응용통계연구 제 **15**권 *1*호, *2002*년, *pp. 45-55*

3개의 교체그룹을 갖는 2수준 교체표본설계에서의 복합추정량에 관한 연구

박유성[1] 문원기[2] 김기환[3]

요 약

미국 통계국(Bureau of Census)에서 실시하고 있는 소매동태조사(Monthly Retail Trade Survey)를 중심으로 3개의 교체그룹을 갖는 2수준 교체표본설계를 설명하고, 특성치의 추정을 위한 복합추정량을 다루었다. 설계간격(design gap)이 존재하는 2수준 교체표본설계에서의 일반화복합추정량(generalized composite estimator)과 순환복합 추정량(recursive composite estimator)을 제시하였고, 각 추정량의 분산식을 유도하였다. 또한 공분산구조와 관련된 응답변동(response variability)과 반복응답으로 부터 연유되는 상관관계를 고려하여 복합추정량 사이의 효율을 비교하였다.

주요용어: 다수준교체표본조사, 설계간격, 일반화 복합추정량, 순환복합추정량

1. 서 론

교체표본조사(rotation sample survey)란 시간에 따라 변화하는 모집단의 특성을 조사하기 위한 방법인 반복조사(repeated survey) 방법 중의 하나로 조사시간, 조사비용을 줄이면서도 효율적인 추정과 안정적인 시계열 자료의 구축이 가능한 방법이다. 이러한 교체표본조사는 크게 일수준교체표본조사(one-level rotation sample survey)와 다수준교체표본조사(multi- level rotation sample survey)로 나누어 진다. 이들의 구분기준은 응답자가 한번의 조사에서 과거 어느 시점까지의 자료를 보고하는가에 의한 것으로 일수준교체표본조사의 대표적인 예는 미국 통계국의 CPS(Current Population Survey), 캐나다 통계국의 LFS(Labour Force Survey)이며 다수준교체표본조사의 대표적인 예는 미국통계국의 2수준 교체표본조사인 MRTS(Monthly Retail Trade Survey)와 4수준 교체표본조사인 SIPP(Survey of Income and Program Participation)가 있다. 교체표본조사는 교체표본설계(rotation design)을 어떻게 하느냐에 따라 다양한 형태가 만들어질 수 있고, 특수한 표본구조를 반영하여 모집단의 특성치를 어떻게 추정할 것인가가 주요 관심사가 된다. 교체표본설계에 대한 연구는 Eckler(1955), Rao와 Graham(1964), Cantwell(1990)과 Park과 Kim(2000)에 의하여 수행되었다. 교체표본을 이용한 모집단 특성 추정에는 주로 복합추정

1) (130-070)서울특별시 성북구 안암동 5-1, 고려대학교 통계학과, 교수
E-mail: yspark@mail.korea.ac.kr
2) (130-070)서울특별시 성북구 안암동 5-1, 고려대학교 통계학과 대학원
E-mail: wkmoon@kustat.korea.ac.kr
3) (130-070)서울특별시 성북구 안암동 5-1, 고려대학교 통계연구소, 연구원
E-mail: korpen@ahanet.co.kr

량(composite estimator)이 사용되고 있으며 이는 표본의 조사와 교체과정에서 발생하는 중복표본의 정보를 이용하여 특성 추정량의 효율을 높이는 것으로 다양한 형태가 존재하지만 이들 중 일반화복합추정량(Generalized Composite Estimator, GCE)의 효율이 가장 좋은 것으로 알려져 있다. 교체표본조사에서 사용되는 GCE에 대한 연구는 Breau와 Ernst(1983), Cantwell(1990), Kim과 Park(2000), 박유성외 2인(2001)등이 있다.

본 연구에서는 3수준 교체표본의 형태를 갖으면서도 실제로는 매 조사시점마다 2회의 보고만을 하도록 설계된 MRTS에서 일반화복합추정량을 유도하고 일반화 복합추정량이 수반하는 현실적인 문제를 고려한 순환복합추정량(Recursive Composite Estimator, RCE)을 새롭게 제안하였다. 2절에서는 MRTS에 적용된 교체표본설계의 이해를 위하여 구체적인 설명을 하였으며, 3절에서는 MRTS에서의 GCE 및 RCE를 정의하고 이들의 유도과정을 다루었다. 4절에서는 추정량들의 효율을 비교하였으며, 5절 에서는 각 추정량들의 효율을 비교하고 종합적인 결과를 정리하였다.

2. 소비자동태조사(MRTS)의 디자인

MRTS는 미국 전지역에 걸쳐서 최종소비자에게 상품을 팔고 그에 관련된 서비스를 제공하는 소배점이나 소매회사를 조사대상으로 하고 있으며, 소매점이나 소매회사의 판매액에 대한 추정치를 제공하기 위해서 현재 13,300개의 소매업체를 대상으로 우편조사를 실시하고 있다. 새로운 표본은 5년마다 선택되어 조사되고 최근에 선택된 표본은 1997년 4월부터 조사되고 있다.

그림 2.1: 3개의 교체그룹을 가지는 2수준 교체표본설계

조사월	교체그룹 1	교체그룹 2	교체그룹 3
⋮	⋮	⋮	⋮
$t-5$		$x_{t-5,1}$	$x_{t-5,2}$
$t-4$	$x_{t-4,2}$		$x_{t-4,1}$
$t-3$	$x_{t-3,1}$	$x_{t-3,2}$	
$t-2$		$x_{t-2,1}$	$x_{t-2,2}$
$t-1$	$x_{t-1,2}$		$x_{t-1,1}$
t	$x_{t,1}$	$x_{t,2}$	

MRTS의 표본은 크기가 비슷한 3개의 부표본(subsample)으로 이루어지며 이를 교체그룹(rotation group)이라고 한다. 3개의 교체그룹은 서로 독립이라는 가정 하에서 각 월마다 하나의 교체그룹을 조사하고, 한번의 조사에서 현월과 전월의 매출액을 보고하게 된다. 이러한 형식의 조사를 각 교체그룹에 대하여 3개월 마다 반복적으로 조사하게 된다. 그림

2.1은 MRTS의 교체표본설계를 표현한 것으로 이에 의하면 조사월 t에서 $x_{t,i}, i = 1, 2$는 $(t+i-1)$월에서 보고된 t월의 모집단 특성치에 대한 단순불편추정량을 나타낸다. 즉 조사월 $t-3$에서의 $x_{t-3,1}$은 교체그룹 1에서 $t-3$월에 보고한 자료에 의해 구해진 특성추정치를 나타내며, $x_{t-3,2}$는 교체그룹 2에서 $t-2$월에 회상에 의하여 보고한 자료에 의해 구해진 특성추정치를 나타낸다(이후 내용전개의 단순화를 위하여 $x_{t,i}$ 를 관찰치로 표현하도록 하겠다). 교체그룹 1의 경우 $t-3$월에 $x_{t-3,1}$과 $x_{t-4,2}$에 대한 자료를 제공하고 3개월 후인 t월에 다시 조사에 참여하게 된다. 이때 교체그룹 1이 조사에서 제외되는 3개월을 설계간격(design gap)이라고 하며 설계간격은 교체그룹의 응답부담(response burden) 을 줄여주는 역할을 하게 된다. 설계간격에 의하여 교체그룹이 조사에서 제외되는 기간에는 나머지 교체그룹들이 조사에 참여하게 된다. 그러므로 MRTS는 매 조사에서 각 교체그룹이 현월과 전월에 대한 자료를 제공하고 3개의 교체그룹이 반복적으로 조사되므로 3개의 교체그룹을 갖는 2수준 교체표본설계에 의한 조사라고 할 수 있다.

3. 3개의 교체그룹을 갖는 2-수준교체표본설계에서의 복합추정량

3.1. 일반화 복합추정량

GCE의 형태는 1983년 Bureau와 Ernst에 의하여 제시되었으며, 1990년 Cantwell은 설계간격이 없는 p-수준 교체표본설계하에서의 GCE 및 이의 분산식을 제시하였다. 본 논문에서 고려하고 있는 3개의 교체그룹을 갖는 2-수준 교체표본설계에 적용 가능한 GCE는 다음과 같은 형태로 정의할 수 있다.

$$\begin{aligned} y_t &= \sum_{i=1}^{2} a_i x_{t,i} - k \sum_{i=1}^{2} b_i x_{t-1,i} + k y_{t-1} \\ &= \mathbf{a}'\mathbf{x}_t - k\mathbf{b}'\mathbf{x}_{t-1} + k y_{t-1} \end{aligned} \tag{3.1}$$

여기에서 $\mathbf{x}_t = (x_{t,1}, x_{t,2})'$, $\mathbf{a} = (a_1, a_2)'$, $\mathbf{b} = (b_1, b_2)'$이고 $\mathbf{a}, \mathbf{b}$ 는 제약조건 $\mathbf{a}'\mathbf{1} = \mathbf{b}'\mathbf{1} = 1$을 만족하는 계수이며 k는 0과 1 사이의 상수이다. 식 (3.1)에서 관찰치 $x_{t,i}$에 의하여 발생하는 공분산구조는 다음과 같이 가정하게 된다. $i, i' = 1, 2$ 이고 $t, t' = 0, \pm1, \pm2, \cdots$ 에 대하여

$$Cov(x_{t,i}, x_{t',i'}) = \begin{cases} d_i^2\sigma^2, & t = t' \text{ 이고 } i = i' \text{ 인 경우.} \\ \rho_{|t-t'|} d_i d_{i'} \sigma^2, & t \neq t' \text{ 이고 } x_{t,i}\text{와 } x_{t',i'}\text{이 같은 교체그룹 내의} \\ & \text{관찰치인 경우.} \\ 0, & \text{나머지 경우.} \end{cases} \tag{3.2}$$

식 (3.2)는 교체그룹간에는 서로 독립이고, 교체그룹내에서의 상관관계는 조사월의 차이에만 의존한다는 정상상관관계에 의하여 정의된 것으로, 이 가정은 Fuller외 2인(1992)에 의하여 실증적으로 뒷받침되고 있다. 여기에서 d_1, d_2는 각각 임의의 조사월에서 동시에 보고되는 현월과 전월 관찰치의 응답변동을 나타내며 $d_1 < d_2$의 조건을 갖는다. ρ는 같은 교체

그룹에 속하는 관찰치 사의의 상관관계를 나타낸다. 식 (3.2)를 행렬을 이용하여 재표현하면

$$\begin{cases} Var(\mathbf{x}_t) = \sigma^2\mathbf{D} \\ Cov(\mathbf{x}_t, \mathbf{x}_{t-r}) = \sigma^2\rho_r\mathbf{D}\mathbf{Q}\mathbf{J}^r\mathbf{Q}'\mathbf{D} \end{cases} \tag{3.3}$$

여기에서 $D = diag(d_1, d_2)$는 응답변동을 나타내는 행렬이며, $\mathbf{J}$는 3개의 교체그룹이 3개월마다 반복적으로 조사됨을 나타내는 순열행렬(permutation matrix)이며, $\mathbf{Q}$는 임의의 조사월에서 자료가 보고되지 않는 교체그룹을 나타내는 행렬로 아래와 같이 정의 된다.

$$\mathbf{J} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \qquad \mathbf{Q} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

이제 식 (3.1)의 GCE 정의와 식 (3.3)의 공분산구조에 의하여 GCE의 분산을 구할 수 있다.

정리 3.1 3개의 교체그룹을 가지는 2수준 교체표본설계에서 공분산구조가 식 (3.3)을 따를 때

$$Var(y_t) = \frac{\sigma^2}{1-k^2}\{\mathbf{a}'\mathbf{D}^2\mathbf{a} + k^2\mathbf{b}'\mathbf{D}^2(\mathbf{b} - 2\mathbf{a}) + 2(\mathbf{a} - k^2\mathbf{b})'\mathbf{D}\mathbf{Z}\mathbf{D}(\mathbf{a} - \mathbf{b})\}$$

이며, 여기서 $\mathbf{Z} = \sum_{i=1}^{\infty} k^i\rho_i\mathbf{Q}\mathbf{J}^i\mathbf{Q}'$이다.

일반적으로 시계열자료 생성을 목적으로 하는 표본조사에서는 특정 월의 특성치 추정뿐만 아니라 월변화, 분기변화, 연변화, 분기합, 연합등에 대한 추정에도 많은 관심을 갖게 된다. 변화에 대한 GCE는 $y_t - y_{t-t^*}, t^* \geq 1$로 표현할 수 있고, 합에 대한 GCE는 $S_{t,t^*} = y_t + y_{t-1} + \cdots + y_{t-t^*+1}$로 표현할 수 있다. 이들 추정량의 분산을 구하기 위해 고정된 $t \geq 1$에 대하여 A_t 를 다음과 같이 정의한다.

$$\begin{aligned} A_t = \ & \mathbf{a}'\sum_{i=0}^{t-1} k^i\rho_{t-i}\mathbf{D}\mathbf{Q}\mathbf{J}^{t-i}\mathbf{Q}'\mathbf{D}\mathbf{a} \\ & + \mathbf{a}'\sum_{i=0}^{t-1}\sum_{j=1}^{\infty} k^{i+j}\rho_{t+j-i}\mathbf{D}\mathbf{Q}\mathbf{J}^{t+j-i}\mathbf{Q}'\mathbf{D}(\mathbf{a} - \mathbf{b}) \\ & - \mathbf{b}'\sum_{i=0}^{t-1} k^{i+1}\rho_{t-i-1}\mathbf{D}\mathbf{Q}\mathbf{J}^{t-i-1}\mathbf{Q}'\mathbf{D}\mathbf{a} \\ & - \mathbf{b}'\sum_{i=0}^{t-1}\sum_{j=1}^{\infty} k^{i+j+1}\rho_{t+j-i-1}\mathbf{D}\mathbf{Q}\mathbf{J}^{t+j-i-1}\mathbf{Q}'\mathbf{D}(\mathbf{a} - \mathbf{b}) \end{aligned} \tag{3.4}$$

정리 3.2 정리 3.1과 동일한 가정 하에서 월의 변화와 합에 대한 분산식은 다음과 같다.

(1) $Var(y_t - y_{t-t^*}) = 2(1 - k^{t^*})Var(y_t) - 2\sigma^2 A_t$.

(2) $Var(S_{t,t^*}) = \left\{ t + 2\sum_{i=1}^{t-1}(t-i)k^i \right\} Var(y_t) + 2\sigma^2 \sum_{i=1}^{t-1}(t-i)A_i$

정리 3.1, 정리 3.2의 증명은 Kim과 Park(2000)의 증명과 유사하므로 생략하기로 한다.

3.2. 순환복합추정량

Park외 2인(2001)의 연구에 의하면 일수준 교체표본설계하에서 GCE는 최소분산선형불편추정량(MVLUE)과 비교했을때 거의 동일한 효율을 가지며, 시의성 측면에서는 GCE가 좀 더 우수하다는 것을 지적하고 있다. 시의성 관점에서 식 (3.1)을 보면, t월에서 GCE를 구하기 위해서는 $x_{t,1}$과 $x_{t,2}$가 모두 필요하지만 실제로 $x_{t,2}$는 $t+1$월에서야 얻어지므로 t월에서의 GCE는 $t+1$월까지 조사가 진행되어야만 구할 수 있다는 것을 알 수 있다. 이러한 문제점은 다수준 교체표본조사의 경우 t월에서의 GCE는 실제로 t월 이후에 구할 수 있게 되므로 다수준교체표본조사에서 GCE의 사용은 시의성 결여라는 문제를 항상 수반하게 된다. 이러한 문제의 해결책으로 Wolter(1979)가 제안한 2수준 교체표본조사에서의 초기 및 최종 복합추정량을 고려할 수 있다. Wolter가 제시한 복합추정량들은 다음과 같으며

$$\begin{aligned} y_t^p &= (1-\beta)x_{t,1} + \beta(y_t^{p-1} + x_{t,1} - x_{t-1,2}) \\ y_t^f &= (1-\alpha)x_{t,2} + \alpha y_t^p \end{aligned} \tag{3.5}$$

여기서 α, β는 $0 \le \alpha, \beta \le 1$을 만족하는 상수이다. y_t^p는 초기복합추정량으로 t월까지의 모든 정보를 이용하여 시의성을 확보하고 있으며, $t+1$월에 이르러 추가된 정보 $x_{t,2}$를 이용하여 최종복합추정량 y_t^f를 구하고 있다. Wolter가 제시한 최종복합추정량(이후 WCE라 표현)은 초기복합추정량으로부터 단계적으로 구해지므로 신뢰성이 높은 초기복합추정량을 구하는 것이 매우 중요해지게 된다. 그러나 신뢰성이 높은 초기복합추정량을 구한다는 관점에서 Wolter의 초기복합추정량 y_t^p는 2가지의 유용한 정보를 사용하지 않았다는 단점을 갖고있다. 첫 번째로 y_t^p를 구하는데 y_{t-1}^p보다 정밀한 y_{t-1}^f이 있음에도 사용하지 않았다는 것이며, 두 번째로 y_t^p를 구할 때 $x_{t-1,1}$을 제외하고 $x_{t-1,2}$만을 사용했다는 것이다. 따라서 위의 2가지 문제에 의하여 Wolter가 제시한 초기복합추정량은 정보의 손실이 있게된다.

본 논문에서는 Wolter의 초기복합추정량의 두 가지 단점을 개선한 순환복합추정량(RCE)을 제안하고자 한다. RCE의 초기, 최종복합추정량은 식 (3.6)처럼 정의한다.

$$\begin{aligned} y_{r,t}^p &= (1-\beta)x_{t,1} + \beta(y_{r,t-1}^f + x_{t,1} - c_1 x_{t-1,1} - c_2 x_{t-1,2}) \\ y_{r,t}^f &= (1-\alpha)x_{t,2} + \alpha y_{r,t}^p \end{aligned} \tag{3.6}$$

여기서 α, β는 $0 \le \alpha, \beta \le 1$을 만족하는 상수이며, c_1, c_2는 제약조건 $c_1 + c_2 = 1$을 만족하는 계수이다. 식 (3.6)에서 알 수 있듯이 t월의 초기복합추정량 $y_{r,t}^p$는 t월에는 얻어지는 $t-1$월의 최종복합추정량 $y_{r,t-1}^f$을 이용하게 되고, 이는 다시 $t+1$월에서 얻어지는 t월의 최종복합추정량 $y_{r,t}^f$에 이용되기 때문에 이를 RCE라고 한다. 식 (3.6)의 초기복합추정량 $y_{r,t}^p$에서 $y_{r,t-1}^f$ 대신에 $y_{r,t-1}^p$를 사용하고 $c_1 = 0, c_2 = 1$이면 식 (3.5) Wolter의 초기복합추정량 y_t^p와 같아지게 된다.

LEMMA 3.1 정리 3.1과 동일한 조건에서 식 (3.6)에 정의된 RCE의 분산식은 다음과 같다.

$$Var(y^p_{r,t}) = \frac{\sigma^2}{1-k_p^2}\left\{\mathbf{a}_p'\mathbf{D}^2\mathbf{a}_p + k_p^2\mathbf{b}_p'\mathbf{D}^2(\mathbf{b}_p - 2\mathbf{a}_p) + 2(\mathbf{a}_p - k_p^2\mathbf{b}_p)'\mathbf{DZD}(\mathbf{a}_p - \mathbf{b}_p)\right\} \quad (3.7)$$

$$Var(y^f_{r,t}) = \frac{\sigma^2}{1-k_f^2}\left\{\mathbf{a}_f'\mathbf{D}^2\mathbf{a}_f + k_f^2\mathbf{b}_f'\mathbf{D}^2(\mathbf{b}_f - 2\mathbf{a}_f) + 2(\mathbf{a}_f - k_f^2\mathbf{b}_f)'\mathbf{DZD}(\mathbf{a}_f - \mathbf{b}_f)\right\} \quad (3.8)$$

여기에서 $k_p = k_f = \alpha\beta$, $\mathbf{a}_p = (1,0)'$, $\mathbf{b}_p = (c_1^*, 1-c_1^*)'$, $\mathbf{a}_f = (\alpha, 1-\alpha)'$, $\mathbf{b}_f = (c_1, c_2)'$이다.

증명: 식 (3.6)에서

$$\begin{aligned} y^p_{r,t} =& (1-\beta)x_{t,1} + \beta(y^f_{r,t-1} + x_{t,1} - c_1x_{t-1,1} - c_2x_{t-1,2}) \\ =& x_{t,1} + \beta\{\alpha y^p_{r,t-1} + (1-\alpha)x_{t-1,2} - c_1x_{t-1,1} - (1-c_1)x_{t-1,2}\} \\ =& x_{t,1} + \beta\{\alpha y^p_{r,t-1} - c_1x_{t-1,1} - (\alpha - c_1)x_{t-1,2}\} \end{aligned}$$

식 (3.7)에서 $c_1 = \alpha c_1^*$라고 했을 때

$$\begin{aligned} =& x_{t,1} + \alpha\beta\{y^p_{r,t-1} - c_1^*x_{t-1,1} - (1-c_1^*)x_{t-1,2}\} \\ =& (1-\alpha\beta)x_{t,1} + \alpha\beta\{y^p_{r,t-1} + x_{t,1} - c_1^*x_{t,1} - (1-c_2^*)x_{t-1,2}\} \end{aligned} \quad (3.9)$$

이고

$$\begin{aligned} y^f_{r,t} =& \alpha y^p_{r,t} + (1-\alpha)x_{t,2} \\ =& \alpha\{(1-\beta)x_{t,1} + \beta(y^f_{r,t-1} + x_{t,1} - c_1c_{t-1,1} - c_2x_{t-1,2})\} + (1-\alpha)x_{t,2} \\ =& (1-\alpha\beta)\mathbf{a}_f'\mathbf{x}_t + \alpha\beta(y^f_{r,t-1} + \mathbf{a}_f'\mathbf{x}_t - \mathbf{b}_f'\mathbf{x}_{t-1}) \end{aligned} \quad (3.10)$$

이다. 또한 식 (3.1)에서 정의된 GCE는 다음과 같이 재표현된다.

$$\begin{aligned} y_t =& \mathbf{a}'\mathbf{x}_t + k(y_{t-1} - \mathbf{b}'\mathbf{x}_{t-1}) \\ =& (1-k)\mathbf{a}'\mathbf{x}_t + k(y_{t-1} + \mathbf{a}'\mathbf{x}_t - \mathbf{b}'\mathbf{x}_{t-1}). \end{aligned} \quad (3.11)$$

따라서 식 (3.9)과 식 (3.10)의 초기, 최종 RCE는 GCE의 범주에 속하게 됨을 알 수 있다. 예를 들어 $y^f_{r,t}$는 식 (3.11)에서 $k = \alpha\beta$, $\mathbf{a} = (\alpha, 1-\alpha)'$, $\mathbf{b} = (c_1, c_2)'$인 경우이다. 그러므로 정리 3.1에 의하여 초기, 최종 RCE는 각각 식 (3.7)과 식 (3.8)을 따르게 된다. □

RCE의 변화와 합에 대한 추정량 $y^f_{r,t} - y^f_{r,t-t^*}$와 $S_{r,t,t^*} = y^f_{r,t} + y^f_{r,t-1} + \cdots + y^f_{r,t-t^*+1}$의 분산식도 정리 3.2를 이용하고 보조정리 3.1과 동일한 증명과정을 통해서 유도할 수 있

다. 따라서 어떤 고정된 $t \geq 1$에 대하여 $A_{r,t}$를 다음과 같이 정의하기로 한다.

$$\begin{aligned} A_{r,t} = \ & \mathbf{a}'_f \sum_{i=0}^{t-1} k_f^i \rho_{t-i} \mathbf{DQJ}^{t-i}\mathbf{Q}'\mathbf{Da}_f \\ & + \mathbf{a}'_f \sum_{i=0}^{t-1}\sum_{j=1}^{\infty} k_f^{i+j} \rho_{t+j-i} \mathbf{DQJ}^{t+j-i}\mathbf{Q}'\mathbf{D}(\mathbf{a}_f - \mathbf{b}_f) \\ & - \mathbf{b}'_f \sum_{i=0}^{t-1} k_f^{i+1} \rho_{t-i-1} \mathbf{DQJ}^{t-i-1}\mathbf{Q}'\mathbf{Da}_f \\ & - \mathbf{b}'_f \sum_{i=0}^{t-1}\sum_{j=1}^{\infty} k_f^{i+j+1} \rho_{t+j-i-1} \mathbf{DQJ}^{t+j-i-1}\mathbf{Q}'\mathbf{D}(\mathbf{a}_f - \mathbf{b}_f) \end{aligned} \tag{3.12}$$

LEMMA 3.2 정리 3.1과 동일한 가정 하에서 (3.6)에 정의된 RCE의 변화와 합에 대한 분산식은 다음과 같다.

(1) $Var(y^f_{r,t} - y^f_{r,t-1}) = 2(1 - k^t_f)Var(y^f_{r,t}) - 2\sigma^2 A_{r,t}$

(2) $Var(S_{r,t,t^*}) = \{t + 2\sum_{i=1}^{t-1}(t-i)k^i_f\}Var(y^f_{r,t}) + 2\sigma^2 \sum_{i=1}^{t-1} A_{r,i}$.

여기서 $k_f = \alpha\beta$, $\mathbf{a}_f = (\alpha, 1-\alpha)'$, $\mathbf{b}_f = (c_1, c_2)'$이다.

GCE와 RCE의 경우 추정량의 분산을 최소화하도록 교체그룹에 적절한 가중치를 부여하는 형태로 표현되어 있다. 식 (3.11) 를 보면 GCE y_t는 미지의 계수 $\mathbf{a}, \mathbf{b}$, k를 포함하고 있으며 이 중에서 $\mathbf{a}, \mathbf{b}$는 각각 x_t와 x_{t-1}이 구성하고 있는 교체그룹에 부여되는 가중치로 해석되며, k는 GCE 계산시 과거의 관찰치들을 사용하는 정도를 나타낸다. 그러므로 미지의 계수를 결정하는 문제가 중요하게 된다. 계수의 결정은 각 GCE의 분산식을 목적함수로 하고 각 계수에 주어진 제약식을 이용하여 라그랑지승수법을 이용하여 풀 수 있다. 다수준 교체표본조사에서 GCE의 계수 유도과정은 Kim과 Park(2000)에서 제시된 것과 동일하므로 생략하기로 한다.

4. 수치 예

본 절에서는 3개의 교체그룹을 갖는 2수준 교체표본설계하에서 3절에서 설명하였던 복합추정량들의 효율을 비교하였다. 모든 조사월에서 얻어지는 판매액의 분산 σ^2는 동일하게 100으로 가정하였으며, 임의의 조사월에서 보고되는 현월 및 회상에 의해 보고되는 전월 판매액에 대한 응답변동 d_1, d_2는 $d_1 < d_2$를 만족한다는 조건하에서 응답변동이 $d_1 = 0.1, d_2 = 0.2$인 경우와 $d_1 = 0.1, d_2 = 0.4$인 경우를 고려하였다. ρ는 임의의 조사월에서 동시에 보고되는 현월과 전월의 판매액 사의의 상관관계를 나타내며 $0.4 \sim 0.8$의 값을 고려하였다. 시차에 따른 ρ 값의 변화는 서로 다른 교체그룹간에는 상관관계가 존재하지 않고, 동일한 교체그룹의 서로 다른 조사월에서 보고되는 판매액 사이의 상관관계는 조사

월의 차이에만 의존한다는 정상상관관계 가정하에서

$$\rho_{t-i} = \rho_t^i, \quad i = 1, 2, \cdots$$

와 같이 시차가 커질 수록 상관관계가 지수적으로 감소하는 형태를 고려하였다. 복합추정량들의 효율 비교을 위하여 RCE와 WCE의 효율을 GCE를 기준으로 각각 Var(GCE)/Var(RCE), Var(GCE)/Var(WCE)로 정의하였다. 그러므로 효율이 1보다 작다는 의미는 RCE나 WCE가 GCE보다 효율이 떨어짐을 의미하게 된다.

표 4.1: 현월 복합추정량의 효율 비교

응답 변동	복합 추정량	ρ_t = 0.4	0.5	0.6	0.7	0.8
$d_1 = 0.1$	RCE	0.9996	0.9995	0.9996	0.9999	0.9999
$d_2 = 0.2$	WCE	0.9994	0.9990	0.9990	0.9996	0.9996
$d_1 = 0.1$	RCE	0.9999	0.9995	0.9999	0.9999	0.9999
$d_2 = 0.4$	WCE	0.9995	0.9994	0.9995	0.9999	0.9999

표 4.1은 현월 판매액에 대한 복합추정량의 효율을 나타내고 있다. 전체적으로 효율값이 1 보다 작게 나타나고 있어 GCE가 항상 RCE, WCE보다 효율적인 추정량임을 알 수 있다. 이는 시의성의 문제는 있으나 GCE가 표본의 정보를 최대한 활용하는 추정량이라는 점에서 당연한 결과라고 할 수 있다. 그러므로 RCE, WCE등이 GCE와의 효율의 차이가 미미하며 시의성 문제를 극복한 추정량임을 고려해볼 때 매우 고무적인 결과임을 알 수 있다. 특히 RCE의 경우 WCE보다 비교적 높은 효율을 보이고 있으며, 응답변동의 차이가 커지는 경우는 효율면에서 GCE와 거의 차이가 없음을 알 수 있다.

표 4.2: 전월대비변화 복합추정량의 효율 비교

응답 변동	복합 추정량	ρ_t = 0.4	0.5	0.6	0.7	0.8
$d_1 = 0.1$	RCE	0.9997	0.9989	0.9977	0.9957	0.9928
$d_2 = 0.2$	WCE	0.9994	0.9978	0.9944	0.9880	0.9760
$d_1 = 0.1$	RCE	0.9999	0.9996	0.9996	0.9995	0.9988
$d_2 = 0.4$	WCE	0.9994	0.9989	0.9981	0.9960	0.9916

표 4.2의 경우는 이전 월과 현재월의 판매액 변화에 대한 복합추정량의 효율을 보여주고 있다. 표 4.1과 마찬가지로 전체적인 효율은 GCE보다 떨어지고 있는 것으로 나타났으

나, 응답변동의 차이가 적고, 상관관계가 커질 수록 RCE는 WCE보다 높은 효율을 보여주고 있다.

표 4.3: 분기합 복합추정량의 효율 비교

응답 변동	복합 추정량	ρ_t 0.4	0.5	0.6	0.7	0.8
$d_1 = 0.1$	RCE	0.9998	0.9993	0.9987	0.9983	0.9975
$d_2 = 0.2$	WCE	0.9996	0.9987	0.9961	0.9903	0.9801
$d_1 = 0.1$	RCE	0.9998	0.9997	0.9984	0.9889	0.9694
$d_2 = 0.4$	WCE	0.9994	0.9984	0.9963	0.9869	0.9671

표 4.3의 경우는 한 분기의 총판매액에 대한 복합추정량의 효율을 보여주고 있다. 전체적으로 GCE의 효율이 좋은 것으로 나타나며, 앞서 제시된 두 경우와 마찬가지로 WCE 보다는 RCE의 효율이 좋은 것으로 나타나고 있다. RCE와 WCE의 효율의 차이도 표 4.1보다 큰 차이를 보이고 있다. 표로 제시하지는 않았지만 6개월간의 총판매액이나 연간 총판매액을 추정하는 반기합, 연합의 경우도 표 4.3의 경우와 비슷한 경향을 보이고 있다.

5. 결론

다수준 교체표본설계및 다수준 교체표본설계하에서 GCE에 대한 기존의 연구들은 설계간격을 고려하지 않았다. 본 연구에서는 설계간격이 고려된 다수준 교체표본조사인 MRTS를 3개의 교체그룹을 갖는 2수준 교체표본조사로 정의하고 이에 대한 다양한 GCE 의 분산식을 제시하였다. 또한 GCE의 시의성 문제를 해결할 수 있는 RCE를 제안하였으며, GCE, RCE, WCE의 효율비교를 통하여 RCE가 GCE와 거의 동일한 효율을 보이면서도 WCE 보다는 우수한 효율을 갖는다는 것을 현월, 월간변화, 분기합등에 대하여 수치 예를 제시하였다. 일수준 교체표본조사에 비하여 다수준 교체표본조사는 비용이나 시간적인 측면에서 더 효율적이라는 장점이 있다. 이러한 장점에 설계간격까지 적절히 고려한다면, 반복조사의 효과를 살리면서 조사의 효율성을 좀더 제고할 수 있을 것이라 생각된다.

참고문헌

[1] Derringer, G. and Suich, R. (1980). *"Simultaneous optimization of several response variables"*, Journal of Quality Technology, 13, 1-45.

[2] Harrington, E. C., Jr. (1965). *"The desirability function"*, Industrial Quality Control, 21, (10), 494-498.

[3] Park, Sung. H. (1981). *"Simultaneous optimization techniques for multi-purpose response functions"*, Journal of Military Operations Research Society of Korea, 7, 118-138.

[4] 박성현, 박준오. (1997). *"Simultaneous optimization of multiple response using weighted desirability function"*, 한국품질경영학회지, 25(1), 56-68.

[2001년 6월 접수, 2001년 9월 채택]

A Study of Composite Estimator in 2-level Rotation Design based on 3 Rotation Groups

YouSung Park[1] Wonki Moon [2] Kee Whan Kim[3]

ABSTRACT

The 2-level rotation design based on 3 rotation groups is discussed in view of Monthly Retail Trade Survey conducted by the Bureau of Census in U.S., and composite estimators for population characteristics are concerned. The generalized composite estimators and the recursive composite estimators are presented at 2-level rotation design with design gap and variance formulas for the composite estimators are provided. Also under the response variability related with covariance structure and correlation structure from repeated response, relative efficiencies of the composite estimators are compared.

1) Associate Professor, Dept. of Statistics, Korea University.
E-mail: yspark@mail.korea.ac.kr
2) Graduate Student, Dept. of Statistics, Korea University.
E-mail: wkmoon@kustat.korea.ac.kr
3) Institute of Statistics, Korea University.
E-mail: korpen@ahanet.co.kr