

## 상관관계와 표본 크기에 따른 BLS 무응답 보정의 효율성 비교

김 석<sup>1</sup> · 신기일<sup>2</sup>

<sup>1</sup>한국외국어대학교 통계학과, <sup>2</sup>한국외국어대학교 통계학과

(2009년 9월 접수, 2009년 10월 채택)

### 요약

미국 노동통계청에서 사용하고 있는 BLS 방법의 효율성과 민감성에 관한 연구 결과에 의하면 표본 틀(Sample frame) 자료와 조사된 자료의 상관관계가 높을수록 BLS 무응답 보정 효과는 커지는 것으로 알려져 있다 (이석진과 신기일, 2008). 그러나 표본 틀 자료와 조사된 자료의 상관계수가 층별로 크기가 다른 경우, BLS 보정 효과는 달라질 수 있다. 따라서 일반적으로 실시되는 표본 설계에서는 층화추출 방법이 사용되기 때문에 각 층의 표본 크기와 상관계수가 다른 경우의 BLS 보정 효과를 살펴보는 것은 매우 중요하다. 본 논문에서는 층의 표본 크기와 상관계수 그리고 무응답 비율에 따른 BLS 무응답 보정 효과를 살펴보았다. 이를 위해 사용된 자료는 노동부의 월별 자료인 2007년 매월노동통계 자료이다.

주요어: 표본설계, 가중치, BLS 사후 보정법, 무응답 보정.

### 1. 서론

표본설계에서 고려해야 할 가장 중요한 부분은 모집단을 잘 대표할 수 있도록 표본의 대표성을 확보하는 것이다. 최근 기업의 잦은 휴폐업과 무응답 그리고 이상치 등 조사과정에서 발생하는 비표본 오차는 커지고 있는 실정이다. 따라서 비표본 오차를 줄이기 위한 노력의 중요성은 커지고 있으며 비표본 오차를 줄이는 통계 추정방법은 매우 중요하다고 하겠다. 비표본 오차를 줄이는 하나의 방법인 결측치(Missing value) 대체법에 관한 연구가 활발히 진행되고 있으며 (Little과 Rubin, 1987, 2002) 국내에서도 결측치 대체법에 관한 연구는 활발히 진행되고 있다 (김규성, 2000). 또한 대체법과는 다른 방법인 갈퀴법을 사용함으로써 비표본 오차를 줄일 수 있다. 갈퀴법에 관한 내용은 Oh와 Scheuren (1987) 또는 신민웅과 이상은 (2001)을 살펴보기 바란다. 그러나 비표본 오차를 줄이는 통계 추정방법은 가중치 보정법을 통해서도 이루어질 수 있기 때문에 가중치에 관한 연구는 매우 의미가 있다고 하겠다.

본 논문에서 연구될 보정 방법은 BLS 방법으로 미국 노동통계청(Bureau of labor statistics)이 사업체 조사의 보정을 위해 만든 것이다. 이 방법을 사용하기 위해서는 기본적으로 표본에 주어진 설계가중치와 표본 틀에 보조변수가 있어야 한다. 기본적으로 사용될 수 있는 보조변수로는 종사자수가 있다. 이 두 수치는 큰 어려움 없이 얻을 수 있다. 국내에서도 사업체 조사를 위해 일반적으로 사업체기초통계조사 표본 틀로 사용되고 있으며 이 속에 종사자수가 포함되어 있다.

표본 틀(Sample frame)을 기초로 한 표본 설계에서는 기본 가중치 또는 설계 가중치를 구할 수 있다. 만약 최신의 표본 틀을 사용한다면 조사 시점의 자료와 실사 자료와는 큰 상관상계를 유지할 수 있

본 연구는 2009년도 한국외국어대학교 학술연구비 지원에 의해 이루어진 것임.

<sup>2</sup>교신저자: (449-791) 경기도 용인시 모현면 왕산리 산 89, 한국외국어대학교 통계학과, 교수.

E-mail: keyshin@hufs.ac.kr

다. 이 경우 BLS 보정 방법을 사용하면 무응답 보정의 효과를 높일 수 있다. BLS 보정 방법은 표본 틀의 정보를 이용하기 때문에 이를 이용하지 않고 단순히 조사된 개수를 이용한 일반 방법에 비해 우수한 결과를 주는 것으로 알려져 있다. BLS 보정 방법의 효과를 보기 위하여 이석진과 신기일(2008)은 BLS의 민감도를 살펴보았다. 결과에 의하면 상관계수가 0.9 이상으로 높은 경우에 BLS를 사용하면 MSE를 기준으로 약 30%의 정밀도를 향상시킬 수 있는 것으로 알려졌다. 그러나 이석진과 신기일(2008)은 일반적인 표본 설계에서처럼 전체 자료를 총화하지 않고 분석하였기 때문에 현실적인 결과로 해석하기에는 다소 무리가 있다. 즉 일반적인 표본 설계는 지역별, 규모별, 대분류별 등 특정 층으로 나눈 후 가중치를 구하게 된다. 이에 본 논문에서는 실제적인 표본설계에서처럼 층을 나누어 설계 가중치가 구해졌을 경우를 고려하여 분석하였다. 특히 패널 조사처럼 매년 또는 매월 자료가 얻어질 경우, 전년 또는 전월 자료가 존재하게 되기 때문에 전년 또는 전월 자료를 이용한 BLS 보정 방법이 적용될 수 있다. 이에 본 연구에서는 노동부의 매월노동통계자료 중 1월 자료와 5월 자료를 이용하여 패널 자료와 같은 구조가 될 수 있도록 만들었으며 각 층별로 상관계수 및 표본 크기에 따른 BLS 보정 방법의 효과를 살펴보았다. 따라서 본 논문에서 얻어진 결과는 일반적인 BLS 보정 방법 뿐만 아니라 패널 자료의 무응답 가중치 보정에서도 적용이 가능하다. 2장에는 미국의 BLS 보정 방법을 간단히 설명하였고 3장에는 무응답 보정에 관한 간단한 이론을 살펴보았다. 4장에는 2007년 매월노동통계 자료를 이용하여 실시한 모의실험 결과가 나와있으며 5장에 결론이 있다.

## 2. BLS(Bureau of Labor Statistics) 보정 방법

미국 노동통계청(Bureau of Labor Statistics)의 BLS 방법은 사업체 조사 시 표본 틀의 보조 자료와 실사에서 얻어진 자료 사이에 상관관계가 높은 경우에 사용하는 방법으로 무응답이 랜덤하게 발생(MAR)한 경우에 사용하는 방법이다. BLS 가중치 방법의 가장 대표적인 보정 단계는 다음의 4단계이다. 즉 집계 후 보정(Re-aggregation Adjustment), 무응답 보정(Non-response Adjustment), 이상점 보정(Outlier Adjustment) 그리고 벤치마크 보정(Benchmark Adjustment)이다. 이외에 다른 보정 방법이 발표되었으나 본 연구에서는 대부분의 표본조사에서 가장 많이 활용되는 위의 4단계만을 간단히 설명하였다. BLS 방법에 관한 자세한 내용은 이석진과 신기일(2008) 또는 Burdette(2003)를 참조하기 바란다. 2장에서는 설명을 간단히 하기 위해 층이 하나인 경우인 경우를 살펴보았다.

### 2.1. 집계 후 보정(Re-aggregation Adjustment; REAG)

REAG는 실사 후 표본 틀 자료와 조사 결과와의 차이가 큰 경우(일반적으로 2배 이상인 경우)에 적용하는 보정 방법으로 다음의 보정 공식을 이용하여 가중치를 보정한다.

$$\text{Reag Factor} = f_i^{\text{REAG}} = \frac{\sum y_i^D}{\sum y_i^O}, \quad (2.1)$$

여기서  $y_i^D$ 는 표본설계 시 사용했던 보조변수의 자료 값(조사 전, Design)이며  $y_i^O$ 는 조사 후 얻어진(Observed) 실사자료의 자료 값이다.

### 2.2. 무응답 보정(Nonresponse Adjustment; NRA)

NRA는 가장 많은 관심을 갖는 방법이며 BLS의 무응답 보정공식은 다음과 같다.

$$\text{NRA factor} = f_i^{\text{NRA}} = \frac{\sum^n w_i y_i^D}{\sum^r w_i y_i^D}, \quad (2.2)$$

여기서  $w_i$ 는 표본설계 시 주어진 기본 가중치이고,  $n$ 은 조사시점에서 존재하지 않는 조사대상자(폐업)를 제외한 모든 조사대상자 수를 나타내며,  $r$ 은 조사시점에서 존재하지 않는 조사대상자(폐업)와 무응답을 한 조사 대상자를 제외한 모든 조사대상자 수를 나타낸다. 즉  $r$ 은 유효 표본 크기이다.

### 2.3. 이상점 보정(Outlier Adjustment Factor; OAF)

OAF는 자료가 이상점으로 판명이 났을 때 이 이상점의 가중값을 보정하기 위한 방법으로 이상점에 부여된 가중치를 “1” 값으로 환원하게 만든다. 이를 위한 공식은 다음과 같다.

$$\text{Outlier factor} = f_i^{OAF} = \frac{1}{w_i \times f_i^{REAG} \times f_i^{NRA}} \quad (2.3)$$

또한 이상점이 아닌 자료의 가중치 보정을 위한 공식은 다음과 같다.

$$\text{Non-outlier factor} = f_i^{OAF(no)} = \frac{\text{ANOF}_1 / \text{ANOF}_2}{f_i^{NRA}}, \quad (2.4)$$

여기서  $\text{ANOF}_1 = \text{NOF}_1 - \sum_i y_i^{D,out}$ ,  $\text{ANOF}_2 = \text{NOF}_2 - \sum_i w_i y_i^{D,out}$ 이고,  $\text{NOF}_1 = \sum^n w_i y_i^D$ ,  $\text{NOF}_2 = \sum^r w_i y_i^D$ 이다. 또한  $y_i^{D,out}$ 은 이상점으로 판명된 자료의 자료 값이다.

### 2.4. 벤치마크 보정(Benchmark Adjustment Factor; BMF)

BMF는 가중치 보정 단계 중에서 마지막으로 사용하는 보정 단계로 조사시점에서 얻어진 최신 정보를 기준으로 가중치를 보정한다. 벤치마크 보정방법은 다음과 같다.

$$\text{Benchmark factor} = f_i^{BMF} = \frac{\text{Target value}_i (\text{Benchmark value}_i)}{\text{Reported value}_i} \quad (2.5)$$

즉 Target value 혹은 Benchmark value가 주어지면 조사시점에서 얻은 추정치와의 비율을 이용하여 보정한다.

### 2.5. 최종 가중치(Final Weight)

전술한 4가지 단계를 마친 후 얻어진 최종 가중치는 다음과 같다.

$$w_i^{Final} = w_i \times f_i^{REAG} \times f_i^{NRF} \times f_i^{OAF} \times f_i^{BMF} \quad (2.6)$$

식 (2.6)에 의해 최종가중치가 정해지면 BLS에 의한 추정량,  $\hat{t}_B$ 은 다음과 같다.

$$\hat{t}_B = \sum_{i=1}^r w_i^{Final} y_i^O. \quad (2.7)$$

## 3. 상관계수와 표본 크기에 따른 BLS의 효율성 분석

본 연구에서는 BLS에서 사용하는 4단계 중에서 큰 관심을 끌고 있는 무응답 보정을 살펴보았다. 일반적으로 관심변수와 상관성이 높다고 생각되는 자료가 있을 경우 식 (2.2)을 사용하는 것은 당연하다. 그러나 한번 정해진 가중치는 관심 있는 변수 하나에 적용되는 것이 아니라 다른 관심 있는 변수에도 사용된다. 예를 들어 노동부에서 생산하는 매월노동통계는 상용근로자수를 추정할 뿐 아니라 일인당 평균 근로시간과 일인당 평균 총임금을 추정한다. 일반적으로 상용근로자수와 평균 근로시간 또는 상용근로자수와 일인당 평균 총임금 사이에는 큰 상관관계가 존재하지 않는다. 따라서 상관관계의 크기에 따른 BLS 무응답 보정효과에서 보정에 사용되는 보조변수,  $Y_i^D$ 의 영향력을 알아보는 것은 매우 중요하다.

### 3.1. 무응답 보정에서의 효율성 분석

본 연구에서는 일반적인 표본설계에서 사용되는 방법인 층화추출로 표본설계가 이루어진 경우를 살펴 보았다. 이 경우  $h$ 층의 설계 가중치  $w_{hi}$ 는  $N_h/n_h = w_h$ 로 구해진다. 여기서  $N_h$ 는  $h$ 층의 모집단 크기이고  $n_h$ 는  $h$ 층의 표본 크기이다. 이제 유효 표본 크기를  $r_h$ 라 하자. 그러면 일반적인 무응답 보정 인자를  $f_u$ 라 할 때  $f_u = n_h/r_h$ 이 되고 최종 가중치는 다음과 같다. 여기서  $U$ 는 “일반적인 방법”이라는 것을 표시하기 위해 사용하였다.

$$w_{hi}(U) = w_h(U) = \frac{N_h}{n_h} \times f_u = \frac{N_h}{n_h} \times \frac{n_h}{r_h} = \frac{N_h}{r_h} \quad (3.1)$$

따라서 일반적인 방법을 이용한  $h$ 층의 총계추정치,  $t_h(U)$ 는 다음의 식에 의해서 구해진다.

$$t_h(U) = \sum_{i=1}^{r_h} \frac{N_h}{r_h} y_{hi}^O \quad (3.2)$$

이제 BLS 방법을 이용한 무응답 보정 최종 가중치를 구하자. 이 최종 가중치를  $w_{hi(B)}$ 라 하자. 그러면 층내 가중치가 모두 같다고 가정하면  $w_{hi} = w_h$ 이고 식 (2.2)에 의해

$$w_{hi(B)} = \frac{N_h}{n_h} \times f_{hi}^{NRA} = \frac{N_h}{n_h} \times \frac{\sum_{i=1}^{r_h} w_{hi} y_{hi}^D}{\sum_{i=1}^{r_h} w_{hi} y_{hi}^D} = \frac{N_h}{n_h} \times \frac{\sum_{i=1}^{r_h} y_{hi}^D}{\sum_{i=1}^{r_h} y_{hi}^D} = w_{h(B)} \quad (3.3)$$

가 된다. 따라서 BLS 방법을 이용한 총계추정치  $t_h(B)$ 는 다음과 같이 구해진다. 여기서  $B$ 는 BLS 방법이라는 것을 표시하기 위해 사용하였다.

$$t_h(B) = \sum_{i=1}^{r_h} w_{h(B)} y_{hi}^O = \sum_{i=1}^{r_h} \frac{N_h}{n_h} \times \frac{\sum_{j=1}^{r_h} y_{hj}^D}{\sum_{j=1}^{r_h} y_{hj}^D} y_{hi}^O = \frac{N_h}{n_h} \times \frac{\sum_{j=1}^{r_h} y_{hj}^D}{\sum_{j=1}^{r_h} y_{hj}^D} \sum_{i=1}^{r_h} y_{hi}^O \quad (3.4)$$

**Case 1.**  $y_{hi}^O$ 와  $y_{hi}^D$ 가 완전 상관인 경우

두 자료가 완전 상관인 경우이므로  $y_{hi}^D = c y_{hi}^O$ 로 표시 될 수 있다. 따라서 이 식을 식 (3.4)에 대입하면

$$t_h(B) = \frac{N_h}{n_h} \times \frac{\sum_{j=1}^{r_h} c y_{hj}^O}{\sum_{j=1}^{r_h} c y_{hj}^O} \sum_{i=1}^{r_h} y_{hi}^O = \frac{N_h}{n_h} \times \sum_{i=1}^{r_h} y_{hi}^O \quad (3.5)$$

가 된다. 따라서 무응답의 수에 상관없이 완전한 자료가 얻어진 결과와 같은 결과를 주고 있다.

**Case 2.**  $y_{hi}^O$ 와  $y_{hi}^D$ 에 일부 상관관계가 있는 경우

두 자료의 상관관계가 “0”이 아니므로  $y_{hi}^D = y_{hi}^O + x_{hi}$ 로 표시할 수 있다. 여기서  $y_{hi}^O$ 와  $x_{hi}$ 는 독립이다. 이제  $\text{Var}(y_{hi}^O) = \sigma_{y_h}^2$ ,  $\text{Var}(x_{hi}) = \sigma_{x_h}^2$ 라 표시하고 상관계수를 구하기 위해 공분산을 구하면 다음과 같다.

$$\text{Cov}(y_{hi}^O, y_{hi}^D) = \text{Cov}(y_{hi}^O, y_{hi}^O + x_{hi}) = \text{Var}(y_{hi}^O) = \sigma_{y_h}^2.$$

또한  $\text{Var}(y_{hi}^D) = \text{Var}(y_{hi}^O + x_{hi}) = \sigma_{y_h}^2 + \sigma_{x_h}^2$ 이므로 상관계수  $\rho_{h,xy}$ 는 다음과 같이 구해진다.

$$\rho_{h,xy} = \frac{\sigma_{y_h}^2}{\sqrt{\sigma_{y_h}^2} \sqrt{\sigma_{y_h}^2 + \sigma_{x_h}^2}} = \frac{1}{\sqrt{1 + \sigma_{x_h}^2 / \sigma_{y_h}^2}}. \quad (3.6)$$

다음으로  $\bar{y}_h^O - \mu_{y_h} = \sigma_{y_h} / \sqrt{r_h} z_{y_h}$ ,  $\bar{x}_h - \mu_{x_h} = \sigma_{x_h} / \sqrt{r_h} z_{x_h}$  라 하자. 여기서  $E(y_{hi}^O) = \mu_{y_h}$ ,  $E(x_{hi}) = \mu_{x_h}$  이다. 그러면  $z_{y_h}, z_{x_h}$  는 독립이고 점근적으로 표준 정규분포를 따른다. 이제 BLS 무응답 보정인자를 살펴보자. 그러면

$$f_h^{NRA} = \frac{\sum_{hi}^{n_h} y_{hi}^D}{\sum_{hi}^{r_h} y_{hi}^D} = \frac{\sum_{hi}^{n_h} y_{hi}^O + \sum_{hi}^{n_h} x_{hi}}{\sum_{hi}^{r_h} y_{hi}^O + \sum_{hi}^{r_h} x_{hi}} \quad (3.7)$$

가 된다. 이석진과 신기일 (2008)의 모의실험에서는  $x_{hi}$ 의 분산과  $\sum^{n_h} x_{hi}, \sum^{r_h} x_{hi}$ 가 매우 크다는 조건 하에서  $y_{hi}^O$ 와  $y_{hi}^D$ 가 독립임을 보였으며 이를 이용하여  $f_h^{NRA}$ 가  $n_h/r_h$ 에 가까워짐을 보였다. 그러나 층으로 모집단이 나누어질 경우 이러한 가정은 현실적으로 어렵다. 즉  $x_{hi}$ 의 분산이 매우 클 경우 실사 자료가 다른 층으로 바뀌게 되는 경우가 발생하기 때문이다. 또한 모집단이 층으로 나누어져 있어서 층 안에 속한 자료의 합이 매우 커지지 않을 수도 있기 때문에  $\sum^{n_h} x_{hi}, \sum^{r_h} x_{hi}$ 이 매우 크게 된다는 가정도 현실적으로 어렵다. 오히려  $\sigma_{y_h}^2 / \sigma_{x_h}^2$ 이 작게 되는 경우가 현실적일 것이다. 이 경우  $\rho_{h,xy}$ 는 작게 되지만 “0”이 되는 경우는 없을 것이다. 이제 식 (3.7)을 다시 적으면 다음과 같다.

$$\begin{aligned} f_h^{NRA} &= \frac{\sum_{hi}^{n_h} y_{hi}^O + \sum_{hi}^{n_h} x_{hi}}{\sum_{hi}^{r_h} y_{hi}^O + \sum_{hi}^{r_h} x_{hi}} = \frac{n_h \bar{y}_{n_h}^O + \bar{x}_{n_h}}{r_h \bar{y}_{r_h}^O + \bar{x}_{r_h}} \\ &= \frac{n_h (\mu_{y_h} + \mu_{x_h}) + \frac{\sigma_{y_h}}{\sqrt{n_h}} z_{y_h} + \frac{\sigma_{x_h}}{\sqrt{n_h}} z_{x_h}}{r_h (\mu_{y_h} + \mu_{x_h}) + \frac{\sigma_{y_h}}{\sqrt{r_h}} z_{y_h} + \frac{\sigma_{x_h}}{\sqrt{r_h}} z_{x_h}} \\ &= \frac{n_h \frac{\mu_{y_h} + \mu_{x_h}}{\sigma_{x_h}} + \frac{\sigma_{y_h}}{\sigma_{x_h} \sqrt{n_h}} z_{y_h} + \frac{1}{\sqrt{n_h}} z_{x_h}}{r_h \frac{\mu_{y_h} + \mu_{x_h}}{\sigma_{x_h}} + \frac{\sigma_{y_h}}{\sigma_{x_h} \sqrt{r_h}} z_{y_h} + \frac{1}{\sqrt{r_h}} z_{x_h}}. \end{aligned} \quad (3.8)$$

이제  $\sigma_{y_h} / \sigma_{x_h}$ 가 매우 작은 경우를 가정한다면 식 (3.8)은 근사적으로 다음과 같다.

$$f_h^{NRA} = \frac{n_h \frac{\mu_{y_h} + \mu_{x_h}}{\sigma_{x_h}} + \frac{1}{\sqrt{n_h}} z_{x_h}}{r_h \frac{\mu_{y_h} + \mu_{x_h}}{\sigma_{x_h}} + \frac{1}{\sqrt{r_h}} z_{x_h}}. \quad (3.9)$$

결론적으로  $n_h, r_h$ 가 상대적으로 충분히 큰 경우에  $f_h^{NRA} \approx n_h/r_h$ 가 성립된다. 즉 근사적으로 BLS 방법은 일반적인 방법과 같은 결과를 준다.

여기서 식 (3.4)를 일반적인 분리비추정량(Ratio estimator) 형태로 만들면 다음과 같다. 즉

$$t_h(B) = N_h \times \frac{\bar{y}_{h_r}^O}{\bar{y}_{h_r}^D} \times \bar{y}_{h_n}^D.$$

이 경우  $\bar{R}_h = \bar{y}_{h_r}^O / \bar{y}_{h_r}^D$  라 하면  $\bar{R}_h$ 는  $h$ 층의 분리비 추정량의 형태가 되어 전체적으로 총계추정을 위한 비추정의 형태가 된다. 그러나 이 결과는 비추정의 형태를 취하고는 있으나 비추정을 사용할 경우의 기본 가정과 다르기 때문에 비추정에서 얻어지는 결과를 직접적으로 적용할 수는 없다.

## 4. 모의실험

### 4.1. 자료 설명

효율성 분석을 위한 모의실험을 실시하였다. 모의실험에 사용된 자료는 2007년 1월과 5월의 매월노동

표 4.1. 층별 상관계수

층	모집단수	상관계수	분산
C1	268	0.1971	0.7
C2	1328	0.5820	1.9
C3	1596	0.6564	3.5
C4	2044	0.7435	31.6
C5	1596	0.9407	379.1
C6	983	0.9469	2895.8
C7	330	0.8996	3123.7
C8	405	0.9980	5493370.9
C9	1718	0.9982	1588170.9
C10	3314	0.9983	878199.6

통계 자료이다. 이 자료는 약 7,000개로 이루어졌으며 월별 상용근로자 수를 모집단으로 가정하고 분석하였다. 또한 1월 자료를  $y_i^D$ 로 5월 자료를  $y_i^O$ 로 가정하였다. 모집단은 각 사업체의 상용근로자 수를 기준으로 7개 층으로 구분 되어있다. 본 모의실험에서는 전체 모집단 총계를 추정하는 것이 목적이 아니라 다양한 상관관계를 갖는 층을 만들어 층별 특성을 보는 것을 목적으로 하였다. 이를 위하여 추가로 몇 개 층을 합하여 새로운 층을 만들었다. 층의 이름과 구분은 다음과 같다. C1: 1-4인, C2: 5-9인, C3: 1-9인, C4: 10-29인, C5: 30-99인, C6: 100-299인, C7: 300-499인, C8: 500인 이상, C9: 100인 이상, C10: 30인 이상. 다음으로 각 층에서 자료 수 100, 150, 200, 250, 300, 350, 400인 자료를 추출하였으며 이 중에서 일부를 결측 처리하였다. 먼저 자료수 100에서는 결측치 수를 10, 30, 50, 70로 정하여 결측 비율이 매우 큰 경우를 살펴보았다. 특히 50개와 70개를 결측한 이유는 C10이 식 (3.5)을 만족하는지 여부를 확인하기 위해서이다. 다음으로 일정 비율의 결측치를 만들었는데 이때 사용한 결측 비율은 현실적으로 나올 수 있는 3, 5, 10, 15, 20%이다. 각 층별 상관계수와 분산을 계산한 결과는 표 4.1과 같다.

상관계수를 살펴보면 종사자 수가 큰 사업체 규모일수록 상관관계도 커지는 것을 알 수 있으며 식 (3.6)에서도 알 수 있듯이  $\sigma_y^2$ 이 커지면서 일반적으로 상관계수도 증가하는 것을 확인 할 수 있다. C1과 C2에서는 분산의 크기에는 차이가 별로 없지만 상관계수에는 큰 변화가 있는 것을 확인 할 수 있다.

#### 4.2. 비교 통계량

일반 방법과 BLS 보정 방법의 효과를 비교하기 위해 사용된 비교 통계량은 Rao (2003)에서 제안된 통계량을 사용하였으며 정의는 다음과 같다.

$$\begin{aligned}
 \text{MSE} &= \frac{1}{3000} \sum_k^{3000} (t_{B,k} - \hat{t}_{B,k})^2 \\
 \text{ABSBIAS} &= \frac{1}{3000} \sum_k^{3000} |t_{B,k} - \hat{t}_{B,k}| \\
 \text{MSPE} &= \frac{1}{3000} \sum_k^{3000} \left( \frac{t_{B,k} - \hat{t}_{B,k}}{t_{B,k}} \right)^2 \\
 \text{RRMSE} &= \left\{ \frac{1}{3000} \sum_k^{3000} \left( \frac{t_{B,k} - \hat{t}_{B,k}}{t_{B,k}} \right)^2 \right\}^{\frac{1}{2}}
 \end{aligned}$$

표 4.2. 표본 크기 100인 경우(무응답 수 고정)

층	무응답수	BLS방법				일반방법				Ratio mse
		Mse	Abs bias	Mspe	Rrmse	Mse	Abs bias	Mspe	Rrmse	
C1	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.9933
	30	2.116	1.500	2.116	1.453	2.121	1.500	2.122	1.455	0.9953
	50	4.022	2.077	4.023	2.003	3.987	2.077	3.987	1.994	0.9840
	70	8.694	3.115	8.695	2.945	8.695	3.077	8.698	2.945	0.9930
C2	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.0251
	30	1.132	1.046	1.132	1.063	1.207	1.092	1.207	1.099	1.0934
	50	1.411	1.131	1.411	1.186	1.634	1.243	1.634	1.281	1.1871
	70	1.986	1.298	1.986	1.407	2.434	1.496	2.434	1.561	1.2561
C3	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.1009
	30	1.350	1.150	1.350	1.161	1.620	1.275	1.620	1.271	1.3212
	50	2.075	1.401	2.077	1.442	2.686	1.636	2.688	1.639	1.4248
	70	3.759	1.819	3.761	1.936	5.304	2.246	5.306	2.304	1.5533
C4	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.0886
	30	1.156	1.057	1.156	1.074	1.372	1.189	1.373	1.173	1.2918
	50	1.544	1.171	1.543	1.243	2.158	1.481	2.159	1.470	1.5218
	70	2.649	1.402	2.649	1.627	4.271	2.036	4.272	2.069	1.7550
C5	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.1755
	30	1.057	1.031	1.057	1.029	1.390	1.171	1.391	1.182	1.5461
	50	1.152	1.071	1.153	1.074	2.199	1.453	2.199	1.485	2.2432
	70	1.389	1.164	1.390	1.181	4.040	1.975	4.041	2.012	3.4186
C6	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.2613
	30	1.080	1.044	1.081	1.042	1.553	1.246	1.554	1.245	1.8135
	50	1.216	1.108	1.218	1.103	2.563	1.597	2.565	1.598	2.6575
	70	1.520	1.248	1.521	1.234	4.936	2.205	4.940	2.220	4.0961
C7	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.1677
	30	1.085	1.037	1.090	1.040	1.419	1.177	1.425	1.185	1.5280
	50	1.287	1.110	1.290	1.128	2.290	1.468	2.293	1.511	2.0775
	70	1.710	1.247	1.716	1.304	4.202	1.990	4.210	2.044	2.8704
C8	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.0921
	30	1.001	1.001	1.001	1.000	1.236	1.080	1.236	1.112	1.3482
	50	1.003	1.001	1.003	1.001	1.685	1.211	1.685	1.298	1.8350
	70	0.999	0.999	0.999	0.999	2.657	1.452	2.657	1.630	2.9035
C9	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.0913
	30	1.000	1.000	1.000	1.000	1.300	1.097	1.300	1.140	1.4199
	50	0.996	0.999	0.996	0.998	1.843	1.260	1.843	1.358	2.0185
	70	0.998	0.998	0.998	0.999	3.056	1.506	3.056	1.749	3.3420
C10	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.1151
	30	1.001	1.000	1.001	1.001	1.321	1.131	1.321	1.149	1.4721
	50	1.001	1.000	1.001	1.000	1.959	1.345	1.959	1.400	2.1833
	70	0.993	0.998	0.993	0.997	3.451	1.664	3.451	1.858	3.8741

$$\text{Ratio - MSE} = \frac{\text{MSE}_{\text{일반방법}}}{\text{MSE}_{\text{BLS방법}}}$$

여기서  $t_{B,k}$ 는 모집단 총계이고,  $\hat{t}_{B,k}$ 는  $k$ 번째 반복에서 구해진 총계 추정치이다. 그리고 비교통계량을 구하기 위해 3,000번을 반복하였다.

표 4.3. 표본 크기 100인 경우(무응답 비율로 고정)

층	무응답 비율	BLS방법				일반방법				Ratio mse
		Mse	Abs bias	Mspe	Rrmse	Mse	Abs bias	Mspe	Rrmse	
C1	5	1.067	1.045	1.068	1.033	1.076	1.045	1.077	1.037	1.0014
	10	1.350	1.182	1.351	1.164	1.356	1.182	1.357	1.164	0.9971
	15	1.663	1.318	1.663	1.290	1.670	1.318	1.670	1.291	0.9974
	20	1.931	1.455	1.931	1.390	1.937	1.455	1.937	1.392	0.9962
C2	5	1.017	1.004	1.017	1.008	1.029	1.011	1.029	1.016	1.0171
	10	1.049	1.014	1.050	1.024	1.080	1.032	1.081	1.040	1.0351
	15	1.085	1.025	1.084	1.043	1.130	1.050	1.130	1.064	1.0474
	20	1.122	1.036	1.121	1.059	1.165	1.054	1.165	1.080	1.0441
C3	5	1.020	1.009	1.020	1.008	1.041	1.023	1.040	1.020	1.0495
	10	1.106	1.051	1.105	1.051	1.190	1.097	1.189	1.090	1.1073
	15	1.177	1.089	1.177	1.083	1.323	1.153	1.323	1.152	1.1568
	20	1.295	1.145	1.295	1.134	1.504	1.236	1.502	1.227	1.1952
C4	5	1.015	1.007	1.015	1.005	1.019	1.011	1.019	1.008	1.0270
	10	1.058	1.023	1.058	1.027	1.121	1.062	1.121	1.059	1.0837
	15	1.123	1.042	1.123	1.057	1.243	1.125	1.243	1.114	1.1328
	20	1.148	1.058	1.148	1.071	1.329	1.168	1.329	1.151	1.1847
C5	5	1.004	1.000	1.003	1.000	1.029	1.015	1.028	1.016	1.0690
	10	1.005	0.999	1.005	1.000	1.100	1.041	1.100	1.048	1.1411
	15	1.015	1.004	1.015	1.006	1.207	1.093	1.206	1.098	1.2398
	20	1.037	1.017	1.037	1.016	1.299	1.130	1.299	1.140	1.3067
C6	5	1.005	1.003	1.004	1.005	1.064	1.029	1.063	1.027	1.1220
	10	1.022	1.010	1.022	1.014	1.178	1.084	1.178	1.082	1.2225
	15	1.033	1.020	1.033	1.019	1.307	1.141	1.305	1.142	1.3411
	20	1.040	1.025	1.040	1.024	1.431	1.197	1.431	1.192	1.4588
C7	5	1.004	1.002	1.007	1.000	1.050	1.018	1.051	1.024	1.0982
	10	1.033	1.012	1.034	1.016	1.136	1.055	1.134	1.064	1.1552
	15	1.055	1.023	1.060	1.033	1.227	1.087	1.223	1.112	1.2213
	20	1.085	1.035	1.087	1.041	1.401	1.152	1.401	1.184	1.3557
C8	5	1.000	1.000	1.000	1.000	1.017	1.008	1.017	1.009	1.0412
	10	1.000	1.000	1.000	1.000	1.053	1.014	1.053	1.026	1.0769
	15	1.000	1.000	1.000	1.001	1.107	1.033	1.107	1.052	1.1324
	20	1.002	1.001	1.002	1.001	1.185	1.063	1.185	1.088	1.2102
C9	5	0.999	1.000	0.999	1.000	1.018	1.009	1.018	1.009	1.0447
	10	1.000	1.000	1.000	1.000	1.056	1.017	1.056	1.028	1.0826
	15	1.000	1.000	1.000	1.000	1.106	1.037	1.106	1.051	1.1336
	20	1.000	1.000	1.000	1.000	1.142	1.047	1.142	1.068	1.1708
C10	5	0.999	1.000	0.999	1.000	1.007	0.998	1.007	1.004	1.0515
	10	0.997	0.999	0.997	0.998	1.045	1.003	1.045	1.023	1.0942
	15	1.000	1.000	1.000	1.000	1.151	1.046	1.151	1.073	1.2010
	20	1.000	0.999	1.000	1.000	1.168	1.054	1.168	1.081	1.2188

#### 4.3. 모의실험 결과

표 4.2는 4.2절의 비교통계량을 이용하여 얻은 결과이다. 표 4.2의 결과를 살펴보면 다음과 같다. 먼저 표본 크기 100인 표 4.2의 층 C1-C10 결과를 살펴보자. 무응답의 수는 10, 30, 50 그리고 70개이며 무

표 4.4. 표본 크기 200인 경우의 결과

층	무응답비율	BLS방법				일반방법			
		Mse	Abs bias	Mspe	Rrmse	Mse	Abs bias	Mspe	Rrmse
C3	5	1.024	1.021	1.022	1.012	1.035	1.021	1.034	1.018
	10	1.130	1.079	1.130	1.066	1.200	1.104	1.197	1.094
	15	1.247	1.136	1.245	1.120	1.364	1.181	1.362	1.171
	20	1.328	1.171	1.329	1.157	1.547	1.250	1.545	1.247
C4	5	1.006	1.002	1.006	1.004	1.042	1.026	1.042	1.019
	10	1.062	1.023	1.062	1.030	1.135	1.077	1.135	1.064
	15	1.079	1.026	1.078	1.042	1.205	1.127	1.205	1.098
	20	1.149	1.052	1.148	1.072	1.313	1.185	1.313	1.143
C5	5	1.001	1.001	1.002	1.000	1.015	1.007	1.016	1.008
	10	1.013	1.009	1.013	1.004	1.091	1.058	1.092	1.047
	15	1.018	1.014	1.018	1.008	1.162	1.093	1.164	1.079
	20	1.027	1.017	1.028	1.012	1.236	1.129	1.237	1.115
C9	5	1.000	1.000	1.000	1.000	1.027	1.008	1.027	1.013
	10	1.002	1.000	1.002	1.001	1.088	1.027	1.088	1.043
	15	1.001	1.000	1.001	1.000	1.125	1.031	1.125	1.060
	20	1.000	1.000	1.000	0.999	1.194	1.051	1.194	1.093
C10	5	0.999	1.000	0.999	1.000	1.011	1.004	1.011	1.006
	10	0.999	0.999	0.999	1.000	1.070	1.029	1.070	1.035
	15	0.999	0.999	0.999	1.000	1.141	1.057	1.141	1.068
	20	0.999	0.999	0.999	1.000	1.230	1.092	1.230	1.109

응답 비율이 커질 때 BLS 방법과 일반적인 방법에서 얻어진 결과가 나와 있다.

먼저 층 C1 결과를 살펴보면 무응답의 수가 증가 할수록 MSE를 포함한 모든 비교 통계량이 급격히 증가하는 것을 확인 할 수 있다. 이러한 경향은 BLS 방법 뿐만 아니라 일반적인 방법에서도 나타나고 있다. 다음으로 BLS 방법과 일반적인 방법을 비교해 보면 어떤 방법을 사용하여도 큰 차이가 없음을 확인 할 수 있다. 이는  $y_{hi}^D$ 와  $y_{hi}^O$ 의 상관계수가 0.19로 작기 때문에 나타나는 결과이다.

다음으로 층 C2, C3, C4 결과를 살펴보자. 이 층들의 상관계수는 0.58에서 0.75사이인 경우이다. 이 경우의 비교 통계량 결과를 살펴보면 층 C1에 비해 급격히 증가하고 있지는 않다. 그러나 무응답이 70개인 경우, MSE 기준으로 약 2배에서 4배까지 증가하고 있다. 이와 같은 결과는 일반적인 방법에서도 나타나고 있다. MSE 기준으로 살펴보면 일반적인 방법에서 무응답이 70개인 경우 약 2.4배에서 5.3배로 더욱 나빠지는 것을 확인 할 수 있다.

또한 층 C5, C6, C7의 경우는 상관계수가 약 0.9에서 0.95 사이의 결과로 무응답의 수가 증가하여도 크게 비교통계량의 결과가 나빠지고 있지 않다. MSE 기준으로 보면 약 1.3배에서 1.7로 나빠지는 것을 확인 할 수 있다. 그러나 일반적인 방법의 경우 보조 정보를 사용하고 있지 않기 때문에 결과가 개선되고 있지 않다.

표 4.3에서 표 4.5는 무응답 비율을 3, 5, 10, 15, 20%로 고정하고 표본 크기를 100에서 400으로 증가시켰을 경우의 결과표이며 관계상 3% 결과는 생략하였다.

끝으로 층 C8, C9, C10 결과를 살펴보자. 결과를 보면 무응답 수에 상관없이 모든 비교 통계량의 값이 “1”에 가까운 것을 확인 할 수 있다. 상관계수를 보면 모두 0.998 이상으로 나와 있어 상관계수가 “1”에 가까운 경우 무응답 수에 상관없이 좋은 결과를 얻을 수 있음을 확인 할 수 있다. 물론 일반적인 방법인

표 4.5. 표본 크기 400인 경우의 결과

층	무응답비율	BLS방법				일반방법			
		Mse	Abs bias	Mspe	Rrmse	Mse	Abs bias	Mspe	Rrmse
C3	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10	1.038	1.019	1.042	1.017	1.049	1.029	1.047	1.025
	15	1.133	1.058	1.133	1.058	1.204	1.096	1.201	1.098
	20	1.235	1.107	1.238	1.108	1.394	1.173	1.396	1.180
C4	5	1.343	1.146	1.343	1.158	1.542	1.231	1.544	1.246
	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	15	1.017	1.005	1.015	1.006	1.034	1.009	1.032	1.018
	20	1.073	1.023	1.074	1.030	1.143	1.062	1.143	1.072
C5	5	1.127	1.029	1.125	1.061	1.258	1.104	1.257	1.126
	10	1.189	1.050	1.188	1.085	1.365	1.155	1.364	1.168
	15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	20	1.003	1.001	1.007	1.000	1.048	1.021	1.045	1.024
C9	5	1.025	1.010	1.028	1.017	1.199	1.085	1.199	1.096
	10	1.055	1.021	1.056	1.025	1.375	1.152	1.372	1.168
	15	1.083	1.037	1.083	1.042	1.543	1.212	1.538	1.240
	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
C10	5	1.001	1.000	1.001	1.001	1.072	1.040	1.072	1.036
	10	0.999	1.000	0.999	1.000	1.240	1.130	1.240	1.114
	15	1.003	1.002	1.003	1.003	1.390	1.205	1.390	1.180
	20	1.002	1.001	1.002	1.001	1.545	1.279	1.544	1.244

경우는 크게 개선되지 않음을 알 수 있다.

이제 Ratio-MSE 기준 결과를 살펴보자. 먼저 상관계수가 0.197인 C1을 살펴보면 오히려 일반적인 방법의 MSE가 작은 결과를 보여주고 있다. 물론 그 차이가 매우 미세하여 어떤 방법을 사용하여도 무방하다고 할 수 있다. 그러나 층 C2에서 C10을 보면 상황은 달라진다. 특히 상관계수가 높은 C8에서 C10을 살펴보면 약 3배에서 4배 정도로 BLS 방법이 효과적인 것을 확인할 수 있다. 따라서 상관계수가 매우 작은 경우에는 어떤 방법을 사용하여도 좋으나 상관계수가 존재하거나 높은 경우에는 BLS 방법을 사용하는 것이 타당할 것이다.

다음으로 표 4.3을 살펴보자. 일반적인 표본조사에서는 무응답 비율이 20%이내이므로 이 구간에서의 변화를 자세히 살펴볼 수 있도록 무응답 비율을 조정하였다. 전체적인 결과는 표 4.2의 결과와 유사하다. 즉 상관계수가 높을수록 일반적인 방법에 비해 BLS의 효과는 높아지고 있으며 특히 상관계수가 높은 경우 무응답 비율에 상관없이 매우 우수한 결과를 주고 있다.

표 4.4와 4.5는 표본 크기의 영향력을 살펴보기 위해 실시한 모의실험 결과이다. 실제로 표본 크기를 150, 200, 250, 300, 350 그리고 400개를 이용하여 실시하였으나 결과가 비슷하여 표본 크기가 200과 400인 경우의 결과만을 표로 작성하였다. 또한 대표적인 층인 C3, C4, C5, C9, C10 만을 살펴보았으며 C1의 경우는 모집단 크기가 268개로 작아 표본 크기에 따른 영향력을 살펴보는 모의실험에서 제외했다.

표 4.6은 무응답 비율별, 표본 크기별로 BLS 방법과 일반적인 방법의 MSE 비율인 Ratio-MSE 결과를 정리한 표이며 그림 4.1과 4.2는 층별, 무응답 비율별, 표본 크기별 Ratio-MSE 중에서 3%와 15%를 그림으로 표현한 것이다.

표와 그림을 살펴보면 무응답 비율이 증가할수록, 상관계수가 증가할수록 BLS 보정 방법이 효과적임을

표 4.6. Ratio-MSE

층	무응답비율	표본 크기						
		100	150	200	250	300	350	400
C3	3	1.0290	1.0206	1.0462	1.0186	1.0366	1.0159	1.0406
	5	1.0495	1.0334	1.0583	1.0447	1.0580	1.0240	1.0518
	10	1.1073	1.0617	1.1105	1.0825	1.1041	1.0306	1.1053
	15	1.1568	1.1196	1.1447	1.1373	1.1387	1.0584	1.1741
	20	1.1952	1.1392	1.2189	1.1895	1.2153	1.0653	1.1947
C4	3	1.0231	1.0270	1.0179	1.0410	1.0319	1.0528	1.0329
	5	1.0270	1.0369	1.0541	1.0828	1.0551	1.1063	1.0509
	10	1.0837	1.0620	1.0876	1.1313	1.1150	1.1792	1.1006
	15	1.1328	1.1072	1.1370	1.1911	1.1436	1.2620	1.1530
	20	1.1847	1.1349	1.1631	1.2608	1.2303	1.3551	1.1858
C5	3	1.0429	1.0461	1.0394	1.0444	1.0550	1.0168	1.0774
	5	1.0690	1.0811	1.0534	1.0721	1.0960	1.0247	1.1252
	10	1.1411	1.1476	1.1193	1.1098	1.1839	1.0515	1.2605
	15	1.2398	1.2498	1.1870	1.2122	1.2837	1.0875	1.4040
	20	1.3067	1.3222	1.2510	1.3225	1.4368	1.1001	1.5344
C9	3	1.0319	1.0252	1.0234	1.0181	1.0264	1.0890	1.0587
	5	1.0540	1.0447	1.0506	1.0534	1.0442	1.1361	1.1342
	10	1.1055	1.0826	1.1114	1.1145	1.0927	1.3063	1.3134
	15	1.1808	1.1336	1.1502	1.1851	1.1525	1.4810	1.4670
	20	1.2565	1.1708	1.2222	1.2491	1.2221	1.7060	1.6317
C10	3	1.0351	1.0432	1.0340	1.0522	1.0231	1.0428	1.0774
	5	1.0738	1.0515	1.0468	1.0745	1.0315	1.0657	1.1312
	10	1.1147	1.0942	1.1083	1.1815	1.1047	1.1488	1.2850
	15	1.2249	1.2010	1.1814	1.2962	1.1215	1.2768	1.4514
	20	1.3092	1.2188	1.2728	1.4071	1.1941	1.3434	1.6027

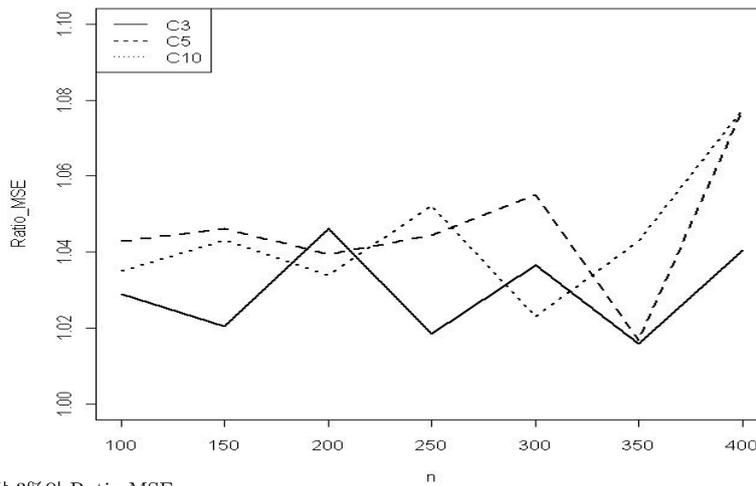


그림 4.1. 무응답 3%의 Ratio-MSE

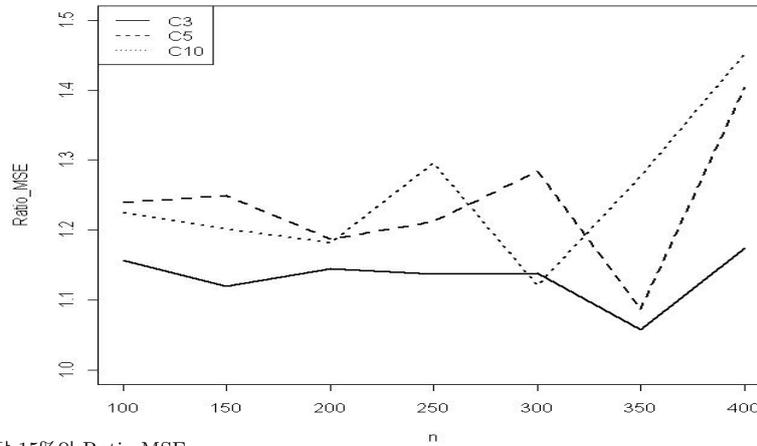


그림 4.2. 무응답 15%의 Ratio-MSE

보여주고 있다. 그러나 무응답 비율이 일정할 경우 표본 크기에 따른 BLS 보정 방법의 효과는 존재하지 않는 것을 확인 할 수 있다.

## 5. 결론

본 논문에서는 무응답이 발생했을 경우 가중치를 이용하여 보정하는 방법을 살펴보았다. 현재 일반적으로 사용하는 단위 무응답 가중치 보정 방법은 모집단 크기와 유효 표본 크기를 이용하여 가중치를 보정하고 있다 (김호진 등, 2008). 물론 이 방법은 항목무응답에서도 적용하고 있다. 그러나 항목 무응답 가중치 계산에 필요한 보조 정보가 있을 때 이 정보를 이용한 BLS 방법을 가중치 계산에 사용한다면 더 좋은 결과를 줄 수 있다. 이 결과는 보조자료와 실사 자료와의 상관관계가 존재한다는 조건이 필요하다. 본 논문에서는 상관관계가 높은 층이 있을 경우에는 BLS 방법이 매우 좋은 결과를 주는 것을 확인할 수 있었으며 많은 경우 표본 설계는 층화추출법을 사용하고 있기 때문에 본 논문에서 얻어진 결과는 매우 유용하리라 생각된다. 이러한 결론은 비추정의 분리비추정과 결합비 추정과의 관계를 이해하면 쉽게 얻어질 수 있는 결론이다.

## 참고문헌

- 김규성 (2000). 무응답대체방법과 대체효과, <조사연구>, **1**, 1-14.
- 김호진, 류정진, 장영석, 류기섭 (2008). <제1차 장애인고용패널조사>, 한국장애인고용촉진공단, 고용개발원.
- 신민웅, 이상은 (2001). <표본조사를 위한 표본설계>, 교우사.
- 이석진, 신기일 (2008). BLS 보정방법의 민감도에 관한 연구, <한국통계학회논문집>, **15**, 843-858.
- Burdete, T. (2003). *Survey of Occupational Injuries and Illnesses, Sample Design*, Bureau of Labor Statistics, manuscript.
- Little, R. J. A. and Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, New York.
- Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, John Wiley & Sons, New Jersey.
- Oh, H. L. and Scheuren, F. (1987). Modified raking estimation, *Survey Methodology*, **13**, 209-219.
- Rao, J. N. K. (2003). *Small Area Estimation*, John Wiley & Sons, New York.

# A Study on the Efficiency of the BLS Nonresponse Adjustment According to the Correlation and Sample Size

Seok Kim<sup>1</sup> · Key-Il Shin<sup>2</sup>

<sup>1</sup>Department of statistics, Hankuk University of Foreign Studies

<sup>2</sup>Department of statistics, Hankuk University of Foreign Studies

(Received September 2009; accepted October 2009)

---

## Abstract

Efficiency and sensitivity of BLS adjustment method have been studied and the method is known to provide more accurate estimate of total by using properly adjusted weights of samples. However, BLS methods provide different efficiencies according to the magnitudes of correlation coefficients and the sizes of samples in strata. In this paper we study the efficiency of the BLS adjustment according to the sample sizes and correlations in strata. For this study, 2007 monthly labor survey data is used.

Keywords: Sampling survey, weight, non-response adjustment, outlier adjustment.

---

---

This research was supported by the research fund of Hankuk University of Foreign Studies 2009.

<sup>2</sup>Corresponding author: Professor, Department of statistics, Hankuk University of Foreign Studies, Yongin-si, Kyonggy 449-791, Korea. E-mail: keyshin@hufs.ac.kr