# 이상적(異常的) 가중치를 줄이는 가중치 조정 방법 연구
# Weighing adjustment avoiding extreme weights

김재광*

가중치 조정은 표본추출 확률의 역수로 계산된 기본 가중치 외에 보정승수를 곱해 줌으로써 추정단계에서 보조변수를 활용하여 추정치의 효율을 높이는 방법이다. 가중치 조정의 대표적인 예로는 사후층화를 들 수가 있는데 이는 회귀추정의 특수한 경우이나 회귀추정보다 계산이 편리하여 실제로 많이 사용되고 있다. 이러한 경우 보조변수를 많이 사용하게 되면 보정승수가 지나치게 크거나 또는 지나치게 작아지게 되는 경우가 있는데 이렇게 되면 추정치의 편향(bias)이 커지게 된다. 본 연구에서는 적절히 추정치의 효율도 높이면서 편향을 줄이는 가중치 조정 방법을 제안한다. 또한 시뮬레이션을 통하여 제안된 추정치의 성질을 확인하였다.

Weighting adjustment is a method of improving the efficiency of the estimator by incorporating auxiliary variables at the estimation stage. One commonly used method of weighting adjustment is the poststratification, which is a special case of regression estimation but is relatively feasible in terms of actual implementation. If too many auxiliary variables are used in the poststratification, the bias of the resulting point estimator is no longer negligible and the final weights may have extreme weights. In this study, we propose a method of weighting adjustment that compromises the efficiency and the bias of the point estimator. A limited simulation study is also presented.

## 1. 서론

표본조사에서 사용되는 대부분의 추정량은 표본 관측치의 가중 합으로 종종 표현되는데 이렇게 관측치의 일차 가중합으로 추정량을 구현하는 방법은 하나의 가중치가 여러 개의 항목에 공통적으로 쓰이게 됨으로써 특히 다목적 조사(multi-purpose survey)의 추론에 편리하다. 또한, 예를 들어 총수입과 같은 항목은 여러 가지 세부 수입 항목들의 합으로 표현되는데 이렇게 세부 수입 항목치에 대한 통계치 합이 총수입 항목의 통계치와 같아짐으로써 일관성있는 통계치가 구현되기 위해서는 그 통계 추정을 일차 가중합으로 구현해야 할 것이다. 이러한 성질을 통계량의 내적 일관성(internal consistency)이라고도 하는데 이는 누구나 같은 결과를 얻을 수 있다는 점에서 특히 자료가 일반에게 공개되는 주요 국가 통계에서 반드시 갖추어야할 중요 성질이라 할 수 있을 것이다.

* 한국외국어대학교 통계학과 교수

이 때 사용되는 가중치는 그 표본의 추출확률의 역수로 계산되는 것이 일반적인 방법이다. 이렇게 해서 구현되는 통계량을 Horvitz-Thompson 추정량(이하 HT 추정량)이라고 불리는데 이 추정량의 가장 큰 장점은 추정량의 편향(bias)이 없다는 점이다. 그러나 이 HT 추정량은 표본이 모집단의 주요 특성을 대표하지 못할 경우 그 주요 특성들에 대해서 현실과는 동떨어진 통계량을 구현할 위험이 있고 또 표본 추출 후에 얻어지는 여러 가지 보조정보들을 반영하지 못함으로써 추정치의 효율이 떨어지는 단점이 있다. 이를 극복하지 위해서 표본 추출확률의 역수로 계산된 가중치를 기본가중치(base weight)로 하고 거기에 가중치 보정 승수(weighting adjustment factor)를 곱해 줌으로써 최종 가중치(final weight)를 구현하는 가중치 조정 방법이 등장하게 되었다. 즉, 가중치 조정 방법은 다음과 같이 표현 될 수 있다.

$$\text{최종 가중치}(w_i) \;=\; \text{기본 가중치}(d_i) * \text{보정승수}(F_i)$$

이 때 보정승수를 어떻게 결정하는가는 여러 가지 방법이 있겠지만 기본적으로는 보조변수에 대한 모집단 값을 알 때 그 보조변수의 모수(예를 들면 모평균)에 대해 일치하는 통계량을 구현하도록 최종 가중치를 결정하는 것이 일반적이다. 즉, 보조변수를 **x** 라고 하고 표본원소 집합을 S, 모집단 원소 집합을 U 라고 하면

$$\sum_{i \in S} w_i x_i = \sum_{i \in U} x_i \qquad (1)$$

을 만족하도록 최종 가중치($w_i$)를 결정해 준다. 위의 성질을 북미에서는 주로 Calibration property 라고 하고 유럽에서는 Benchmarking 이라고도 하는데 그 의미는 우리가 확실히 알고 있는 모집단의 보조 정보에 대해서는 오차가 없는 통계가 구현된다는 입장에서 바람직한 성질이라 할 수 있을 것이다. Cochran (1977) 등에서 다루는 비추정(ratio estimation) 이나 회귀추정(regression estimation) 등은 (1)의 성질을 만족하는 대표적인 추정법이라 할수 있다. 이러한 성질은 위에서 살펴본 HT 추정량의 두가지 단점, 즉 표본의 대표성 결여시의 위험성과 추정량의 효율성 결여, 에 대한 해결책을 제시해 준다. 즉, X 가 모집단의 주요 특성을 나타낸다고 할 때 추정치가 (1)을 만족하면 가중치 작업을 통해서 그 변수에 대한 대표성을 확보할 수 있고 또한 우리가 관심변수(Y)와 상관관계가 높은 X에 대하여 (1)의 성질을 만족시키도록 추정치를 구현하면 Y의 추정량에 대한 분산이 현저히 줄어들게 된다.

이러한 바람직한 성질에도 불구하고 (1)을 만족시키는 가중치 조정법에도 한계점이 있게 된다. 첫째로 최종가중치가 기본가중치보다 너무 동떨어진 값으로 변해버린 경우 추정

량의 불편성(unbiasedness)이 깨어진다는 단점이 있다. 즉, 분산은 줄어들어도 편향(bias)이 늘어나게 된다는 점이다. 이를 위해서 최종가중치와 기본가중치의 거리를 최소화하는 제약조건을 넣어주는 것이 일반적이다. 즉, 제약조건 (1)을 만족하는 것 중에서

$$\text{minimize} \sum_{i \in S} Q(d_i, w_i) \tag{2}$$

을 만족하는 최종가중치를 찾는 방법론을 생각해 볼 수 있는데 이 때, $Q(d_i, w_i)$ 는 기본가중치$(d_i)$와 최종 가중치$(w_i)$ 의 거리를 정의하는 함수이다. Deville and Särndal (1992)은 여러 가지 형태의 거리함수 Q(,)에서 가중치 구현 방법에 대해 소개하였다. 그러나, 보조변수의 개수가 많아질 경우 이러한 제약 조건 (1)의 해 자체가 존재하지 않을 수 있고 또 있더라도 이상적인 값을 갖는 최종가중치가 발생할수 있는데 가장 극단적인 경우는 최종가중치가 음수가 되는 경우일 것이다. Huang and Fuller (1978)은 이 경우 최종가중치의 최대값과 최소값을 주고 그 범위 내에서 (2)의 조건을 만족시키는 알고리즘을 제안하였다. 그 이후로 여러 학자들에 의해서 많은 연구가 진행되었고 최근 Fuller (2002) 는 이러한 분야에 대한 연구들을 정리하였다.

대략적으로 말해서 가중치 작업에는 두 가지 다른 목표가 있게 된다. 첫 번째 목표는 추정량의 불편성이다. 이는 최종 가중치가 기본 가중치와 같은 경우에 가장 잘 성취가 될 것이고 이는 또한 제약조건 (1)이 없이 (2)를 만족하는 해를 찾는 것과 동일한 문제이다. 두 번째 목표는 추정량의 효율성이다. 이는 제약조건 (1)을 만족하는 X를 많이 넣으면 넣을수록 추정량의 효율성의 측면에서는 더 효과를 보게 된다.

이 두 가지 상충적인 목표를 달성하기 위해서 어느 하나를 희생하느냐에 따라 크게 두 가지 접근법으로 분류할 수 있다. Huang and Fuller (1978) 과 Deville and Särndal (1992) 는 제약 조건 (1)을 만족하는 것 중에서 (2)를 찾는 방법론을 고려하였으므로 불편성을 희생하고 추정량의 효율성을 우선시한 것으로 분류할 수 있고 반면 Rao and Singh (1997) 과 Chen et al (2001) 은 제약 조건 (1)을 다음과 같이 바꾸어서

$$\left\| \sum_{i \in S} w_i x_i - \sum_{i \in U} x_i \right\| < \delta$$

추정량의 효율성을 희생하여 불편성을 우선시한 것으로 분류할수 있다.

본 연구에서는 그 두개의 관점이 어떻게 절충될 수 있는지를 살펴보고자 한다. 특히 사후층화처럼 일반적으로 많이 사용되고 있는 방법에서 제안된 방법론이 실제적으로 어떻게

구현될 수 있는지를 다루고 마지막 절에서는 시뮬레이션을 통하여 그 방법론의 효과를 살펴보기로 한다.

## 2. 방법론

이해를 돕기 위하여 간단한 예로 시작하고자 한다. 제약 조건이

$$\text{minimize} \sum_{i \in S} d_i(w_i - d_i)^2 \qquad \text{subject to} \sum_{i \in S} w_i(1, x_i) = (1, \bar{x}_N)$$

이라고 할 때 그 해는 Lagrange multiplier method를 이용하여 다음과 같이 얻어진다.

$$w_i = d_i + (\bar{x}_N - \bar{x}_n)\frac{d_i(x_i - \bar{x}_n)}{\sum_{i \in S} d_i(x_i - \bar{x}_n)^2}, \tag{3}$$

이고 $\bar{x}_N = \frac{1}{N}\sum_{i \in U} x_i$ .

이때 $\bar{x}_n = \sum_{i \in S} d_i x_i$ 이다. (3)의 형태에서 볼 수 있듯이 x 값들의 변동이 크면 클수록 최종 가중치의 변동이 커지고 심지어는 음의 값을 가지는 최종 가중치도 가능하다. 예를 들어, 단순임의추출 하에서는 $d_i = 1/n$ 이고 이 경우에 특정 원소의 x 값이 그 표본평균보다 훨씬 작으면 음의 최종 가중치를 가질 수 있게 된다. 이를 피하기 위해서 우선적으로 생각해 볼수 있는 방법이 제약조건에 $w_i \geq 0$ 을 넣고 계산하는 방법이다. 이러한 알고리즘은 quadratic programming이라는 방법으로 구현되는데 Husain (1969)이 이미 그 방법을 적용하였고 Huang and Fuller(1978) 는 일반적인 벡터 x 에서의 방법론을 포트란 프로그램으로 구현하였다. 그러나 이러한 알고리즘은 보조변수가 많아질 때 그 프로그래밍이 복잡해진다는 단점이 있다. 다른 방법으로는 제약조건을 완화시키는 방법이다. 실제로 위의 경우에

$$\text{minimize} \sum_{i \in S} d_i(w_i - d_i)^2$$

subject to

$$\sum_{i \in S} w_i = 1 \qquad \text{and} \qquad -\delta < \sum_{i \in S} w_i x_i - \bar{x}_N < \delta \tag{4}$$

로 풀면 적당한 $\delta$ 값에서 항상 양수가 되는 최종가중치를 구현할 수 있다. 이 때의 $\delta$는 보조변수 x 에 대한 (오차)허용값(tolerance level)이라고 이름지워 질수 있는데 그 의미는 우리가 허용할 수 있는 보조변수 x의 추정치 오차 범위로 해석할 수 있다. 이 오차허용값이 0에 가까우면 우리는 엄격하게 보조변수 x 에 대한 calibration property를 요구하여 (3)의 해를 갖고 이 오차허용값이 무한대에 가까우면 $w_i = d_i$ 를 갖게 될 것이다. 이러한 오차허용값이 주어졌을 때의 일반적인 최종 가중치 구현 방법은 Chen et al(2001)등에 의해서 연구되었으나 이 오차허용값을 주어진 데이터에서 어떻게 구하는지는 아직 연구되지 않았다.

그렇다면 어떠한 오차허용값을 사용해야 할 것인가 ? 이 문제에 대한 직관적인 대답은 아마도 이럴 것이다: 만약 보조변수 x 가 관심변수 y 에 높은 상관관계가 있다면 이 $\delta$는 작아야 할 것이고 그렇지 않다면 우리는 이 $\delta$를 큰 값으로 허용하여도 좋다. 즉, 주요 관심 변수와의 상관관계가 고려되어야 할 것이다. 좀 더 구체적으로 알아보기 위하여 조금 다른 접근을 시도하자.

일반적으로 두 가지 종류의 보조 변수를 생각하자. 가령 x를 엄격하게 calibration 이 되기를 원하는 보조변수라 하고 z를 calibration 적용이 완화될 수 있는 보조변수라고 하면 x 만을 calibration equation 에 적용하여 얻어진 추정량 $\hat{\theta}_1$과 (x,z)를 calibration equation 에 적용하여 얻어진 추정량 $\hat{\theta}_2$을 얻을 수 있는데 이 경우 보조변수 z를 고려할 것인가 아닌가 하는 방법은 변수 선택의 관점에서 접근할 수도 있고 ( Silva and Skinner, 1997) 또는 능형회귀(Ridge regression)와 같은 Shrinkage 방법을 사용할 수도 있다. (Rao and Singh, 1997) 여기서 우리는 Shrinkage 방법의 한 형태로서 다음과 같은 복합 추정량 (composite estimator)를 생각해 볼수 있을 것이다.

$$\hat{\theta}_\alpha = \alpha \hat{\theta}_1 + (1-\alpha)\hat{\theta}_2 \tag{5}$$

즉, $\hat{\theta}_1$과 $\hat{\theta}_2$과의 가중평균의 형태로 복합추정량을 정의하면 적절한 계수 $\alpha$에 대하여 최적의 성질을 가지는 추정량을 구현할 수 있다. 여기서 $\alpha$는 $\hat{\theta}_2$ 을 $\hat{\theta}_1$ 방향으로 값을 보정해주는 Shrinkage 계수로도 불리 울 수 있다. 이 복합추정량의 분산을 최소화하도록 그 계수를 구하면

$$\alpha^* = \frac{Var(\hat{\theta}_2) - Cov(\hat{\theta}_1, \hat{\theta}_2)}{Var(\hat{\theta}_1) + Var(\hat{\theta}_2) - 2Cov(\hat{\theta}_1, \hat{\theta}_2)} \tag{6}$$

으로 표현된다. 여기서 복합추정량의 편향은 근사적으로 $n^{-1}$의 크기이기에 무시할 수 있다. 분산과 공분산의 불편추정치를 구하여 대입하면 분산 및 근사적 오차제곱평균(MSE)를 최소화하는 복합 추정량을 구현할 수 있을 것이다.

이러한 복합 추정량은 $\hat{\theta}_1$이나 $\hat{\theta}_2$ 중 택일하여 사용하는 것보다 더 효율적이다. 왜냐하면 $\hat{\theta}_1$과 $\hat{\theta}_2$ 모두 (5)의 형태로 표현되고 (6)에서 정의된 최적계수 $\alpha^*$을 사용한 복합추정량은 그러한 (5)의 형태를 가지는 추정량 중에서 가장 효율적인 추정량이기 때문이다. 여기서 주의할 점은 이 최적 복합추정량이 관측치의 가중 합으로 표현되지 않는다는 점이다. 만약 $\alpha^*$가 우리가 기존 조사 등을 통해서 아는 값이고 $\hat{\theta}_1$과 $\hat{\theta}_2$이 모두 관측치의 가중 합으로 표현된다면

$$\hat{\theta}_{\alpha^*} = \sum_{i \in S} w_i^* y_i$$

로 표현되고 이 때

$$w_i^* = \alpha^* w_{1i} + (1 - \alpha^*) w_{2i} \tag{7}$$

으로 표현될 수 있을 것이다. (여기서 $w_{1i}$ 는 $\hat{\theta}_1$ 에서 사용된 가중치이고 $w_{2i}$ 는 $\hat{\theta}_2$ 에서 사용된 가중치이다.) 이 경우에는 최적 추정치가 관측치의 가중 합으로 표현된다고 말할 수도 있겠지만 실제로는 $\alpha^*$의 추정치를 관측치로부터 구해서 사용하게 되므로 관측치의 비선형 함수가 된다.

이런 의미에서 두 가지 문제점을 생각할 수 있는데 첫째로는 가중치 작업이 관측치를 얻어내기 전에 실시되는 경우 $\alpha^*$의 추정치를 계산할 수 없다는 점이고 둘째로는 가중치 작업에서 관측치를 얻을 수 있다 하더라도 관심변수가 여러 개인 경우에는 최적 $\alpha^*$ 값이 관심변수 마다 달라지므로 (7)로 표현되는 최종 가중치가 변수마다 달라질 수 있다는 점이다. 첫 번째의 경우에 우리가 할 수 있는 방법은 문제의 접근 방법을 가중치 작업으로 보지 않고 일반적인 추정 문제로 확대하여 보는 방법이다. 이러한 경우에는 추정치가 비선형 함수라는 것이 크게 문제가 되지 않는다. 두 번째 문제점에는 우리가 몇 개의 주요

관심변수를 관측하여 각각에서 최적의 $\alpha^*$를 계산한 후에 그 것들의 산술 평균 또는 가중 평균을 사용하여 최종 $\alpha^*$을 결정하는 방법이다. 4절에서 살펴볼 시뮬레이션 결과를 보면 이 방법이 각각의 변수에 다른 최적 $\alpha^*$를 계산하는 방법에 비해 크게 효율이 차이가 나지 않았다.

## 3. 사후 층화에서의 적용

사후 층화는 사후 층에 대한 지시변수들로 구성된 벡터를 보조변수로 사용하는 회귀추정의 특별한 경우이다. 사후 층화를 적용하려면 무엇보다도 먼저 사후 층에 대한 모집단 크기의 참값을 알아야 한다. 총 G 개의 사후 층이 있고 각 사후층 g 에서의 모집단 크기를 $N_g$라고 할 때 사후 층화 추정량은 다음과 같다.

$$\hat{\theta}_{post} = \sum_{g=1}^{G} N_g \frac{\hat{Y}_g}{\hat{N}_g} \tag{8}$$

여기서 $\hat{N}_g$ 는 표본으로부터 얻어진 $N_g$의 추정량이고 $\hat{Y}_g$ 는 사후층 g 에서의 Y의 총계에 대한 추정량이다. 따라서 $F_g = N_g/\hat{N}_g$ 는 사후층화에 의한 기본가중치에 곱해지는 보정승수가 된다. 층의 수(G)가 표본수에 비해 상대적으로 큰 경우에는 이 보정승수 $F_g$ 의 값이 큰 변동을 보이게 되고 극단적인 경우에는 그 사후층에 해당되는 표본이 없는 경우도 있게 된다. 이와는 반대로 층의 수가 너무 작은 경우에는 사후 층화의 효과를 크게 보지 못하게 될 것이다. 따라서 2절에서 다룬 복합 추정법은 두 가지 극단을 어떻게 절충할 수 있을지에 대해 방법론을 제시해 준다. 이 경우의 좀 더 자세한 적용은 4절에서 가상의 데이터를 가지고 시뮬레이션에서 예를 들어서 설명하고자 하므로 여기서는 생략하기로 한다.

복합 추정량의 사후층화에 대한 또 다른 적용으로는 (사후)층화 변수가 여러 개일 때 적용이 될 수 있다. 예를 들어 두 개의 층화 변수(예, 연령대와 성별)가 있다고 하자. 만약 우리가 모집단의 (연령대 * 성별) 에 대한 모집단 크기를 안다면 이 경우는 층화변수가 한 개인 경우로 바꾸어서 사후 층화 추정식 (8)을 적용시킬 수 있을 것이다. 그러나 우리가 연령대 별에 대한 모집단 크기와 성별에 대한 모집단 크기를 각각 알지만 그 성별 연령대별의 각각에 대한 모집단 크기를 알지 못할 경우에는 사후 층화 추정식을 직접 적용하지 못하고 Raking ratio 추정법 등을 사용하는 것이 일반적이다(예 ; Särndal et al, *1992, p283*). 그러나 이보다는 성별을 이용한 사후 층화 추정치 $\hat{\theta}_1$ 와 연령대별을 이용한

사후 층화 추정치 $\hat{\theta}_2$ 를 이용하여 (5)의 형태로 최적 복합추정량을 구현하는 것이 더 효율적일 것이다.

## 4. 시뮬레이션 결과

2절에서 제안된 추정량의 효율을 알아보기 위해서 가상의 모집단(모집단 크기 N=10,000)을 만들었다. 그 모집단 내에서 두개의 이산형 지시변수$(X_1, X_2)$를 다음과 같이 만들었다.

$$X_1 \sim Bernoulli(0.5),\ X_2 \sim Bernoulli(0.5)$$

이 때 $X_1$ 과 $X_2$ 는 독립이다. 이 지시변수 $(X_1, X_2)$는 사후 층을 만드는 변수로 여기면 예를 들어 $X_1$는 남자에 대한 지시 변수로 여길 수 있고 $X_2$는 Young/Old 에 대한 지시 변수로 여길 수 있을 것이다. 이러한 지시 변수 외에도 다음과 같은 가상적인 관심 변수를 만들었다.

$$\begin{aligned} Y_1 &= 5 + 2X_1 + e_1, \quad e_1 \sim N(0, 2^2) \\ Y_2 &= 5 + 2X_2 + e_2, \quad e_2 \sim N(0, 2^2) \end{aligned} \tag{9}$$

이렇게 만들어진 모집단에서 n=50 개의 표본을 단순임의 추출로 뽑아 다음과 같은 추정량들을 계산하였다.

A. 단순 평균
B. 사후 층화 추정량 1 : $X_1$ 만을 이용하여 사후 층화 추정량을 계산함.
C. 사후 층화 추정량 2 : $(X_1, X_2)$ 을 이용하여 사후 층화 추정량을 계산함.
D. 복합 추정량 1 : 사후 층화 추정량 1 과 사후 층화 추정량 2를 이용하여 최적 복합 추정량을 $Y_1$과 $Y_2$에 대하여 각각 계산함.
E. 복합 추정량 2: 사후 층화 추정량 1과 사후 층화 추정량 2를 이용하여 최적 복합 추정량을 $Y_1$에 대해서 구현한 $\alpha$ 값을 $\alpha_1^*$ 이라고 하고 $Y_2$에 대해서 구현한 $\alpha$ 값을 $\alpha_2^*$ 이라고 할 때 최종값을 $\alpha^* = (\alpha_1^* + \alpha_2^*)/2$ 으로 계산하여 사용함.

이렇게 각각의 표본에서 5개의 추정량을 계산하는 것을 B=5,000 번 반복하여 그 값들을 통해서 Monte Carlo 방법으로 평균, 분산, 그리고 평균제곱오차(MSE)를 구하였다. 5가지 모든 추정치의 Monte Carlo Bias 가 유의하지 않았으므로 MSE만 살펴보면 다음과 같다. <표 1>은 표본 평균의 분산을 100으로 놓았을 때의 각 추정치의 상대 MSE값을 보여준다.

<표 1> 표준화된 MSE값의 비교

| 추정량 | $Y_1$ 변수 | $Y_2$ 변수 |
|---|---|---|
| A | 100 | 100 |
| B | 81 | 102 |
| C | 86 | 87 |
| D | 83 | 90 |
| E | 82 | 96 |

위의 결과로부터 다음과 같은 결론을 얻을 수 있었다.

1. 추정량 A 와 그 나머지 추정량을 비교해 보면 알 수 있듯이 대체적으로 사후 추정을 하는 것이 사후 추정을 하지 않는 것 보다 효율적이지만 그렇지 않은 경우도 있었다. $Y_2$ 변수는 (9)의 모형을 따르므로 $X_2$ 에 대한 사후층화에는 효율이 크겠지만 $X_1$ 에 대해 사후 층화를 한 경우(추정량 B를 사용한 경우)에는 전혀 도움이 되지 못하고 오히려 분산이 2% 증가하였다. 이 분산의 증가는 가중치의 변동에 기인하는 것으로도 볼 수 있다.

2. 추정량 B와 C를 비교해 보면 $X_1$ 에 대해 사후 층화를 한 추정량 B의 경우에는 자료의 모형이 $X_1$ 에 대해 유의한 모형인 $Y_1$ 변수의 사후 층화 추정치에는 제일 효율적이지만 그렇지 않은 $Y_2$ 변수의 추정에는 효율이 가장 좋았다. 추정량 C의 경우에는 두 변수 모두에 상당히 효율적이었으나 $Y_1$변수에 대해서는 효율이 다소 떨어지는 편이었다.

3. 추정량 C와 D를 비교해 보면 거의 대등한 효율을 보여준다. 이는 두 추정량 모두 최적효율을 구현하는 추정량이기 때문이고 추정량 E도 추정량 D에 비해서는 효율이 떨어지지만 추정량 B에 비해서는 더 나은 효율을 보여준다. 결론적으로 말하자면 복합 추정량 E는 사후 추정량 B와 사후층화 추정량 C를 절충하는 효과를 보여주었다.

## 참고 문헌

Chen, J., Sitter, R.R., and Wu, C. 2002. "Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys." *Biometrika* 89: 230-237.

Cochran, W.G. 1977. Sampling Techniques, 3rd ed., New York: John Wiley & Sons.

Deville, J. and Samdal, C.-E. 1992. "Calibration estimators in survey sampling." *Journal of the American Statistical Association* 87: 376-382.

Fuller, W. A. 2002. "Regression Estimation for Survey Samples." *Survey Methodology* 28: 5-23.

Huang and Fuller. 1978. "Nonnegative regression estimation for sample survey data." *Proceedings of the social statistics section.* American Statistical Association: 300-305.

Husain. 1969. Construction of regression weights for estimation in sample surveys. Unpublished M.S. thesis, Iowa State University.

Rao, J.N.K. and Singh, A.C. 1997. "A ridge shrinkage method for range restricted weight calibration in survey sampling." *Proceedings of the section on survey research methods.* American Statistical Association, : 57-64.

Särndal, C-E., Swensson, B., and wretman, J. 1992. Model Assisted Survey Sampling. Springer - Verlag.

Silva, P.L.D.N., and Skinner, C.J. 1997. "Variable selection for regression estimation in finite populations," *Survey Methodology* 23: 23-32.