survey) is better than that without design gaps. Ignorance of the second-order correlation is significant in three-way balanced design as we expected.

References

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

Breau, P. and Ernst, L. (1983). Alternative estimators to the current composite estimators. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 397-492.

Cantwell, P.J. (1990). Variance formulae for composite estimators in rotation designs. *Survey Methodology*, **16**, 153-163.

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*, **14**, 47-59.

Huang, E.T. and Ernst, L. R. (1981). Comparison of an alternative estimator to the current composite estimator in CPS. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 303-308.

Kumar, S. and Lee, H. (1983). Evaluation of composite estimation for the Canadian Labor Force Survey. *Survey Methodology*, 9, 403-408.

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

Rao, J.N.K. and Graham, J.E. (1964). Rotation designs for sampling on repeated occasions. *Journal of the American Statistical Association*, **59**, 492-509.

Wolter, K. (1979). Composite estimation in finite population. *Journal of the American Statistical Association*, **74**, 604-613.