응용통계연구 제 17권 1호, 2004년, pp. 61-73

l/G 교체표본디자인에서의 일반화복합추정량과 평균제곱오차에 관한 연구 *

김기환[1)] 박유성[2)] 남궁재은[3)]

요 약

교체표본조사에서는 모든 표본단위를 복수 개($=G$)의 교체그룹으로 나누고 일정 횟수만큼 조사한 후 표본단위의 교체를 하는 경우와 조사기간 동안 동일한 표본단위를 조사한 후 교체그룹 자체를 교체하는 두 가지 경우가 있다. 본 연구는 후자의 경우를 일반화하는 것으로, 매 조사월에서 하나의 교체그룹이 조사되고 이 교체그룹에 속한 모든 표본단위는 최근 l개월 동안의 정보를 제공하는 l−수준 교체표본설계이다. 표본단위 교체가 오직 교체그룹의 총 개수인 G와 회상 개월 수인 l에 의해 결정되므로 이를 l/G 교체표본설계로 일반화하였으며 일반화복합추정량의 분산과 두 가지 형태의 편향(bias)하에서 MSE를 구하고 절충 GCE(compromise GCE)의 계수를 유도하였다. 또한 GCE의 분산과 MSE를 상관계수, 편향, 표본조사단위의 분산의 형태, 그리고 설계간격(design gap)의 형태에 따라 분석하였다.

주요용어: 다수준교체표본조사, 설계간격, 일반화 복합추정량, 편향.

1. 서 론

교체표본조사(rotation sampling)는 시간에 의존하는 모집단의 특성을 조사하기 위한 반복조사(repeated survey)의 종류인 고정표본조사(fixed panel survey)와 독립표본조사 (independent sample survey)의 장점을 적절히 수용한 조사방법이다. 이는 일정기간 반복조사 후 정해진 규칙에 따라 새로운 표본들로 교체하는 방법으로 조사시점 사이에 중복되는 표본에 의해 발생되는 자기상관관계를 이용하여 모집단의 변화에 대한 추정치의 표본오차를 줄일 수 있고, 표본교체에 의하여 응답자들에게 부담감을 분산시켜 응답률을 유지할 수 있다.

교체표본조사는 보고하는 수준(level)에 따라 일수준교체표본조사와 다수준교체표본조사로 분류된다. 여기서 수준이란, 응답자가 한 번의 조사에서 조사자에게 보고하는 자료들의 시점 수를 말한다. 일수준교체표본조사에서 응답자는 현시점의 자료만을 보고하며, 그 예로는 미국 통계국의 CPS(Current Population Survey)와 캐나다 통계국의 LFS(Labour

* 이 논문은 2003년도 고려대학교 박사후연수과정 연구비 지원에 의하여 수행되었음.

1) (136-701) 서울시 성북구 안암동 5가 1, 고려대학교 통계연구소, 연구원
E-mail: korpen@korea.ac.kr
2) (136-701) 서울시 성북구 안암동 5가 1, 고려대학교 통계학과, 교수
E-mail: yspark@korea.ac.kr
3) (136-701) 서울시 성북구 안암동 5가 1, 고려대학교 통계학과 대학원
E-mail: statis1000@hotmail.com

Force Survey)가 있다. 다수준교체표본조사는 응답자가 한번의 인터뷰에서 조사시점의 자료뿐만 아니라 조사시점 이전의 일정기간에 대한 자료를 동시에 보고하며, 그 예로는 미국 통계국의 2-수준 교체표본조사인 MRTS(Monthly Retail Trade Survey)와 4-수준 교체표본 조사인 SIPP(Survey of Income and Program Participation)가 있다.

일수준 교체표본추출방법의 일반화에 대한 연구는 Rao and Graham (1964)에 의해 시도 되었고 Park, Kim and Choi (2001)에 의하여 일반화 되었다. Park, Kim and Choi의 연구는 표본의 교체방법의 일반화 뿐 아니라 표본중복에 의해 발생되는 상관구조를 정확히 밝힘으로써 일반화복합추정량(Generalized Composite Estimator, GCE)의 분산식을 명확한 형태로 제시하였다. 다수준 교체표본추출방법의 기존의 연구로는 Bershad (1953), Wolter (1979), Cantwell (1990), Caldwell and Cantwell (1996), 박유성, 배경화, 김기환(2001), 박유성, 문원기, 김기환 (2002)등이 있다.

본 논문은 다수준교체표본추출방법에서 G개의 교체그룹을 갖는 l-수준교체표본추출방법인 l/G 디자인을 고려하였으며 l/G디자인의 이해를 위하여 구체적인 설명을 2절에서 하였다. 교체그룹으로부터 정보가 제공되지 않는 월이 존재하면 이 설계를 설계간격이 있다고 한다. Cantwell (1990)은 설계간격(design gap)이 존재하지 않는 4-수준 교체표본설계인 SIPP에서 일반화복합추정량을 이용하여 현월추정량(monthly level)과 월간차 추정량(monthly change)의 분산을 유도하였다. 그러나 다수준교체표본에서 가장 중요한 특성치인 연 평균과 연 평균의 차에 대한 분산은 구해지지 않았다. 이에 3절에서는 설계간격의 존재여부에 관계없이, 교체표본조사에서 관심을 가지고 있는 모든 추정량에 대해 다수준 교체표본 설계를 일반화한 l/G 디자인에서 GCE를 적용하고 이의 분산, 다수준교체표본설계에서 발생하는 두 가지 형태의 편향을 고려한 MSE, 이들을 최소화하는 최적계수를 유도하였다. 4절에서는 각 추정량의 효율을 단순추정량과 비교하고 설계간격의 형태에 따른 디자인의 효율을 비교하였으며, 5절에서는 각 추정량의 효율을 비교한 결과를 정리하였다.

2. l/G교체표본디자인

다수준교체표본조사에서 가장 널리 쓰이고 있는 교체규칙은 표본을 복수개의 교체그룹으로 나눈 후, 매 조사시점마다 오직 하나의 교체그룹만을 조사하고 일정한 규칙에 따라 교체그룹을 교체하는 조사방법으로 응답자는 한번의 인터뷰에서 조사시점의 자료뿐만 아니라 조사시점 이전의 일정기간에 대한 자료를 동시에 보고한다. 미국의 SIPP에서는 표본을 4개의 교체그룹으로 나눈 후 매월 하나의 교체그룹을 조사하고 조사되는 교체그룹 내에 있는 모든 표본단위는 현월과 과거 3개월의 정보를 제공한다. 즉, 각 교체그룹은 4개월마다 조사되나 회상에 의해 조사되지 않는 시점의 정보도 모두 구할 수 있게 된다. 그림 2.1은 SIPP의 교체규칙을 그림으로 나타낸 것으로 $x_t^{(j)}$는 어떤 관심 있는 특성치에 관한 $t+j$월에서 j개월 전을 회상하여 보고 된 t월의 관찰치 이다. 예를 들어 교체그룹 1은 시점 $t-4$와 t에서 조사가 이루어지고 있음을 알 수 있고 시점 t에서 $t-1, t-2, t-3$에 대한 정보까지 $x_{t-1}^{(1)}, x_{t-2}^{(2)}, x_{t-3}^{(3)}$에 의해 각각 보고 되고 있음을 알 수 있다. 같은 방법으로 교체그룹 2는 $t-3$과 $t+1$에서, 교체그룹 3은 $t-2$와 $t+2$에서, 교체그룹 4는 $t-1$과 $t+3$에서 조사됨

으로서, 매 조사시점에서는 오직 1개의 교체그룹이 조사되나 회상에 의해 모든 교체그룹에 대한 정보를 갖게 됨을 알 수 있다. MRTS는 3개의 교체그룹이 있으나 조사 시 현월과 전

조사월	교체그룹			
	1	2	3	4
⋮	⋮	⋮	⋮	⋮
t-4	$x_{t-4}^{(0)}$	$x_{t-4}^{(1)}$	$x_{t-4}^{(2)}$	$x_{t-4}^{(3)}$
t-3	$x_{t-3}^{(3)}$	$x_{t-3}^{(0)}$	$x_{t-3}^{(1)}$	$x_{t-3}^{(2)}$
t-2	$x_{t-2}^{(2)}$	$x_{t-2}^{(3)}$	$x_{t-2}^{(0)}$	$x_{t-2}^{(1)}$
t-1	$x_{t-1}^{(1)}$	$x_{t-1}^{(2)}$	$x_{t-1}^{(3)}$	$x_{t-1}^{(0)}$
t	$x_{t}^{(0)}$	$x_{t}^{(1)}$	$x_{t}^{(2)}$	$x_{t}^{(3)}$

그림 2.1: 4개의 교체그룹을 가지는 4-수준 표본설계 SIPP 디자인

월 2개월 동안의 정보만을 제공함으로써 매월 1개의 교체그룹에 대한 정보는 생략되게 된다. 그림 2.2는 MRTS의 교체규칙을 나타내고 있다. 이와 같이 매월 1개 이상의 교체그룹에

조사월	교체그룹		
	1	2	3
⋮	⋮	⋮	⋮
t-4	$x_{t-4}^{(1)}$		$x_{t-4}^{(0)}$
t-3	$x_{t-3}^{(0)}$	$x_{t-3}^{(1)}$	
t-2		$x_{t-2}^{(0)}$	$x_{t-2}^{(1)}$
t-1	$x_{t-1}^{(1)}$		$x_{t-1}^{(0)}$
t	$x_{t}^{(0)}$	$x_{t}^{(1)}$	

그림 2.2: 3개의 교체그룹을 가지는 2-수준 교체표본설계 MRTS 디자인

대한 정보가 조사되지 않을 때 이를 설계간격이 있다고 한다.

이제, 앞에서 살펴본 SIPP와 MRTS의 교체규칙을 일반화하여 보자. 모든 표본단위를 G개의 교체그룹으로 나눈 후, 일정한 순서에 따라 매월 1개의 교체그룹을 조사하며 이를 매 G개월마다 반복한다고 하자. 또한 조사시점에서는 조사시점을 포함하여 과거의 $l-1$개월 동안의 시점, 즉, 총 l개월의 정보를 제공하는 교체규칙을 고려한다면 이를 l/G 교체 디자인이라고 정의할 수 있다. 예를 들어 SIPP는 4/4디자인으로 표기되며 MRTS는 2/3디자인으로 표현된다. 따라서 $l=G$이면 SIPP와 같은 설계간격이 없는 교체표본디자인이 되며, $l<G$이면 MRTS와 같이 설계간격이 존재하는 교체표본디자인이 됨을 알 수 있다.

3. l/G 교체표본디자인에서의 일반화복합추정량

교체표본조사에서 얻어진 자료는 서로 다른 시기에 조사된 동일한 표본의 반복에서 생기는 중복을 이용하여 과거 시점의 자료를 현시점의 모집단 특성추정에 이용하여 모집단의 특성을 좀 더 정확하게 추정할 수 있게 해 준다.

Y_t를 t월에서 조사하려는 특성치라 하면 Breau and Ernst (1983)가 제시한 일반화복합추정량(generalized composite estimator, GCE)은 식(3.1)과 같이 고정된 t시점에서 현재와 과거 자료의 선형결합으로 이루어져 있으며 과거의 모든 데이터를 이용할 수 있는 장점이 있다.

$$y_t = \sum_{j=0}^{l-1} a_j x_t^{(j)} - \omega \sum_{j=0}^{l-1} b_j x_{t-1}^{(j)} + \omega y_{t-1} = \mathbf{a}'\mathbf{X}_t - \omega \mathbf{b}'\mathbf{X}_{t-1} + \omega y_{t-1} \tag{3.1}$$

여기서, y_t는 Y_t의 추정량을 나타내고 $x_t^{(j)}$는 $t+j$월에서 j개월 전을 회상하여 보고 된 t월의 관찰치를 나타낸다. 여기서 $\mathbf{a} = (a_0, a_1, \cdots, a_{l-1})'$, $\mathbf{b} = (b_0, b_1, \cdots, b_{l-1})'$ 그리고 $\mathbf{X}_t = (x_t^{(0)}, x_t^{(1)}, \cdots, x_t^{(l-1)})'$ 이며 $\mathbf{a}'\mathbf{1} = \mathbf{b}'\mathbf{1} = 1$ 이고 $0 \le \omega < 1$이라는 제약조건을 만족한다. 만약 $a_j = 1/l$이고 $w = 0$으로 고정시키는 경우 y_t는 현재 교체표본조사에서 실무적으로 많이 사용되는 단순추정량(simple estimator)이 된다.

3.1. 일반화복합추정량의 편향과 분산

다수준교체표본조사에서 응답자는 현월의 정보뿐 아니라 과거 여러 시점의 자료를 회상에 의해 동시에 얻게 되므로 보고하는 수준이 증가할수록 비용은 줄어들게 되나 응답자가 과거 시점에 대한 정확한 데이터를 보고하는 것이 어렵게 된다. 이에 회상에 대한 혼란이 가중되어 회상편향(recall bias)이 발생할 수 있다. 한편, Cantwell and Caldwell (1998)이 지적하였듯이, 서로 다른 교체그룹은 특성치에 대해 서로 다른 기대값을 가질 수 있어 교체그룹편향(rotation group bias)이 발생할 수 있다. 즉 $x_t^{(j)}$가 g번째 교체그룹으로부터 관찰치라고 할 때 $x_t^{(j)}$의 기대값은

$$E(x_t^{(j)}) = \mu_t + \tau_g + \xi_j\ ,\quad g = 1, \cdots, G,\quad j = 0, \cdots, l-1$$

으로 표기하며 여기서 μ_t는 구하고자 하는 참 평균, τ_g는 g번째 교체그룹편향, ξ_j는 j번째 회상에 의한 편향을 나타낸다. y_t의 편향을 구하기 위하여 다음의 두 가지 행렬을 정의하였다.

$$(I_{l\times G})_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}, \qquad (L_{G\times G})_{i,j} = \begin{cases} 1 & \text{if } j = i+1 \text{ or } i = G \text{ and } j = 1 \\ 0 & \text{otherwise} \end{cases}$$

즉, $I_{l\times G}$행렬은 $I_{G\times G}$ 단위행렬로부터 마지막 $G-l$행들을 없앤 것이며 $L_{G\times G}$행렬은 전방위순환행렬(forward permutation matrix)이다. 이 두 가지 행렬을 이용하여 y_t의 기대값을 구하면 다음과 같다.

정리 3.1 l/G 디자인에서 $t=0$에서 첫 번째 교체그룹이 조사된다고 하자. 그러면 y_t의 기대값은 아래와 같다.

$$E(y_t)=\mu_t+\frac{1}{1-\omega^G}\mathbf{a}'\sum_{k=0}^{G-1}\omega^k IL^{t-k}\boldsymbol{\tau}-\frac{\omega}{1-\omega^G}\mathbf{b}'\sum_{k=0}^{G-1}\omega^k IL^{t-k-1}\boldsymbol{\tau}+\frac{1}{1-\omega}(\mathbf{a}'-\omega\mathbf{b}')\boldsymbol{\xi}$$

여기에서, $\boldsymbol{\tau}=(\tau_1,\cdots,\tau_G)'$이고, $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_{l-1})'$이다.

증명: l/G디자인의 교체 규칙에 의해 $t=0$일 때 교체그룹은 1은 $x_t^{(0)}$, 교체그룹 2는 $x_t^{(1)}$, $\cdots$, 교체그룹 l은 $x_t^{(l-1)}$을 제공한다. $t=1$일 때 $t=0$에서 교체그룹이 1개씩 밀려서 조사되는 형태를 따르므로 즉, 교체그룹 2는 $x_1^{(0)}$, 교체그룹 3은 $x_1^{(1)}$, $\cdots$, 교체그룹 $l+1$은 $x_1^{(l-1)}$을 제공한다. 동일한 방법으로 1개월이 지남에 따라 하나씩 교체그룹이 밀리는 형태가 되고 마지막 G 교체그룹이 사용되면 다음으로 교체그룹이 1이 다시 되돌아오는 순환형태를 띄게 된다. 즉 교체그룹을 $(1,2,\cdots,G)$ 벡터로 표현할 때 $IL(1,2,\cdots,G)'=(2,3,\cdots,l+1)$이 되어 이는 $t=1$일 때의 교체그룹을 나타내며 $IL^2(1,2,\cdots,G)'=(3,4,\cdots,l+2)$가 되어 $t=2$일 때 사용한 교체그룹을 나타낸다. 그러므로 일반적으로, $IL^t\boldsymbol{\tau}$는 시점 t에서 사용되는 교체그룹을 나타내며 $IL^t\boldsymbol{\tau}$에서 첫번째 행의 교체그룹은 $x_t^{(0)}$를, 두번째행의 교체그룹을 $x_t^{(1)}$을, $\cdots$, l번째 행의 교체그룹은 $x_t^{(l-1)}$이 된다. 그러므로 $\mathbf{X}_t=(x_t^{(0)},x_t^{(1)},\cdots,x_t^{(l-1)})'$의 기대값은 $E(\mathbf{X}_t)=\mu_t\mathbf{1}+IL^t\boldsymbol{\tau}+\boldsymbol{\xi}$ 로 여기에서 $\mathbf{1}$는 $l\times 1$ 단위벡터이다. 식(3.1)에 의하여

$$\begin{aligned}
E(y_t)&=\sum_{k=0}^{\infty}\omega^k\mathbf{a}'E(\mathbf{X}_{t-k})-\sum_{k=0}^{\infty}\omega^{k+1}\mathbf{b}'E(\mathbf{X}_{t-k-1})\\
&=\sum_{k=0}^{\infty}\omega^k\mathbf{a}'(\mu_{t-k}\mathbf{1}+IL^{t-k}\boldsymbol{\tau}+\boldsymbol{\xi})-\sum_{k=0}^{\infty}\omega^{k+1}\mathbf{b}'(\mu_{t-k-1}\mathbf{1}+IL^{t-k-1}\boldsymbol{\tau}+\boldsymbol{\xi})\\
&=\sum_{k=0}^{\infty}\omega^k\mathbf{a}'\mathbf{1}(\mu_{t-k})-\sum_{k=0}^{\infty}\omega^{k+1}\mathbf{b}'\mathbf{1}(\mu_{t-k-1})\\
&\quad+\sum_{k=0}^{\infty}\omega^k\mathbf{a}'(IL^{t-k}\boldsymbol{\tau}+\boldsymbol{\xi})-\sum_{k=0}^{\infty}\omega^{k-1}\mathbf{b}'(IL^{t-k-1}\boldsymbol{\tau}+\boldsymbol{\xi})\ .
\end{aligned}$$

이 되고 제약조건 $\mathbf{1}'\mathbf{a}=\mathbf{1}'\mathbf{b}=1$과 $L^t=L^{t+G}$임을 이용하면

$$E(y_t)=\mu_t+\frac{1}{1-\omega^G}\mathbf{a}'\sum_{k=0}^{G-1}\omega^k IL^{t-k}\boldsymbol{\tau}-\frac{\omega}{1-\omega^G}\mathbf{b}'\sum_{k=0}^{G-1}\omega^k IL^{t-k-1}\boldsymbol{\tau}+\frac{1}{1-\omega}(\mathbf{a}'-\omega\mathbf{b}')\boldsymbol{\xi}\ .$$

를 얻을 수 있다. □

l/G디자인에서는 동일한 교체그룹이 매 G개월마다 반복 조사되므로 이들 간의 상관관계가 존재하게 된다. 따라서 조사월 t와 $t+t'$의 상관관계를 $\rho_{t'}$로 표기하고 이를 이용하여 $x_t^{(j)}$와

$x_{t+t'}^{(j)}$의 공분산 구조를 다음과 같이 가정하자.

$$Cov\big(x_t^{(j)},\ x_{t+t'}^{(j')}\big) = \begin{cases} \sigma_j^2, & t' = 0 \text{ 이고 } j = j' \text{ 인 경우.} \\ \rho_{t'}\sigma_j\sigma_{j'}, & t \neq t' \text{ 이고 같은 교체그룹일 경우} \\ 0, & \text{나머지 경우.} \end{cases} \tag{3.2}$$

즉, 다른 교체그룹 간에는 무상관 관계를 가정하고 있으며, 회상시간에 의한 정보의 질은 일반적으로 다를 것으로 예상되어 분산이 회상시간 j에 의존하도록 하고 있다. $x_t^{(j)}$와 $x_{t+t'}^{(j')}$가 같은 교체그룹에서 속하는지는 다음으로부터 판단할 수 있다. $x_t^{(j)}$는 $t+j$시점에서 조사된 교체그룹으로부터 구한 관찰치 이고 $x_{t+t'}^{(j')}$는 $t+t'+j'$시점에서 조사된 교체그룹으로부터 구한 관찰치 라고 하면, l/G디자인에서는 모든 교체그룹이 G개월마다 조사되므로 만약 $|t+t'+j'-(t+j)|$가 G의 배수이면 $x_t^{(j)}$와 $x_{t+t'}^{(j')}$은 같은 교체그룹에 속하는 관찰치 이고 G의 배수가 아니면 두개의 관찰치는 다른 교체그룹에 속하는 것으로 서로 간에 무상관을 의미하게 된다. 이 관계를 이용하여 식(3.2)의 공분산 구조를 쓰면

$$Cov\big(x_t^{(j)},\ x_{t+t'}^{(j')}\big) = \rho_{t'}\sigma_j\sigma_{j'}I_{[mod_G(t'+j'-j),0]}$$

가 된다. 여기에서 $mod_G(|t'+j'-j|)=0$이면 $I_{[mod_G(t'+j'-j),0]}=1$이고 $mod_G(|t'+j'-j|)\neq 0$이면 $I_{[mod_G(t'+j'-j),0]}=0$이다. 이제 $\mathbf{X}_t$와 $\mathbf{X_{t+t'}}$의 공분산을 $l\times l$ 블럭 행렬인 $V_{t'}$로 표현하면

$$Cov\big(\mathbf{X_t},\mathbf{X_{t+t'}}\big) = V_{t'}\ ,\quad t'=0,1,\cdots$$

이 되고, 여기에서 $V_{t'}$의 (i,j)번째 요소는 $Cov\big(x_t^{(i-1)},x_{t+t'}^{(j-1)}\big)$임을 나타낸다.

일반적으로 시계열자료 생성을 목적으로 하는 표본조사에서는 특정 월의 특성치 추정뿐만 아니라 월변화, 분기변화, 연 합, 연 합의 차등에 대한 추정에도 많은 관심을 갖게 되므로 우선, l/G 교체 디자인에서 현월추정량의 분산을 유도하면 다음과 같다.

정리 3.2 식(3.2) 하에서, 현월 일반화복합추정량 y_t의 분산은 $n_1\geq n_2$에 대해서, $B_{n_1,n_2}=\sum_{k=0}^{\infty}\omega^{k+n_2}\ V_{n_1+k-n_2}$ 일 때 다음과 같다.

$$Var(y_t) = \mathbf{a}'Q_1\mathbf{a} + \mathbf{b}'Q_3\mathbf{b} + \mathbf{b}'Q_2\mathbf{a}$$

여기서, $Q_1=(1-\omega^2)^{-1}(V_0+2\omega B_{1,0})$, $Q_2=(1-\omega^2)^{-1}\left(B_{1,0}+2\omega B_{1,1}'\right)$, $Q_3=\omega^2Q_2$ 이다.

월간차, 일정기간의 합, 일정기간의 합과 합의 차에 대한 추정량의 분산을 구하기 위해서 $P_1(t_0)=2(1-\omega t^{t_0})Q_1-2\sum_{n=0}^{t_0-1}B_{t_0,n}$, $P_2(t_0)=2(1-\omega t^{t_0})Q_2+2\sum_{n=0}^{t_0-1}(\omega B_{t_0+1,n}+B_{t_0,n+1}^{'})$, $P_3(t_0)=\omega^2P_1(t_0)$, $t_0\geq 1$ 로 놓고 월간차의 분산을 구하면 다음과 같다.

정리 3.3 식(3.2) 하에서, 월간차 일반화복합추정량 $y_t-y_{t_0}$, $t_0\geq 1$의 분산은 다음과 같다.

$$Var(y_t-y_{t_0}) = \mathbf{a}'P_1(t_0)\mathbf{a} + \mathbf{b}'P_2(t_0)\mathbf{b} + \mathbf{b}'P_3(t_0)\mathbf{b}$$

다음의 일정기간의 합에 대한 추정량 $S_t^{t_0}=y_t+y_{t-1}+\cdots+y_{t-t_0+1}=\sum_{t'=0}^{t_0-1}y_{t-t'}, t_0\geq 1$을 구하면 다음과 같다.

정리 3.4 식(3.2) 하에서, 일정기간의 합 일반화복합추정량 $S_t^{t_0}$ 의 분산은 다음과 같다.

$$\begin{aligned} Var(S_t^{t_0}) =& \mathbf{a}'\Big(t_0^2 Q_1 - \sum_{t'=1}^{t_0-1}(t_0-t')P_1(t')\Big)\mathbf{a} \\ &+ \mathbf{b}'\Big(t_0^2 Q_2 - \sum_{t'=1}^{t_0-1}(t_0-t^*)P_2(t')\Big)\mathbf{a} + \mathbf{b}'\Big(t_0^2 Q_3 - \sum_{t'=1}^{t_0-1}(t_0-t')P_3(t')\Big)\mathbf{a} \end{aligned}$$

마지막으로 합과 합의 변동에 관한 일반화복합추정량 $S_t^{t_0} - S_{t-t1}^{t_0} = \sum_{i=0}^{t_0-1} y_{t-i} - \sum_{i=0}^{t_0-1} y_{t-t_1-i}$, $t_1 \geq t_0$를 구하면 다음과 같다.

정리 3.5 식(3.2) 하에서, 합과 합의 변동에 관한 일반화복합추정량 $S_t^{t_0} - S_{t-t1}^{t_0}, t_1 \geq t_0$ 의 분산은 다음과 같다.

$$\begin{aligned} Var(S_t^{t_0} - S_{t-t_1}^{t_0}) = \mathbf{a}' &\left(\sum_{t'=-t_0+1}^{t_0-1} (t_0 - |t'|)P_1(t_1-t') - 2\sum_{t'=1}^{t_0-1}(t_0-t')P_1(t') \right) \mathbf{a} \\ + \mathbf{b}' &\left(\sum_{t'=-t_0+1}^{t_0-1} (t_0 - |t'|)P_2(t_1-t') - 2\sum_{t'=1}^{t_0-1}(t_0-t')P_2(t') \right) \mathbf{a} \\ + \mathbf{b}' &\left(\sum_{t'=-t_0+1}^{t_0-1} (t_0 - |t'|)P_3(t_1-t') - 2\sum_{t'=1}^{t_0-1}(t_0-t')P_3(t') \right) \mathbf{b}. \end{aligned}$$

정리 3.2~정리 3.5의 증명은 Park, Kim and Choi(2001)의 결과로부터 쉽게 유도되므로 생략하였다.

3.2. 일반화복합추정량의 최적계수

일반화복합추정량 y_t는 미지 계수 $\mathbf{a}$, $\mathbf{b}$, ω를 포함하고 있다. 여기서, $\mathbf{a}$, $\mathbf{b}$는 각각 $\mathbf{X_t}$와 $\mathbf{X_{t-1}}$이 구성하고 있는 교체그룹에 부여되는 가중치로 해석되며, ω는 보고 되는 관찰치 사이의 상관관계가 고려된 일반화복합추정량을 계산함에 있어서 과거의 관찰치 들을 사용하는 정도를 나타낸다. 따라서 미지 계수의 실질적인 값을 결정하는 것은 매우 중요하다. 이와 관련 된 또 하나의 문제는 여러 종류의 추정량들에 관심이 있을 경우, 모든 추정량에 대해 각각 다른 계수들을 사용한다는 것은 너무 복잡해져 현실적인 활용성이 떨어지게 된다. 이에 관심 있는 추정량 y_t, $y_t - y_{t-t_0}$, $S_t^{t_0}$, $S_t^{t_0} - S_{t-t1}^{t_0}$에 대한 분산, MSE의 가중 합을 최소화 함으로써 얻어지는 절충복합계수(compromised composite coefficient)가 제안되었다(Park, Kim and Choi, 2001). 따라서 관심 있는 H개의 일반화복합추정량이 있다고 할 때 이때의 일반화복합추정량을 $z_{th}, h = 1, 2, \cdots, H$라 표기하고 제약조건 $\mathbf{a}'\mathbf{1} = \mathbf{b}'\mathbf{1} = 1$을 고려하여 라그랑즈 승수법에 의해 MSE를 최소하기 위한 목적함수 O는 다음과 같이 놓을 수 있다.

$$O = \sum_{h=1}^{H} \delta_h MSE(z_{th}) - \lambda_1(\mathbf{1}'\mathbf{a} - 1) - \lambda_2(\mathbf{1}'\mathbf{b} - 1) \tag{3.3}$$

여기서, λ_1, λ_2는 라그랑즈 상수이며 δ_h는 대응되는 추정량의 상대중요도를 나타내는 가중치이다. 식(3.3)의 $MSE(z_{th})$는 식(3.4)와 같으며 여기서 z_{th}를

$$MSE(z_{th}) = \mathbf{a}'C_{1h}\mathbf{a} + \mathbf{b}'C_{2h}\mathbf{a} + \mathbf{b}'C_{3h}\mathbf{b} \tag{3.4}$$

4개의 추정량 즉, $h = 1, \cdots, 4$에 대하여 $z_{t1} = y_t$, $z_{t2} = y_t - y_{t-t_0}$, $z_{t3} = S_t^{t_0}$, $z_{t4} = S_t^{t_0} - S_{t-t1}^{t_0}$ 라면 $k = 1, 2, 3$에 대하여 $C_{kh} = P_{kh} + Bias_{kh}^2$이다. 여기서, $Bias = 0$ 일때 식(3.3)은 분산을 최소화하는 라그랑즈 함수이다. 따라서 식(3.4)는 $Var(z_{th})$의 형태와 같음을 알 수가 있다. 위의 식(3.3)으로부터 일반화복합추정량의 MSE를 최소화하는 복합계수 $\hat{\mathbf{a}}$와 $\hat{\mathbf{b}}$를 유도하면 다음과 같다.

정리 3.6 l/G 교체 디자인에서, 가중치 $\delta_h, h = 1, 2, \cdots, H$ 에 대하여 $\sum_{h=1}^{H} \delta_h MSE(z_{th})$를 최소화하는 복합 계수 $\hat{\mathbf{a}}, \hat{\mathbf{b}}$는 다음과 같이 주어진다.

$$\begin{pmatrix} \hat{\mathbf{a}} \\ \hat{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} C_1 + C_1' & C_2' - s_1^{-1}\mathbf{1}\mathbf{1}'(C_1 + C_1')^{-1}C_2' \\ C_2 - s_3^{-1}\mathbf{1}\mathbf{1}'(C_3 + C_3')^{-1}C_2 & C_3 + C_3' \end{pmatrix} \begin{pmatrix} s_1^{-1}\mathbf{1} \\ s_3^{-1}\mathbf{1} \end{pmatrix}$$

여기서, $\mathbf{1}$은 $l \times 1$ 단위벡터이며 $s_1 = \mathbf{1}'(C_1 + C_1')\mathbf{1}$, $s_3 = \mathbf{1}'(C_3 + C_3')\mathbf{1}$이다.

증명은 정리 3.2~정리 3.5에서와 같은 이유로 생략하였다.

4. 수치 예

이론적으로는 무한히 많은 l/G 교체 디자인이 존재할 수 있으나, 미국에서 현재 사용되고 있는 MRTS (3개의 교체그룹)와 SIPP(4개의 교체그룹)를 중심으로 일반화복합추정량의 분산 및 MSE를 상관계수, 편향의 형태, 분산의 형태, 설계간격의 길이에 따라 분석하였고 일반화복합추정량의 특수한 경우인 단순추정량을 기준으로 하여 상대효율을 비교하였다. 표본의 중복으로 발생하는 상관관계는 조사월의 차이 t'에만 의존한다는 가정 하에 지수적 감소형태를 고려하여, $\rho_{t'} = \rho^{t'}$, $t' = 1, 2, \cdots$, $\rho = 0.4, 0.6, 0.8$을 부여하였다. 교체그룹의 분산은 변화에 따른 효과를 알아보기 위하여 $j = 0, \cdots, l-1$에서 $Var(x_t^{(j)}) = (\sigma_j)^2 = (10)^2$의 동일한 패턴(common pattern), $(10 + 2(j))^2$의 증가하는 패턴(increasing pattern), $(10 - 2(j))^2$의 감소하는 패턴(decreasing pattern) 세 가지를 고려하였다. l/G 교체표본디자인에서는 4개의 추정량 중 일정기간과 합과 합의 변동에 관심이 있으므로 다음의 $y_t, y_t - y_{t-1}$의 가중치로는 각각 1/8, $\bar{S}_t^{12}$는 1/2, $\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$는 1/4의 가중치를 부여하였으며 선형결합된 추정량을 최소로 하는 일반화복합추정량의 분산 및 MSE를 구하였다.

표4.1은 var(일반화복합추정량)/var(단순추정량)을 계산한 결과로 대부분 1 보다 작게 나오는 것을 알 수가 있으며 이는 일반화복합추정량이 단순추정량보다 분산이 작고 효율적인 추정량임을 의미한다. 그러나 동일한 분산의 패턴을 가지는 경우 $\bar{S}_t^{12}$과 $\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$에 대해서는 상대효율이 1에 가깝게 나와 단순추정량과 일반화복합추정량에 효율의 차이가 없음을 알 수 있다. 동일 분산의 경우와는 달리 분산의 증가, 감소형태에서는 설계간격이 작을수록 일반화복합추정량의 효율이 단순추정량에 비하여 좋은 것으로 나타났다.

표 4.1: 단순추정량에 대한 일반화복합추정량의 상대효율

	common variance								
	2/4 design			3/4 design			4/4 design		
Var	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$
y_t	0.96	0.90	0.78	0.97	0.92	0.85	1.00	1.00	1.00
$y_t - y_{t-1}$	0.94	0.84	0.61	0.94	0.81	0.55	1.00	1.00	1.00
$\bar{S}_t^{12}$	1.01	1.02	1.02	1.00	1.01	1.01	1.00	1.00	1.00
$\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$	1.00	1.00	1.02	1.00	1.01	1.01	1.00	1.00	1.00
	increasing variance								
	2/4 design			3/4 design			4/4 design		
Var	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$
y_t	0.91	0.83	0.67	0.88	0.81	0.68	0.87	0.84	0.70
$y_t - y_{t-1}$	0.90	0.77	0.51	0.85	0.72	0.49	0.88	0.87	0.87
$\bar{S}_t^{12}$	0.97	0.94	0.84	0.90	0.86	0.69	0.85	0.76	0.51
$\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$	0.96	0.94	0.86	0.90	0.86	0.74	0.85	0.77	0.56
	decreasing variance								
	2/4 design			3/4 design			4/4 design		
Var	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$
y_t	0.93	0.90	0.84	0.83	0.82	0.81	0.61	0.54	0.37
$y_t - y_{t-1}$	0.91	0.84	0.70	0.82	0.77	0.67	0.63	0.61	0.53
$\bar{S}_t^{12}$	0.97	0.97	0.93	0.83	0.79	0.62	0.55	0.41	0.19
$\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$	0.96	0.96	0.94	0.84	0.81	0.69	0.56	0.43	0.23

표 4.2: 4/4 디자인을 기준으로 level에 대한 효과

	common variance								
	1/4 design			2/4 design			3/4 design		
Var	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$
y_t	4.00	4.00	4.00	1.92	1.79	1.57	1.29	1.23	1.14
$y_t - y_{t-1}$	6.67	10.00	20.00	2.52	2.94	3.64	1.53	1.62	1.70
$\bar{S}_t^{12}$	1.93	1.40	1.12	1.37	1.18	1.07	1.11	1.06	1.02
$\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$	2.00	1.47	1.18	1.38	1.19	1.10	1.11	1.06	1.03
	increasing variance								
	1/4 design			2/4 design			3/4 design		
Var	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$
y_t	2.64	2.74	3.29	1.46	1.38	1.34	1.13	1.08	1.10
$y_t - y_{t-1}$	4.26	6.27	11.67	1.87	2.08	2.20	1.31	1.35	1.32
$\bar{S}_t^{12}$	1.33	1.09	1.31	1.10	1.02	1.25	1.00	1.00	1.18
$\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$	1.37	1.12	1.24	1.11	1.03	1.17	1.00	1.00	1.14
	decreasing variance								
	1/4 design			2/4 design			3/4 design		
Var	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$
y_t	12.20	13.68	19.77	4.65	5.02	6.80	2.26	2.50	3.56
$y_t - y_{t-1}$	18.33	26.22	47.97	5.52	6.40	8.35	2.48	2.75	3.48
$\bar{S}_t^{12}$	6.90	6.95	12.29	3.81	4.54	8.68	2.14	2.65	4.42
$\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$	7.03	6.91	10.65	3.79	4.35	7.39	2.14	2.59	4.07

표4.2는 설계간격이 없는 4/4 디자인을 기준으로 회상 개월 수에 따른 디자인의 상대효율을 나타낸다. 표에 나타난 모든 효율 값이 1 보다 크게 나타나고 있어 설계간격이 있는 디자인보다 설계간격이 없는 디자인이 더 효율적인 디자인임을 의미하며 교체그룹의 수와 회상 개월 수의 차가 작으면 작을수록 효율이 더 좋아짐을 알 수 있다. 이는 회상 개월 수가 많을수록 이용할 수 있는 정보가 많아지게 되므로 당연한 결과라 할 수 있다.

표 4.3: 단순추정량에 대한 일반화복합추정량의 MSE 상대효율 : increasing variance

	$\xi = (-10,-7,-4,-1)'$, $\boldsymbol{\tau} = (-10,-5,5,10)'$								
	2/4 design			3/4 design			4/4 design		
MSE	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
y_t	0.71	0.67	0.61	0.61	0.60	0.60	0.68	0.69	0.73
$y_t - y_{t-1}$	0.85	0.77	0.62	1.07	1.11	1.16	0.93	0.92	0.89
$\bar{S}_t^{12}$	0.48	0.43	0.41	0.34	0.36	0.44	0.32	0.40	0.40
$\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$	0.97	0.95	0.91	1.22	1.11	1.04	1.26	1.18	1.15
	$\xi = (-1,-4,-7,-10)'$, $\boldsymbol{\tau} = (-10,-5,5,10)'$								
	2/4 design			3/4 design			4/4 design		
MSE	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
y_t	0.46	0.38	0.27	0.50	0.43	0.33	0.57	0.52	0.42
$y_t - y_{t-1}$	0.69	0.56	0.36	0.89	0.74	0.49	0.93	0.92	0.89
$\bar{S}_t^{12}$	0.68	0.64	0.60	0.36	0.37	0.37	0.20	0.22	0.23
$\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$	1.07	0.99	0.87	0.97	0.87	0.74	0.90	0.76	0.60

표4.3은 일반화복합추정량 대 단순추정량의 비율을 고려한 MSE의 상대효율로 회상편향의 절대값은 교체그룹이 3개일 때에는 $j = 0, \cdots, l-1$에서 회상편향의 절대값이 $\boldsymbol{\xi} = (-7,-4,-1)'$, $\boldsymbol{\xi} = (-1,-4,-7)'$ 으로 감소하는 패턴, 증가하는 패턴을 고려하였고 교체그룹이 4개일 때에는 $\boldsymbol{\xi} = (-10,-7,-4,-1)'$, $\boldsymbol{\xi} = (-1,-4,-7,-10)'$의 감소하는 패턴, 증가하는 패턴을 고려하였다. 교체그룹편향 $\boldsymbol{\tau}$는 교체그룹이 3개일 경우 $\boldsymbol{\tau} = (-5,0,5)'$, 교체그룹이 4개일 경우 $\boldsymbol{\tau} = (-10,-5,5,10)'$ 으로 교체그룹편향의 합이 0이 되는 값을 사용하였다. y_t, $\bar{S}_t^{12}$에서 상대효율이 1 보다 작게 나오는 것을 알 수 있으며 이는 일반화복합추정량이 단순추정량보다 MSE가 작고 효율적인 추정량임을 의미한다. 그러나 $y_t - y_{t-1}$과 $\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$에서는 1 보다 크게 나오므로 단순추정량이 효율적인 추정량임을 알 수 있다. 회상편향의 절대값이 $\boldsymbol{\xi} = (-10,-7,-4,-1)'$으로 감소하는 경우 y_t, $\bar{S}_t^{12}$이 $y_t - y_{t-1}$, $\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$에 비해 단순추정량에 대한 효율이 상대적으로 높음을 알 수 있다. 표의 첫 번째 행의 4/4 디자인의 회상편향의 절대값이 $\xi = (-10,-7,-4,-1)'$으로 감소할 때의 $\bar{S}_t^{12}$와 $\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$ 중 $\bar{S}_t^{12}$의 효율이 상대적으로 높음을 쉽게 볼 수 있다. 이런 현상의 이유는 4개의 추정량 중 $y_t - y_{t-1}$과 $\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$의 추정량은 특성상 회상편향에는 의존하지 않고 교체그룹편향에만 관계가 있기 때문으로 판단된다.

표4.4는 4/4 디자인을 기준으로 수준에 따른 상대효율로 회상편향의 절대값이 증가할 경우에는 표4.2의 패턴과 동일함을 알 수 있다. 그러나 회상편향의 절대값이 감소할 경우에는 $\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$이 1보다 작게 나오며 회상 개월 수가 증가할수록 즉, 설계간격이 작으면 작을수록 디자인의 효율이 더 좋아짐을 알 수가 있다.

표 4.4: 4/4 디자인을 기준으로 level에 대한 효과(MSE) : increasing variance

$\xi = (-10,-7,-4,-1)'$, $\tau = (-10,-5,5,10)'$									
	2/4 design			3/4 design			4/4 design		
MSE	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$
y_t	9.93	9.86	9.33	4.50	4.19	3.48	1.89	1.84	1.75
$y_t - y_{t-1}$	12.05	17.75	34.00	2.64	3.22	4.58	1.61	2.03	3.25
$\bar{S}_t^{12}$	8.92	6.53	4.07	3.16	2.08	1.29	1.57	1.27	1.06
$\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$	0.93	0.73	0.61	0.76	0.68	0.61	0.92	0.85	0.79
$\xi = (-1,-4,-7,-10)'$, $\tau = (-10,-5,5,10)'$									
	2/4 design			3/4 design			4/4 design		
MSE	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$
y_t	5.22	5.72	7.19	1.76	1.57	1.41	1.22	1.14	1.10
$y_t - y_{t-1}$	12.05	17.82	34.02	2.15	2.37	2.69	1.33	1.37	1.37
$\bar{S}_t^{12}$	1.28	1.19	1.24	1.23	1.14	1.10	1.10	1.09	1.06
$\bar{S}_t^{12} - \bar{S}_{t-12}^{12}$	1.31	1.14	1.17	1.18	1.10	1.11	1.03	1.03	1.08

5. 결론

본 논문에서는 설계간격의 존재 여부에 관계없이 다수준교체표본조사를 일반화한 G개의 교체그룹을 갖고 수준 l을 갖는 l/G 교체 디자인 하에서 GCE의 분산과 MSE를 제시하였다. 또한, GCE의 분산과 MSE를 분산의 형태, 설계간격의 형태에 따라 분석하고 GCE와 GCE의 특수한 경우인 단순추정량의 효율 비교를 현월, 월 변화, 연 합, 연 합의 차에 의해 제시하였다. 수치 예의 결과 일반화복합추정량이 단순추정량보다 효율적인 추정량이며 설계간격이 존재하지 않는 디자인이 설계간격이 있는 디자인보다 효율이 더 좋음을 볼 수 있었다.

참고문헌

[1] 박유성, 문원기, 김기환. (2002). 3개의 교체그룹을 가지는 2수준 교체표본설계에서의 복합추정량에 관한 연구, <응용통계연구>, 제 15권 1호, 45-55.

[2] 박유성, 배경화, 김기환. (2001). P-수준 교체표본에서 교체그룹내 상관관계를 고려한 일반화 복합추정량, <응용통계연구>, 제 14권 1호, 81-90.

[3] Bailar, B. (1975). The Effects of Rotation Group Bias on Estimates from Panel Survey, *Journal of the American Statistical Association* 70, 23-30.

[4] Bershad, M. A. (1953). The Sample Survey of Retail Stores, *Time in Sample Survey Methods and Theory, vol. I Methods and Applications,* (edited by Hansen, Hurwitz and Madow), John Wiley & sons, New York, 516-557.

[5] Breau, P. and Ernst, L.R. (1983). Alternative Estimators to the Current Composite Esti-

mator, *Proceedings of the American Statistical Association, Section on Social Statistics,* 397-402.

[6] Cantwell, P. J. (1990). Variance Fomulae for Composite Estimators Rotation Designs, *Survey Methodology,* 16, 153-163.

[7] 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.

[8] 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 the American Statistical Association,* 96, 1483-1496.

[9] Rao, J. N. K. and Graham, J. E. (1964). Rotation Design for Sampling on Repeated Occasions, *Journal of the American Statistical Association,* 59, 492-509.

[10] Wolter, K. M. (1979). Composite Estimation in Finite Populations, *Journal of the American Statistical Association,* 74, 604-613.

[2003년 5월 접수, 2003년 9월 채택]

Generalized Composite Estimators and Mean Squared Errors for l/G Rotation Design

KeeWhan Kim [1)] YouSung Park [2)] JaeEun NamKoong [3)]

ABSTRACT

Rotation sampling designs may be classified into two categories. The first type uses the same sample unit for the entire life of the survey. The second type uses the sample unit only for a fixed number of times. In both type of designs, the entire sample is partitioned into a finite number($=G$) of rotation groups. This paper is generalization of the first type designs. Since the generalized design can be identified by only G rotation groups and recall level l, we denote this rotation system as l/G rotation design. Under l/G rotation design, variance and mean squared error (MSE) of generalized composite estimator are derived, incorporating two type of biases and exponentially decaying correlation pattern. Compromising MSE's of some selected l/G designs, we investigate design efficiency, design gap effect, ans the effects of correlation and bias.

Keywords: multi-level rotation sampling; design gap; generalised composite estimator; bias.

1) Post-Doc., Institute of Statistics, Korea University. E-mail: korpen@korea.ac.kr
2) Professor, Dept. of Statistics, Korea University. E-mail: yspark@korea.ac.kr
3) Graduate Student, Dept. of Statistics, Korea University. E-mail: statis1000@hotmail.com