



제1장

두 통계자료를 이용한 통합 추정 방법 연구 - 고용조사의 Calibration 추정 -

임경은·김서영·김재광

제1절 서론

1. 연구의 필요성

통계가 경제사회현상을 설명하는 중요한 정보원이 된 이래, 우리는 다양한 방법으로 다양한 분야의 통계를 작성하고 있다. 공식통계(official statistics)는 주로 조사와 행정자료에 의해 작성되고 있다. 때로는 같은 시점에서 동일한 목적을 갖는 여러 개의 통계가 작성되는 경우도 있다.

하나의 예로 같은 시점에서 조사되는 전수조사와 표본조사를 들 수 있다. 전수조사는 전국단위조사로 전국, 시도, 시군구 및 이하의 소지역까지 통계가 작성될 수 있다. 표본조사는 일반적으로 적은 표본수를 이용하여 전수조사보다는 넓은 지역이나 영역에서 통계를 작성한다. 다른 예로 같은 시점에서 두 개의 표본조사를 들 수 있다. 이들 두 조사는 전반적인 조사목적과 활용도는 다를 수 있지만, 특정 항목 또는 변수들은 같은 내용을 측정하는 경우가 있다. 이렇게 같은 시점에서 두 개 이상의 통계가 작성될 경우 동일한 모수에 대해 통계가 일치되지 않은 상태로 발표되는 경우가 발생할 수 있다.

동일한 모집단에 대한 서로 다른 조사에서 얻어지는 두 통계량을 일치시키는 것은 공식통계에서는 매우 중요한 작업이다. 미국에서는 1940년대에 이미 이런 작업들을 거침으로써 통계의 신뢰성을 유지해 왔다. 공식통계를 작성하는 통계기관에서는 이점에 특별히 신경을 쓸 필요성이 제기되고 있다. 두 조사의 통계량을 일치시키기 위해 가중치 작업을 해야 하는 중요한 명분을 3가지만 나열해 보자. 첫째, 하나의 모수에 하나의 통

계량이 부여됨으로써 통계의 일관성을 확보하고 통계작성기관의 신뢰성을 높인다. 즉, 서로 다른 통계는 사용자 입장에서는 상당히 혼란스러울 수 있고, 통계작성기관의 신뢰성을 현저히 떨어뜨릴 수 있다. 둘째, 이러한 통계 가중치 작업은 다른 측정치에서의 분산을 낮추는 효과를 발생시킨다. 가중치 작업을 잘 수행하면 상대표준오차, CV값이 낮아진다. 마지막으로, 두 통계가 전국 단위에서 일치되지 않은 채 공표작업을 할 경우, 무응답 등에 의해 나중에 그 불일치도가 더 커지게 되어 통계 전반적으로 문제를 일으킬 위험이 있다.

통계청은 2008년 10월 시점에서 우리나라 고용통계와 관련한 두 개의 통계를 발표하였다. 경제활동인구조사(이하 경활조사)와 2008년 10월 처음으로 실시된 지역별 고용조사(이하 고용조사)가 그것이다. 두 조사는 모두 우리나라 인구의 실업, 취업 등의 경제활동상황을 조사한다는 점에서는 동일한 목적을 갖는다. 반면에 본 연구내용과 관련하여 두 조사의 차이점은 공표단위에 있다. 경제활동인구조사는 전국과 시도 단위를 기본으로, 고용조사는 시군 단위를 기본으로 통계를 작성한다. 두 조사간 불일치는 여기서 발생한다. 고용조사가 시군 단위 통계작성이 목적이지만 시도와 전국 추정치도 자연스럽게 작성될 수 있다. 그러나 이 값들은 경활조사의 전국 또는 시도 단위 통계치와 달라, 동일 모집단에 대해 두 개의 통계가 존재하게 된 셈이다.

결국, 통계청은 2008년 10월 시점에서 우리나라 전국(시도 단위)취업자, 실업자수에 대해 서로 다른 두 개의 통계를 발표하였다. 이는 이용자 입장에서는 상당히 혼란스러운 일이며 나아가서는 통계청의 신뢰도를 떨어뜨릴 위험이 크다. 일반적으로 동일 모집단을 대상으로 하는 두 통계치는 일치시키는 것이 바람직하며 반드시 필요한 작업이다. 이에 본 연구는 두 통계를 일치시키는 통계적 방법을 찾는 데 그 목적을 둔다.

2. 연구 범위

본 연구는 크게 두 가지 목표를 갖는다. 하나는 고용조사와 경활조사의 통합추정을 위한 가장 좋은 추정법을 찾는 것이고, 다른 하나는 다양한 방법론을 소개하고 그 중요성과 필요성을 강조하는 것이다. 현재 고용조사와 경활조사의 전국과 시도 단위에서 통계 불일치 현상이 발생하고 있다. 따라서 전국 단위와 시도 단위에서 두 조사의 통계를 일치시키는 방법을 연구한다.

통계청에서 실시하는 경활조사와 고용조사는 거의 동일한 항목을 조사한다. 그렇지만 고용조사는 표본이 큰(약 17만 가구) 대신 비표본 오차가 크다. 반면에 경활조사는 표본은 고용조사에 비해 작지만(약 3만 2천 가구) 비표본 오차가 작다. 경활조사의 표본 오차는 전국과 시도 단위에서는 작지만, 시군 단위를 통계를 내기에는 무시할 수 없는



수준이다. 따라서 표본수가 많은 고용조사를 이용하여 시군 단위 통계를 내되, 비표본 오차 측면에서 보다 안정적이라 할 수 있는 경찰조사의 관측치를 반영하여 고용조사의 비표본 오차를 보정할 필요가 있다. 이와 같은 보정작업을 통해 두 조사의 전국과 시도 단위의 통계를 일치시키는 것이 중요하다.

따라서 본 연구는 다음의 두 가지 문제에 초점을 두고 연구를 진행한다.

(1) 통합추정 방법론 검토 연구 : 두 조사를 통합 추정하는 방법론을 소개하고, 각 방법들의 특성 및 장단점을 설명한다. 각 방법들의 이론적 특성에 비추어 우리가 사용하고 자 하는 자료에 가장 부합되는 방법을 일차적으로 검토한다. 이때 이론적 강점과 실용성 측면을 동시에 고려한다.

(2) 반복비례가중방법 적용 연구 : 고용조사와 경찰조사의 편차를 보정하기 위해 반복비례가중방법을 연구한다. 1차원 기준 주변분포와 2차원 기준의 결합분포를 이용한 반복비례가중법을 각각 적용한다. 편차조정에 적합한 보정변수를 찾고 최종적으로 5개의 보정변수를 이용한다.

따라서 본 연구는 위 두 가지에 초점을 두어 서술한다. 첫 번째 목표는 다양한 방법을 소개하는 것으로 충분히 달성될 수 있고, 선행연구 논문들을 통해 비교·정리한다. 연구자들은 본 연구의 많은 부분을 반복비례가중법 연구에 할애하였다. 이는 실용성 측면에서 우리가 보유하고 있는 자료와 가장 부합할 것이라는 연구자들의 판단에 따른 것이다. 또, 이 방법이 일반적으로 적용하기 쉽고 기본적인 수 있다는 측면도 고려하였다. 다른 방법들에 대한 분석연구는 이후 과제로 넘긴다.

3. 연구의 제한과 기대효과

본 연구는 많은 연구에서 그렇듯이 몇 가지 문제점을 가지고 있다. 이와 같은 문제점을 밝히는 이유는 본 연구가 최적의 통합 추정방법을 찾는 연구인만큼 연구진행 과정에서 나타나는 방법들의 특성을 충분히 이해하고, 자료분석 결과를 보다 명확하게 해석하기 위함이다. 우선 다양한 보정변수의 제한을 들 수 있다. 가중치 조정의 경우 보조변수의 선택은 매우 중요하다. 좋은 보조정보를 사용할 수 없을 경우 조정의 효과는 다소 미약할 수 있다. 그리고 두 통계자료를 이용한 통합추정에 대한 인식이 약한 것도 다양한 이론적 접근을 어렵게 하고 있다.

연구결과는 동일 모집단에 대한 두 통계의 불일치 문제를 해결함으로써 이용자의 혼란을 없애고, 통계청의 신뢰도를 높이는 데 기여할 수 있을 것이다. 특히, 공식통계에서 보다 넓고 깊게 고려되어야 할 가중치 작업에 관한 연구라는 점에서 본 연구의 가치가

있다. 또, 고용조사의 가중치 조정은 전국 또는 시도 단위에서 일치된 통계를 제공할 수 있는 좋은 방법이 될 것이다.

제2절 Calibration을 위한 가중치 작업

1. 개념

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

가중치 작업은 표본 추출을 통해서 얻어지는 표본에 대하여 적절한 가중치를 부여해 줌으로써 추정량의 효율성과 신뢰성을 제고하고자 하는 방법으로 표본 조사에서의 필수적인 요소이다. 가중치 작업은 표본 추출 확률을 이용한 설계 가중치(design weight) 외에도 외부 보조 변수와의 일치성을 위해 조정해 주는 calibration, 단위 무응답 자료의 처리를 위한 무응답 가중치 조정, 그리고 지나치게 크거나 작은 가중치 이상값(outlier)을 판별하여 이를 처리하는 가중치 이상값 처리와 같은 요소들로 구성되어 있다. 이를 정리하면 다음과 같다.

설계 가중치 작성 → 무응답 가중치 조정 → Calibration → 가중치 이상값 처리

설계 가중치는 일차 표본 포함 확률(first-order inclusion probability)의 역수를 사용하여 계산되는데 이 설계 가중치를 사용한 추정량은 Horvitz-Thompson 추정량 (이하 HT 추정량)이 된다. 무응답 가중치 조정은 응답 자료만을 사용할 때 발생할 수 있는 무응답 오차(nonresponse error)를 줄여주기 위해서 가중치를 응답 확률을 추정하여 그 역수로 사용한다. 무응답 가중치 조정에 관하여는 Sarndal and Lundstrom(2004)을 참고하기 바란다.



calibration은 보조변수에 대한 모집단 값을 알 때 그 보조변수의 모수(예를 들면 모평균)에 대해 일치하는 통계량을 구현하도록 최종 가중치를 결정하는 방법론이다. 즉, 보조변수를 x 라고 하고 표본원소 집합을 S , 모집단 원소 집합을 U 라고 하면

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

을 만족하도록 최종 가중치(w_i)를 결정해 준다. 이러한 가중치를 사용하는 통계량 $\hat{Y}_w = \sum_{i \in S} w_i y_i$ 은 y_i 가 $y_i = \mathbf{b}'\mathbf{x}_i$ 의 선형관계를 만족하는 경우에는 Y 의 총계 추정치가 참값과 같아지는 (즉, $\hat{Y}_w = \sum_{i \in U} y_i$ 가 성립하는) 결과를 가지게 되므로 이러한 선형관계 하에서는 오차가 없는 추정이 가능하게 되므로 바람직하다고 할 수 있을 것이다. 이러한 선형관계가 직접 성립하지 않고 $y_i = \mathbf{b}'\mathbf{x}_i + e_i$ 를 만족하는 경우에도 (여기서 e_i 는 오차) 이러한 모형이 y 에 대한 설명력이 높을 경우에는 (1)을 만족하는 추정이 바람직할 것이다. 즉, calibration 추정을 고려해야 하는 이유는 크게 두 가지이다.

- ① 특정 항목에 대해 미리 알고 있는 참값을 추정에 반영함으로써 추정치간의 일치를 이루고자 하는 통계 조정의 이유.
- ② 설명력이 높은 보조변수를 이용함으로써 추정치의 효율을 향상시키고자 하는 이유

이러한 calibration 추정을 사용하고자 할 때에는 다음과 같은 사항들을 고려해야 한다.

- ① 변수 선택: 어떠한 보조변수(x)를 사용할 것인가?
- ② 방법론 선택: 어떠한 calibration 방법론을 사용할 것인가?
- ③ 평가: 이렇게 얻어진 calibration 추정량의 효율을 어떻게 평가할 것인가?

변수 선택과 관련하여서는 두 가지 측면이 고려된다. 하나는 정책적 측면이고 다른 하나는 통계학적 측면이다. 정책적 측면이라 함은 통계학 외적인 이유로서 이 자료를 외부에 발표하였을 때 기존의 다른 통계들과 일치하는가를 생각하여 그렇지 않은 항목이 있다면 이를 가급적 반영하여 조정하고자 하는 측면이다. 이러한 정책적 측면은 통계청에서 발표하는 다른 통계들과 모순되는 결과가 나오지 않도록 주의해야 하는 이유 때문인데 만약 서로 다른 통계에서 나오는 다른 결과들을 아무런 조정 없이 발표한다면 이

는 통계청의 신뢰성을 저하시키는 요인이 되기 때문이다.

통계학적 측면으로는 어떠한 보조변수를 선택했을 때 추정량이 보다 정확해질 수 있겠는가 하는 문제이다. 이는 관심변수 y 를 잘 설명해주는 보조변수 x 를 결정하는 문제로 생각할 수 있는데 이는 회귀분석에서의 변수 선택의 문제와 동일해진다. 실업률의 추정에서는 관심변수 y 가 실업 유무를 나타내는 지시변수이므로 로지스틱 회귀분석에서의 보조변수 선택문제로 접근할 수 있을 것이다.

calibration 방법론으로는 raking ratio 방법이나 회귀 추정(regression weighting) 방법이 흔히 사용된다. calibration에 대한 보다 자세한 내용은 Kim and Park(2009)을 참고하기 바란다. 우선 raking ratio 방법론은 보조변수가 범주형인 경우에 한해서 사용되는 방법이고 회귀 추정 방법은 보다 일반적으로 사용할 수 있으나 가중치가 음수값을 취할 수 있다는 단점이 있다. 회귀 추정 방법론에 대하여서는 2.2절에서 보다 자세히 다루고자 한다. raking ratio 방법은 2.3절에서 실제 자료 적용과 함께 자세하게 다룰 것이다. 새롭게 제안되는 exponential tilting 방법론은 calibration 방법 중의 하나로 2.3절에서 자세하게 다루게 된다.

또한 본 연구에서는 보조변수 x 의 모집단 총계 $X = \sum_{i \in U} x_i$ 를 알지 못하지만 다른 조사 등을 통하여 그 추정량 \hat{X} 을 알 경우 어떻게 이를 추정에 반영할 수 있을 것인가를 다루고자 한다. 이를 위해 2.4절에서는 모집단 총계의 추정량 \hat{X} 을 사용한 calibration 방법론에 대해 기존의 연구들을 소개한다. 2.6절에서는 고용조사에서 사용할 수 있는 calibration 방법론에 대해 다루기로 한다.

2. 회귀 추정량

관심 변수를 y_i 라 하고 보조 변수를 x_i 라 하며 관심 변수는 표본에서만 관측 가능하지만 보조 변수는 모집단 전체에 대하여 관측 가능한 경우를 생각하자. 이 표본에서의 설계 가중치를 d_i 라고 하면 설계 가중치를 이용한 x_i 총계 추정량 $\hat{X}_d = \sum_{i \in S} d_i x_i$ 은 이

보조 변수의 참값 $X = \sum_{i=1}^N x_i$ 과 반드시 일치하는 것은 아니다. 이러한 x 의 모집단 총계 값은 표본 외에서 얻어지는 보조 정보로써 y 의 모집단 총계 추정에 이용될 수 있다. 회귀 추정량은 calibration 추정량의 대표적인 형태로 다음과 같이 표현될 수 있다.

$$\hat{Y}_{reg} = \hat{Y}_d + (X - \hat{X}_d)' \hat{b} \quad (2)$$



이 때

$$\hat{\mathbf{b}} = \left\{ \sum_{i \in S} d_i (\mathbf{x}_i - \bar{\mathbf{x}}_d) (\mathbf{x}_i - \bar{\mathbf{x}}_d)' \right\}^{-1} \sum_{i \in S} d_i (\mathbf{x}_i - \bar{\mathbf{x}}_d) y_i$$

으로 계산된다. 이 회귀 추정량은 최종 가중치를

$$w_i = d_i \{1 + \hat{\mathbf{c}}' (\mathbf{x}_i - \bar{\mathbf{x}}_d)\} \quad (3)$$

으로 사용하는 선형 통계량으로 구현될 수 있는데 여기서

$$\hat{\mathbf{c}} = \left\{ \sum_{i \in S} d_i (\mathbf{x}_i - \bar{\mathbf{x}}_d) (\mathbf{x}_i - \bar{\mathbf{x}}_d)' \right\}^{-1} (\mathbf{X} - \widehat{\mathbf{X}}_d)$$

으로 표현된다. 만약 $\widehat{\mathbf{X}}_d = \mathbf{X}$ 가 성립하면 $\hat{\mathbf{c}} = \mathbf{0}$ 이고 따라서 최종 가중치는 설계 가중치와 같아지나 그렇지 않은 경우에는 $\hat{\mathbf{c}} \neq \mathbf{0}$ 이고 $\widehat{\mathbf{X}}_d$ 이 모집단 참값과 떨어져 있는 정도에 따라 달라진다. 또한 식 (3)의 최종 가중치는

$$\sum_{i \in S} w_i(1, \mathbf{x}_i) = \left(\sum_{i \in S} d_i, \mathbf{X} \right) \quad (4)$$

을 만족시킴을 확인할 수 있다. 만약 모집단 크기인 N 을 알고 있다면 식 (4) 대신에

$$\sum_{i \in S} w_i(1, \mathbf{x}_i) = (N, \mathbf{X})$$

이 만족되기를 원할 수 있고, 이런 경우 회귀 추정량은 식 (2) 대신에

$$\widehat{Y}_{reg} = \widehat{Y}_d^* + (\mathbf{X} - \widehat{\mathbf{X}}_d^*)' \hat{\mathbf{b}} \quad (5)$$

을 사용하면 된다. 여기서

$$\left(\widehat{Y}_d^*, \widehat{\mathbf{X}}_d^* \right) = \left(\frac{N}{\sum_{i \in S} d_i} \right) \sum_{i \in S} d_i (y_i, \mathbf{x}_i)$$

이다.

식 (2)나 식 (5)로 정의되는 회귀 추정량은 여러 가지 장점이 있다. 먼저 회귀 추정량은 calibration 조건인 식 (1)을 만족한다. 회귀 추정량은 bias가 있는 추정량이지만 이 bias는 표본수가 어느 정도 커지면 무시할 수 있을 정도로 작아진다. 반면 회귀 추정량의 분산은 대략적으로

$$V(\hat{Y}_{reg}) = V(\hat{Y}_d) * (1 - R^2)$$

으로 표현되는데 여기서 R^2 는 y 를 종속변수로 하고 \mathbf{x} 를 독립변수로 하는 회귀분석에서의 결정계수이다. 극단적인 경우 모든 원소에 대해 $y_i = \mathbf{b}'\mathbf{x}_i$ 의 선형관계를 만족하는 경우에는 Y 의 총계 추정치가 참값과 같아지고 위의 분산은 0이 될 것이다. 즉, 회귀 추정량은 약간의 편향이 발생하지만 y 에 대한 보조 변수 \mathbf{x} 의 설명력이 좋으면 그 분산이 작아지게 되므로 추정량의 효율이 높아지게 된다. 만약에 R^2 값이 0.5가 되는 \mathbf{x} 를 찾을 수 있다면 그것을 이용한 회귀 추정량은 그렇지 않은 경우보다 분산이 반으로 줄어든다. 이는 \hat{Y}_d 를 사용하는 경우 표본수를 2배로 늘인 것과 동일한 효율을 가져오게 되는 것이다.

회귀 추정량의 단점은 최종 가중치가 지나치게 커지거나 작아질 수 있다는 것이다. 식 (3)으로 표현된 최종 가중치는 \mathbf{x}_i 의 선형함수로 표현된다. 이는 \mathbf{x}_i 중 어느 하나라도 지나치게 크거나 작은 경우 최종 가중치가 지나치게 커지거나 지나치게 작아지게 될 수 있다는 것을 의미한다. 극단적인 경우로는 가중치가 음수로 될 수 있다. 이런 경우 가중치가 음수가 되지 않도록 회귀 추정량을 약간 수정해야 할 필요가 있다.

3. Exponential Tilting 추정량

2.2 절에서 살펴본 회귀 추정량의 최종 가중치는 식 (4)의 조건을 만족하는 가중치 중에서 다음의 목적함수를 최소화하는 것으로 이해할 수 있다.

$$Q(w) = \sum_{i \in S} d_i \left(\frac{w_i}{d_i} - 1 \right)^2 \quad (6)$$

만약 위의 식 (6)의 목적함수 대신에

$$Q(w) = \sum_{i \in S} w_i \log \left(\frac{w_i}{d_i} \right)$$



을 최소화하는 calibration 추정량을 찾으면

$$w_i = d_i \exp(\lambda_0 + \lambda_1 \mathbf{x}_i) \quad (7)$$

로 계산될 수 있는데 여기서 (λ_0, λ_1) 은 식 (4)의 조건을 만족하도록 계산된다. 즉,

$$\frac{\sum_{i \in S} d_i \exp(\lambda_1 \mathbf{x}_i) \mathbf{x}_i}{\sum_{i \in S} d_i \exp(\lambda_1 \mathbf{x}_i)} = \frac{1}{N} \sum_{i \in U} \mathbf{x}_i \quad (8)$$

을 만족하도록 λ_1 을 찾는다. 이러한 비선형 방정식의 해는 항상 존재하고 그 해는 Newton-Raphson 방법에 의해 다음과 같은 반복계산법으로 계산되어질 수 있다.

$$\lambda_1^{(t+1)} = \lambda_1^{(t)} + \{S_\lambda^{(t)}\}^{-1} (\mathbf{X} - \hat{\mathbf{X}}_\lambda^{(t)})$$

여기서 $\hat{\mathbf{X}}_\lambda^{(t)}$ 은 가중치 $w_i^{(t)} = d_i \exp(\lambda_0^{(t)} + \lambda_1^{(t)} \mathbf{x}_i)$ 을 이용하여 얻어지는 x 의 총계 추정량이고 $S_\lambda^{(t)} = \sum_{i \in S} w_i^{(t)} (\mathbf{x}_i - \bar{\mathbf{x}}_{(t)}) (\mathbf{x}_i - \bar{\mathbf{x}}_{(t)})'$ 이며 $\bar{\mathbf{x}}_{(t)}$ 은 가중치 $w_i^{(t)}$ 를 이용하여 얻어지는 x 의 평균추정량이다. 초기치는 $\lambda_1^{(0)} = \mathbf{0}$ 으로 하여 계산하고 $\lambda_0^{(t)}$ 값은 최종가중치의 합이 N 과 같아지도록 결정한다. 만약 $\hat{\mathbf{X}}_d = \mathbf{X}$ 가 성립하면 $\hat{\lambda}_1 = \mathbf{0}$ 이고, 따라서 최종가중치는 실제 가중치와 같아지나 그렇지 않은 경우에는 $\hat{\lambda}_1 \neq \mathbf{0}$ 이고 $\hat{\mathbf{X}}_d$ 이 모집단 참값과 떨어져 있는 정도에 따라 다르다.

식 (7)로 정의되는 exponential tilting (ET) calibration 추정량은 2.2절에서 다룬 회귀 추정량과 거의 비슷한 통계적 속성을 가진다. 즉, ET 추정량은 bias가 있는 추정량이지만 이 bias는 표본수가 어느 정도 커지면 무시할 수 있을 정도로 작아진다. ET 추정량의 분산은 대략적으로

$$V(\hat{Y}_{ET}) = V(\hat{Y}_d) * (1 - R^2)$$

으로 표현되는데 여기서 R^2 는 y 를 종속변수로 하고 x 를 독립변수로 하는 회귀 분석에서의 결정계수이다.

식 (7)로 표현되는 ET calibration 가중치는 항상 0보다 큰 값을 갖고 \mathbf{x}_i 값이 극단적

인 값을 취할수록 가중치의 값이 커지게 된다. 가중치가 지나치게 커지는 것을 막기 위해서는 \mathbf{x}_i 대신에 극단값을 갖지 않도록 \mathbf{x}_i 값을 상한과 하한을 정한 후에 그 범위 내에서 값을 취하도록 변환한 값 $\mathbf{z}_i = \mathbf{T}(\mathbf{x}_i)$ 을 구한 후 식 (7) 대신에

$$w_i = d_i \exp(\lambda_0 + \lambda_1 \mathbf{z}_i) \quad (9)$$

으로 표현한 후

$$\frac{\sum_{i \in S} d_i \exp(\lambda_1 \mathbf{z}_i) \mathbf{x}_i}{\sum_{i \in S} d_i \exp(\lambda_1 \mathbf{z}_i)} = \frac{1}{N} \sum_{i \in U} \mathbf{x}_i$$

을 풀면 된다. 식 (9)로 표현되는 가중치는 \mathbf{z}_i 가 bounded 되었기 때문에 가중치도 bounded 되어서 지나치게 큰 값을 갖지 않게 된다.

결론적으로 ET 추정량의 성질을 정리하면 다음과 같다.

- ① ET 추정량은 회귀 추정량처럼 calibration 성질을 만족하는 새로 제안되는 추정량이다.
- ② ET 추정량은 회귀 추정량과 근사적으로 동일한 성질을 가지므로 회귀 추정량처럼 약간의 편향이 발생한다. 하지만 y 에 대한 보조 변수 \mathbf{x} 의 설명력이 좋으면 그 분산이 작아지게 되므로 추정량의 효율이 높아지게 된다.
- ③ ET 추정량의 가중치는 회귀 추정량의 가중치와는 달리 극단적인 값을 취하지 않으므로 실무에 유용하게 사용될 수 있다.

4. Composite Calibration 추정량

다음으로는 보조변수 x 의 모집단 평균을 모르지만 다른 조사 등을 통하여 그 추정량을 알 수 있을 경우 이를 추정에 어떻게 반영할 것인가 하는 방법에 대하여 다루고자 한다. Zieschang(1990)은 미국 노동 통계국(Bureau of Labor Statistics)에서 실시하는 Consumer Expenditure Survey(CES)의 가중치 작업에서 이러한 문제를 다루었다. CES에서는 똑같은 모집단에 대하여 두 개의 표본 집단을 독립적으로 추출하고 각각의 표본에 다른 조사를 실시하였는데 하나는 가계부를 기입하는 방식으로 조사하였고 다른 하나는 면접 조사를 실시하였다. 이처럼 같은 항목이 두 조사에 걸쳐서 나타나는 경우 각각 다른 표본에서 다른 결과가 나타날 수 있을 것이다. 이러한 경우 어떻게 두 개의 다른 통



계량을 잘 결합하여 일치된 결과를 얻어낼 수 있을 것인가는 현실적으로 매우 중요한 문제일 것이다.

이를 설명하기 위하여 동일한 모집단에서 단순 임의 추출로 두 개의 독립적인 표본 S_1 과 S_2 를 추출하였다고 하자. 또한 S_1 에서는 항목 x 와 z 를 관측하고 S_2 에서는 항목 x 와 y 를 관측한다고 하자. 이러한 경우 S_1 에서 얻어진 \bar{X}_N 의 불편 추정량을 \bar{x}_1 이라하고 S_2 에서 얻어진 \bar{X}_N 의 불편 추정량을 \bar{x}_2 이라 하고 우리의 관심 모수는 y 의 모집단 평균이라고 하자. 이러한 경우 (2)의 회귀 추정량을 사용하고자 한다면 \bar{X}_N 대신에 \bar{x}_1 을 사용하면

$$\hat{\theta}_{sr} = \bar{y}_2 + (\bar{x}_1 - \bar{x}_2)b \quad (10)$$

의 형태로 표현될 수 있을 것이다. 이 단순 회귀 추정량은 x 에 대하여 같은 추정량 \bar{x}_1 을 구현해 낸다는 일종의 calibration 성질을 지니는 장점이 있지만 전통적인 회귀 추정량과는 달리 식 (10)의 추정량의 분산이 y 의 단순 추정량 \bar{y}_2 의 분산보다 더 작게 된다는 보장이 없게 된다.

이러한 문제점을 해결하기 위하여 Zieschang(1990)과 Renssen and Nieuwenbroek(1997)은 \bar{x}_1 을 control로 사용하는 대신

$$\bar{x}_\alpha = \alpha \bar{x}_1 + (1 - \alpha) \bar{x}_2 \quad (11)$$

을 control로 사용하는 회귀 추정량을 제안하였다. 즉, 식 (11)을 사용한 회귀 추정량은

$$\hat{\theta}_{reg} = \bar{y}_2 + (\bar{x}_\alpha - \bar{x}_2)b \quad (12)$$

의 형태로 표현되며 $\alpha = 0$ 인 경우에는 식 (12)의 회귀 추정량이 단순 추정량 \bar{y}_2 와 동일하게 되고 $\alpha = 1$ 인 경우에는 식 (10)의 단순 회귀 추정량이 된다. 최적 계수 α 는 식 (12)의 분산을 최소화 하도록 결정하면 되는데 이 최적값은 두 표본이 독립인 경우

$$\alpha^* = \frac{1/V(\bar{x}_1)}{\{1/V(\bar{x}_1) + 1/V(\bar{x}_2)\}} \quad (13)$$

으로 표현된다. 즉, 식 (13)의 최적 계수는 식 (12)의 분산을 최소화하는 값이다. 이는 식 (11)에서 x 의 모평균의 최적 추정치를 calibration에 구현한 것과 동일해진다.

식 (12)의 추정량은 또한 다음과 같이 표현된다.

$$\hat{\theta}_{reg} = \alpha \hat{\theta}_{sr} + (1 - \alpha) \bar{y}_2 \quad (14)$$

즉, 식 (14)는 식 (10)의 단순 회귀 추정량과 단순 추정량 \bar{y}_2 와의 가중평균인 일종의 복합 추정량(composite estimator)의 형태가 된다. 이 복합 추정량은 x 의 추정에 대해 calibration을 유지하므로 최적 계수 α 를 사용한 복합 추정량은 calibration을 유지하면서 분산을 최소화 하는 추정량이 될 것이다. 여기서 α 는 $\hat{\theta}_{sr}$ 을 \bar{y}_2 방향으로 값을 보정해주는 Shrinkage 계수로도 불릴 수 있다.

두 조사를 이용한 calibration 최적 추정 방법은 다음과 같은 방법을 사용하면 된다.

(Step 1) 두 조사를 통합하여 X 의 총계 (또는 평균)에 대한 최적 추정량을 식 (11)과

식 (13)을 이용하여 구현한다. 이를 X_{opt} 라고 하자.

(Step 2) 식 (4)의 calibration 식을 다음과 같이 바꾼다.

$$\sum_{i \in S} w_i(1, \mathbf{x}_i) = \left(\sum_{i \in S} d_i, \mathbf{X}_{opt} \right) \quad (15)$$

(Step 3) 식 (15)의 calibration 식을 만족하는 calibration 추정량을 구현한다. 여기서 사용되는 calibration 추정량은 2.2절에서 다룬 회귀 추정량을 사용하거나 2.3절에서 다룬 ET 추정량을 사용한다.

이렇게 해서 얻어지는 calibration 추정량의 분산은

$$V(\widehat{Y}_{reg}) = V(\widehat{Y}_d) * (1 - R^2) + B' V(X_{opt}) B$$

으로 표현될 수 있는데 첫 번째 항은 X 총계를 정확하게 알 때 얻어지는 calibration 추정량의 분산이고 두 번째 항은 X 총계가 추정되기 때문에 발생하는 추가적인 분산의 증가값을 나타낸다.

5. Projection 방법

또 하나의 방법으로 projection 방법이 있다(Kim and Rao, 2009). 이 방법은 imputation 방법과 비슷한 개념이다. 즉 두 조사가 있다고 가정하자. 조사1에서는 보조변수 x 만 관



측하고, 조사2에서는 보조변수 x 와 관심변수 y 를 모두 관측한다고 하자. 이때 조사1의 표본이 조사2의 표본보다 더 크다. 추정방법은 먼저 조사2에서 보조변수 x 를 이용하여 합성한 새로운 y 값을 구한다. 새롭게 구한 y 값과 조사1의 표본가중치를 곱해서 y 의 예측값을 구한다. 이 예측치 y 값을 조사1의 y 값으로 대체한다.

즉, 고용조사 자료에서 x 를 관측하고 y 가 관측되지 않을 경우, 적절한 모형을 이용하여 y 를 예측하고, 이 예측치를 고용조사의 y 값으로 대체하는 방법이다. 이러한 projection 추정량의 효율은 y 를 예측하는 모형이 얼마나 좋은가에 달려있다.

고용조사의 경우, 먼저 모형선택을 통해 설명력이 높은 보조변수를 찾는다. 보조 변수가 지역*성별*연령이 선택되었다고 하자. 이 보조변수로 cell을 먼저 나눈 후 경찰조사에는 y (예: 경찰조사의 실업자)가 취하는 3가지 값들의 분포를 cell별로 구하고 (이를 f_1, f_2, f_3 라고 하자) 고용조사에서는 x (예: 고용조사의 실업자)가 취하는 세 가지 값들의 분포를 구한 후 (이를 g_1, g_2, g_3 라고 하자) 그 분포 값들의 비(ratio)를 $r_i = f_i/g_i$ 로 구한다. 그 후 고용조사의 각 y_2 값에 비를 곱하여 y_1 의 예측치를 구현할 수 있고 이러한 예측치를 사용하여 추정량을 구현할 수 있다.

Kim and Rao(2009)는 projection 방법론을 사용하여 두 서베이를 통합하는 방법을 제안하였다. 이 방법론의 아이디어는 결국 imputation 방법을 사용하여 관측되지 않은 변수를 만들어 내는 것이다. 즉, 기존의 imputation 방법은 자료의 일부가 missing인 경우에 사용되는 방법인데 이 projection 방법론은 자료 전체에서 해당 항목의 값이 관측되지 않는 경우에 사용되는 방법으로 이해될 수 있다. Kim and Rao는 두 서베이 자료의 크기가 차이가 많이 나는 경우, 큰 서베이 자료에 projection 방법론을 이용하는 것이 작은 서베이 자료에서 가중치 조정을 이용하는 추정보다 소지역 추정에서 더 좋은 결과를 가져다 준다고 주장하였다.

구현방법은 우선, projection을 통해서 최적의 모형을 구한 후, 가중치 작업을 통해 가중치 추정량을 구현하면 될 것이다. 이는 Wu and Sitter(2001)의 model calibration과 동일한 개념이다.

지금까지 설명한 가중치 조정 방법을 요약해 보면 <표 1-1>과 같다.

참고로 <표 1-1>에서 설명한 raking ratio를 포함하여 반복비례가중법은 본 연구에서 우선적으로 적용해 본 방법인 만큼 3절에서 자세하게 설명하기로 하겠다(Deville and Sarndal, 1992; Deville et al., 1993). Raking ratio을 제외한 방법들, 특히 회귀 추정량과 관련된 방법들에서는 보조정보 또는 적합한 모형을 찾는 것이 중요하다 하겠다. 가중치 작업에는 기본 가중치를 calibration이라는 과정을 통하여 보정하는 과정이 포함된다. calibration은 벤치마킹하게 될 변수들의 보정된 가중치의 합이 이미 알고 있는 참값과 같

아지도록 하는 것을 지칭하는데 이때 사용되는 변수들을 벤치마킹 변수(또는 설명변수)라고 한다. 이러한 벤치마킹 변수는 두 가지 조건을 만족해야 하는데 첫째 외부 자료 등을 통해 보다 정확한 값을 알고 있어야 하고 둘째 관심변수와 상관관계가 높아야 한다. 따라서 관심변수(Y)의 추정에 도움이 되는 벤치마킹 변수(X)를 찾아내는 방법이 연구되어야 하는데 이는 회귀분석 등에서 사용되는 방법론과 크게 다르지 않다. 다만 통계학 교과서에서 다루는 변수 선택 방법을 그대로 적용하기보다는 보다 실무에 유용한 방법론을 찾아내는 것이 필요할 것이다.

〈표 1-1〉 가중치 방법과 특성

추정량	방법	장점	단점
회귀 추정량	회귀식에 의한 추정	· 보조변수가 좋으면 추정량의 효율이 높음	· 약간의 편향 발생 · 가중치가 지나치게 크거나 작은 경우 발생
Exponential Tilting	회귀추정량과 유사	· 극단적 가중값에 덜 민감	· 범주형 자료에 사용불가
Raking ratio	반복비례가중	· 보조변수에 대해 모집단 비율과 일치하여 추정의 신뢰성 확보 · 범주형 변수 적용 · 계산이 쉬움	· 칸이 비어있거나 적은 경우에 추정 수렴불가
Composite calibration	두 조사를 통합한 새로운 추정량	· 분산을 최소화하는 추정량	· 시군구 통계 구현 불가
Projection	Imputation과 유사	· 두 자료 간 크기 차이가 큰 경우 효과적	· 좋은 프로젝션 모형 찾기가 어려움

6. 고용조사에서의 적용

가. 기존 방법으로부터 선택

지금까지 설명한 방법을 토대로 연구에 사용될 고용조사 자료와 경찰조사 자료에 비추어 적용가능성을 체크해 보자. 각 방법의 사용 가능성과 그 이유는 다음과 같이 정리



해 볼 수 있다.

〈표 1-2〉 추정 방법 사용 가능성

추정량	사용여부	근거
회귀추정량	어려움	· 연속형 보조 변수가 자료에 포함되어 있지 않음
Exponential Tilting	어려움	· 연속형 보조 변수는 자료에 포함되어 있지 않음
Raking ratio (반복비례가중법)	가능	· 모든 보조변수가 범주형으로 가장 쉽게 접근 가능 · 활용성 측면에서 고용조사에 적절
Composite calibration	어려움	· 고용조사자료를 그대로 쓰지 않기 때문에 현실적으로 받아들이기 어려울 것으로 예상
Projection	향후 고려	· 반복비례가중법과 크게 다르지 않음 · 좋은 예측 모형에 대한 충분한 검토 필요

이상과 같이 현재 보유하고 있는 자료의 특성을 고려할 때 raking ratio을 포함한 반복비례가중법이 현실적으로 가장 적합할 것으로 판단된다. 따라서 본 연구는 3절과 4절을 통해 다양한 형태의 반복비례가중법을 적용한 결과를 제시하였다.

나. 새로운 제안

주어진 자료의 특성을 고려하여 기존 방법들을 응용할 수 있는 방법을 생각할 수 있다. 소지역 추정을 이용한 가중치 조정이 그것이다. 여기서는 간단한 아이디어만 제시하고, 실제 자료분석에는 적용하지 않을 것이다. 향후 과제를 통해 이 아이디어에 대한 방법론적 또는 실용적 측면에서 그 가치를 증명할 예정이다.

고용조사는 대규모의 표본에서 여러 가지 보조 정보와 고용 관련 정보를 얻어내고 경찰조사는 작은 규모의 표본에서 동일한 정보를 얻어내지만 비표본 오차가 더 작은 것으로 간주할 수 있다. 경찰조사에서 얻어진 고용 관련 정보를 y 라 하고 고용 조사에서 얻어진 고용 관련 정보를 z 라 하자. 두 조사에서 공통적으로 얻어지는 보조 정보는 x 라고 하자. 즉, 경찰조사에서는 (x, y) 가 관측되고 고용조사에서는 (x, z) 가 관측된다.

이러한 경우 경찰조사 자료를 바탕으로 고용조사값을 보정하는 문제를 생각해 보도록 하자. 고용조사 표본을 S_1 이라고 하고 경찰조사 표본을 S_2 라고 하자. 고용조사의 설

계 가중치를 d_{i1} 이라고 하고 경찰조사의 설계가중치를 d_{i2} 이라고 하자. 또한 고용조사의 최종 가중치를 w_{i1} 이라고 하고 경찰조사의 최종 가중치를 w_{i2} 이라고 하자. 만약 모집단의 x 총계값을 안다면 이러한 최종 가중치는 다음을 만족하도록 결정되어야 한다.

$$\sum_{i \in S_{h1}} w_{i1} x_i = \sum_{i \in S_{h2}} w_{i2} x_i = \sum_{i \in U_h} x_i \quad (16)$$

여기서 h 는 대지역 (시도)를 나타내는 첨자이다. 만약 $X_h = \sum_{i \in U_h} x_i$ 를 모른다면 두 표본을 통합하여 식 (11)과 식 (13)에서처럼 $X_{h,opt}$ 을 구현해야 할 것이다.

또한 y 와 z 는 동일한 항목을 조사하는 것이므로 식 (16)의 calibration 조건 외에도 다음과 같은 calibration constraint을 생각할 수 있다.

$$\sum_{i \in S_{1h}} w_{1i} z_i = \sum_{i \in S_{2h}} w_{2i} y_i = Y_{h,opt}$$

따라서,

$$Y_{h,opt} = \alpha_h \left(\sum_{i \in S_{1h}} d_{i1} \hat{y}_i \right) + (1 - \alpha_h) \left(\sum_{i \in S_{2h}} d_{i2} y_i \right)$$

을 사용한다. 여기서 $\hat{y}_i = z_i R_h$ 는 z_i 를 바탕으로 한 y_i 의 예측치로 ratio 값인 R_h 를 곱해서 구하고, R_h 는 지역 h 에서의 y 평균 추정치를 동일 지역의 z 평균 추정치로 나누어서 계산한다. 최적 계수 α_h^* 는 두 추정량의 분산값을 바탕으로 구현할 수 있으나 고용조사의 표본수가 상대적으로 크므로 $\alpha_h^* \approx 1$ 을 사용하면 된다. 또는 표본수에 비례한 가중 평균을 사용해도 될 것이다.

제3절 반복비례가중법

1. 개요

표본조사를 수행하기 위해서는 먼저 모집단의 특성을 반영한 표본을 선정해야 하며, 모집단과 유사한 형태의 표본을 추출하였을 때 해당 표본이 모집단에 대한 대표성을 갖



는다고 한다. 추출된 표본이 모집단에 대한 대표성을 가지고 있을 때 표본조사 결과를 전체 모집단으로 일반화할 수 있다. 따라서 모집단을 닮은 표본을 추출하기 위한 다양한 형태의 확률 표본 추출 방법들이 표본설계에 이용되고 있다.

모집단에 대한 대표성을 가지는 표본을 추출하기 위해서는 표본조사 목적과 연관관계가 높은 일부 인구사회학적 특성들을 특성지표로 활용하게 되며, 표본 추출에 고려된 특성지표에 대해서는 표본이 모집단에 대한 대표성을 가지게 된다. 그러나 현실적으로 모집단의 모든 인구사회학적 특성을 고려한 표본을 추출한다는 것은 불가능한 일이므로 표본 추출 과정에 고려되지 않은 인구사회학적 변수에 대해서는 표본이 모집단에 대한 대표성을 가지지 못하게 된다. 이러한 차이를 보정하기 위하여 사후에 다양한 형태의 가중치 보정 작업을 수행하게 된다.

잘 알려진 가중치 보정 방법으로는 사후층화(post-stratification)와 반복비례가중법(iterative proportional weighting method) 등이 있다. 사후층화는 모집단의 결합분포와 표본의 결합분포를 비교하여 그 비를 가중치로 이용하는 방법으로, 모집단의 다차원 결합분포가 알려진 경우에 유용하게 이용될 수 있다. 이 방법은 모집단과 표본의 인구사회학적 특성의 다차원 결합분포 내 각 셀의 비율을 조정하는 방법을 이용하므로 칸 가중법(cell weighting)이라고 부르기도 한다. 그러나 모집단의 다차원 결합분포가 알려진 경우가 드물며, 표본의 다차원 결합분포의 칸이 비어있는 경우 이용이 어렵다는 단점이 있다. 반복비례가중법은 모집단의 인구사회학적 특성의 주변분포를 알고 있는 경우 적용할 수 있는 사후 가중치 보정 방법으로, 각 특성 변수의 주변분포를 이용하여 반복적으로 칸 내의 모집단수를 구하는 방법이다. 이 방법은 모집단의 주변분포만 알면 적용이 가능하므로 고려할 특성 변수가 많은 경우에도 활용이 가능하다. 반복비례가중법은 Deming and Stephan(1940)이 처음 제안 하였으며, 적용 방법의 유용성 때문에 다양한 분야에 걸쳐 활발하게 이용되고 있다. 림가중법(rim weighting) 또는 갈퀴법(raking method)으로도 잘 알려진 반복비례가중법의 절차를 세부적으로 살펴보자.

2. 추정과정

반복비례가중법은 모집단이 2개 이상의 특성변수에 의해 분할표 형태로 나뉘고, 모집단의 특성 변수에 대한 1차원 주변분포만 알려져 있는 경우에 각 칸의 표본수와 모집단의 주변분포를 이용하여 1번에 1변수씩 표본의 비율을 모집단의 비율과 일치시키는 과정을 반복적인 알고리즘을 이용하여 구하는 방법이다. 이 방법은 단순임의추출(SRS)에 의한 표본뿐 아니라 일반적인 추출법에 의한 표본에도 적용이 가능하며, 모집단의 1차원 주변분포만 알려져 있다면 적용이 가능하다는 장점이 있다.

반복비례가중법의 추정과정에 대한 이해를 돕기 위하여 2차원 분할표를 이용한 예제를 살펴보자.

모집단을 $U_{R \times C}$ 형태의 2차원 분할표라 하고, $U_{rc}(r=1, \dots, R, c=1, \dots, C)$ 를 크기가 N_{rc} 칸의 모집단수인 모집단의 (r, c) 칸이라고 하자. 표본 s 는 단순임의추출법에 의해 추출되며, $s_{rc} = s \cap U_{rc}$ 는 크기가 n_{rc} 칸의 표본수인 (r, c) 칸의 표본이다. 또한 표집 방법에 의해 결정되는 단위 k 의 조사 가중값을 w_k 라고 하고, 우리의 관심변수를 y 라고 하면, 관심변수의 모집단 총합 $\sum_{k \in U} y_k$ 의 추정량은 $\sum_{k \in s} w_k y_k$ 가 된다. 단순임의추출을 시행한 경우, 단위 k 의 조사 가중값은 $w_k = N/n$ 이 된다.

우리의 목적은 표본수 n_{rc} 로부터 미지인 각 칸의 모집단수, N_{rc} 를 추정하는 것이다. 이 때 모집단의 1차원 주변분포가 알려져 있으므로 다음의 제약식을 이용할 수 있다.

$$N_{r+} = \sum_{c=1}^C \hat{N}_{rc}, \quad N_{+c} = \sum_{r=1}^R \hat{N}_{rc}$$

이들 제약식은 각 칸의 보정된 모집단수의 주변합을 실제 모집단수와 일치시키는 역할을 하게 된다.

단순임의추출에 의한 조사 가중값이 $w_k = N/n$ 이므로, 반복비례추정을 위한 최초의 가중치는 $w_0 = N/n$ 으로 하면, 다음과 같은 반복 과정을 통하여 최종 조정 가중값을 얻을 수 있다.

$$w^{(1,1)} = \left[\frac{N_{r+}}{\sum_{c=1}^C \hat{N}_{rc}^{(0)}} \right] \times w_0$$

$$w^{(1,2)} = \left[\frac{N_{+c}}{\sum_{r=1}^R \hat{N}_{rc}^{(1,1)}} \right] \times w^{(1,1)}$$

$$\vdots$$

여기서 $\hat{N}_{rc}^{(0)}$ 는 w_0 를 이용한 N_{rc} 의 추정값이며, $\hat{N}_{rc}^{(i,j)}$, $(i=1, \dots, j=1, \dots)$ 은 $w^{(i,j)}$ 를 이용한 N_{rc} 의 추정값이다.

위와 같은 알고리즘이 t 번 반복된 시점에서 모집단수에 대한 (r, c) 번째 칸의 반복비



레추정량을 $\hat{N}_{rc}^{(t)}$ 라고 하면, 추정량 $\hat{N}_{rc}^{(t)}$ 가 수렴할 때까지 반복하여 최종 조정 가중값인 $w^{(t)}$ 를 얻게 된다. 이렇게 얻은 최종 조정 가중값을 이용하면 모집단 총합 $\sum_{k \in U} y_k$ 의 추정량을 $\sum_{k \in s} w_k y_k = \sum_{k \in s} w^{(t)} y_k$ 로 구할 수 있다.

반복비례가중법에 대한 이해를 돕기 위하여 모집단의 크기가 900이고 표본의 크기가 120인 2차원 분할표를 생각해보자. 각 특성 변수의 수준은 2개이며 표본조사 결과와 각 특성 변수에 대한 모집단의 주변분포는 <표 1-3>-<표 1-4>와 같다.

<표 1-3> 표본조사 결과

	A1	A2	합계
B1	20	30	50
B2	35	35	70
합계	55	65	120

<표 1-4> 모집단의 주변분포

	A1	A2	합계
B1			400
B2			500
합계	450	450	900

Step1) 초기 가중치 $w_0 = N/n = 900/120 = 7.5$ 를 구한다.

Step2) 표본의 각 칸에 초기 가중치를 곱한 후, A 특성변수에 대한 주변분포를 구한다.

Step3) Step2에서의 A 특성변수의 주변분포와 모집단의 A 특성변수에 대한 주변분포의 비로 각 칸을 재조정 한 후, B 특성변수에 대한 주변분포를 구한다.

Step4) Step2와 Step3가 다음의 수렴조건을 만족할 때까지 반복한다.

$$\left| \sum_{r(c)} \hat{N}_{rc} - N \right| \times 100 < 0.01(\%)$$

위 예제를 9번 반복하면 <표 1-5>와 같은 결과를 얻게 된다.

〈표 1-5〉 반복비례가중법을 이용한 추정값(9번 반복)

	A1	A2	합계
B1	178	222	400
B2	272	228	500
합계	450	450	900

3. 2차원 기준 반복비례가중법

3.2절에서 살펴본 반복비례가중법은 모집단의 특성 변수에 대한 1차원 주변분포를 알고 있는 경우에 활용되는 사후 가중치 보정 방법이다. 그러나 만약 고려하고 있는 모집단의 특성변수에 대한 다차원 결합분포는 알려지지 않으나 각 특성변수들의 2차원 결합분포는 알려져 있는 경우라면, 이들 2차원 결합분포를 반복적으로 적용하여 가중치를 수정해가는 방법을 이용할 수 있을 것이다. 이에 허명회 외(2005)는 2차원 기준 반복비례가중법을 제안하였다.

2차원 반복비례가중법은 관심을 가지고 있는 특성 변수의 1차원 주변분포뿐 아니라 각 특성 변수들의 2차원 결합분포도 알려져 있는 경우에 활용이 가능하다. 즉, 기존의 반복비례가중법을 2차원 결합분포를 이용한 경우로 확장하는 방법이다. 허명회 외가 제안한 2차원 기준 반복비례가중법을 소개하면 다음과 같다.

먼저 변수가 3개인 경우를 고려하자. 이 경우 기존의 반복비례가중법을 적용하기 위해서는 3개 변수의 모집단 주변분포만 알면 된다. 이 때 2차원 결합분포의 모집단수를 알고 있다고 가정하면, 다음과 같은 제약식을 고려할 수 있다.

$$N_{.jk} = \sum_{i=1}^I \hat{N}_{ijk}, \quad N_{i.k} = \sum_{j=1}^J \hat{N}_{ijk}, \quad N_{ij.} = \sum_{k=1}^K \hat{N}_{ijk}$$

2차원 반복비례가중법의 목적 또한 앞 장에서와 마찬가지로 각 칸의 표본수와 위의 제약식을 이용하여 각 칸의 모집단수를 추정하는 것이며, 반복비례가중법의 반복 알고리즘을 2차원 결합분포에 대한 것으로 확장, 적용하면 모집단수에 대한 추정값을 구할 수 있다.

관심을 가지고 있는 특성 변수가 3개인 경우, 3차원 분할표 형태의 모집단수 N_{ijk} 는 1차원적 효과뿐 아니라 2, 3차원의 결합효과도 포함하고 있으므로, 1차원 주변분포만을 이용하는 기존의 반복비례가중법에 비하여 더욱 모집단수에 근접할 가능성이 크다고 할 수 있다. 그러나 1차원 주변분포만을 이용하는 기존의 반복비례추정법과 더불어 2차원



반복비례가중법을 이용한 추정량 또한 모집단수에 대한 일치추정량은 아니다. 2차원 반복비례가중법을 위한 모집단의 2차원 주변분포의 수는 고려 대상인 특성 변수가 4개인 경우 ${}_4C_2 = 6$ 개이다. 즉, 특성 변수의 수가 4개 이상인 경우에도 같은 알고리즘을 적용할 수는 있으나, 이 경우 고려해야 하는 모집단의 2차원 주변분포의 수가 크게 증가함으로 반복 알고리즘의 계산량이 크게 증가될 수 있으며, 이에 따라 추정량을 구하기 위한 시간 또한 크게 증가하는 단점이 있다.

기존의 반복비례가중법은 각 칸의 표본에 모집단의 주변분포를 이용한 비추정량을 곱하는 과정을 반복함으로써 모집단에 대한 칸의 추정량을 구하게 된다. 따라서 표본수가 0인 칸의 최종 모집단 추정값은 0이 된다. 이 경우 기존의 반복비례가중법이나 2차원 기준 반복비례가중법 모두 같은 결과를 가져오게 된다. 만약 1차원 주변분포에서의 칸의 표본수는 0이 아닌데, 2차원 주변분포에서의 칸의 표본수가 0인 경우에는 기존의 반복비례가중법과 2차원 기준 반복비례가중법이 서로 다른 결과를 가져오게 될 것이다. 허명희 외는 이러한 문제가 발생할 경우, 표본수가 0이 되는 일부 2차원 칸의 변수에 대해 기존의 방법을 적용하거나, 표본수가 0인 칸은 모집단수도 매우 작은 경우가 대부분이므로 해당되는 칸의 손실을 감내하는 것도 현실적인 방법이라고 제안하였다. 또 다른 제안으로는 인근 칸과 병합하여 표본수가 0인 칸을 제거하는 방법이 있다.

제4절 사례 분석

1. 분석 자료

고용조사는 지역 고용 정책 수립에 필요한 시군 단위의 세분화된 고용 현황을 파악할 수 있는 기본 통계를 생산하고 제공하기 위한 목적을 가지고 있으며, 시도별 고용 구조 분석 자료 및 산업·직업에 대한 세분화된 자료를 생산하고 있다. 2008년 10월 시작된 고용조사는 매년 동월에 조사가 이루어지며, 조사 대상 기간을 기준으로 대한민국에 상주하는 만 15세 이상인 자¹⁾를 대상으로 경제활동상태와 관련된 항목을 조사하고 있다.

경활조사는 국민의 경제활동 즉, 국민의 취업·실업 등과 같은 경제적 특성을 조사하여 거시경제 분석과 인력자원의 개발 정책 수립에 필요한 기초 자료인 노동공급, 고용구조, 가용노동시간 및 인력자원 활용 정도를 제공하고 정부의 고용정책입안 및 평가에 필요한 기초 자료를 제공하기 위한 목적으로 매월 조사가 이루어지고 있다. 경활조사는 매월 15일 현재 대한민국에 상주하는 만 15세 이상인 자를 대상으로 취업, 실업 및 비경제

1) 현역군인, 경비교도대, 전투의무경찰, 공익근무요원, 교도소수감자 등은 제외

활동 등과 관련된 항목을 조사하고 있으며, 실업자, 취업자 및 비경제활동 현황 등이 특성별로 공표되고 있다.

앞서 소개한 바와 같이 고용조사와 경황조사에서는 경제활동상태 관련 항목들을 포함한 대부분의 조사 항목이 동일하며, 목표모집단과 조사 대상 또한 동일하다. 두 조사의 차이점은 조사 목적과 조사주기 그리고 공표 가능한 지역의 범위가 다르다는 것이다. 이 때 고용조사의 경우 시군 단위의 통계 생산을 목적으로 조사되고 공표도 이루어지지만 시도 단위에서의 통계 생산도 자연스럽게 가능해진다. 따라서 고용조사가 이루어지는 2008년 10월에는 우리나라의 경제활동 관련 국가 통계가 시도 단위에서 두 개가 되며, 이 두 통계는 서로 다른 목적을 가지고 각각의 표본을 이용하여 조사된 것이므로 추정 결과가 일치하지 않게 된다. 이 때 경제활동 관련 통계를 산출하는 두 국가통계의 추정결과가 조사가 이루어지는 동일 시점에 대하여 어느 정도의 일관성을 유지하는 것은 통계에 대한 신뢰나 이용자 혼돈을 방지하는 차원에서 매우 바람직하다. 따라서 고용조사가 이루어진 2008년 10월 현재 16개 시도 단위에서의 경황상태 인구나 경황조사에서의 경황상태 인구가 서로 일관성을 유지하도록 보정하는 작업이 필요하다.

다음 절에서는 두 조사 자료의 경황상태 인구를 보정하는 방안에 대하여 살펴보도록 한다. 먼저 경황상태에 영향을 미칠 것으로 보이는 특성 변수를 선택하고, 이어 인구사회학적 특성 변수인 지역, 성, 연령과 경황조사의 경제상태 인구 및 영향력이 크게 나타난 변수들을 고려한 보정 방안에 대하여 소개한다. 보정 방안은 크게 1:2차원 기준 반복비례가중법과 반복비례가중법에 비조정을 결합시킨 혼합형 조정 방법으로 나누어 진행한다.

2. 변수(인구특성) 선택 및 분석 결과

고용조사와 경황조사의 조사 항목들 중 인구사회학적 특성 변수들은 응답자의 응답 부담 등을 고려하여 극히 제한적으로 이루어지고 있다. 실제 조사가 이루어지는 인구사회학적 특성 변수들은 <표 1-6>과 같다.

경황조사에서 가구 구분(농가여부)을 조사하는 것을 제외하면, 두 조사에서 조사되는 인구사회학적 특성 변수들에는 차이가 없으며 매우 제한적이라는 것을 알 수 있다. 표본 관리명부의 정보를 이용하면 표본에 대한 특성 변수를 몇 가지 더 추가할 수 있으나, 여기서 추가할 수 있는 특성 변수가 거주구분, 주택소유관계, 1인 가구 여부 등 가구에 관한 것이 대부분이므로, 가구원 단위의 경황상태를 보정하는 과정에 큰 역할을 기대할 수 없다.



〈표 1-6〉 고용조사 및 경찰조사의 인구사회학적 특성 변수

구분	고용조사	경찰조사
가구	지역, 총가구원수 (15세 미만/이상)	지역, 총가구원수 (15세 미만/이상), 농가여부
가구원	성별, 연령, 국적, 가구주와의 관계, 교육상태, 혼인상태	성별, 연령, 국적, 가구주와의 관계, 교육상태, 혼인상태

가. 변수선택

〈표 1-6〉에 있는 특성변수들 중 교육상태 및 혼인상태가 경찰상태에 미치는 영향을 확인하고, 해당 변수들이 영향력 있는 특성변수로 확인될 경우 표본 설계에 이용된 지역, 성, 연령과 함께 벤치마킹 변수로 이용하도록 한다.

일반적으로 두 범주형 변수의 연관성 검정은 카이제곱 검정을 이용하게 되나 표본의 수가 큰 경우에는 그 결과에 대한 신뢰도가 크게 떨어진다. 따라서 본 연구에서는 카이제곱 검정 대신 특성변수의 각 수준을 더미변수(가변수, 지시변수)로 변환하여 t검정을 시행하였다.

교육상태 및 혼인상태에 대한 더미변수는 다음과 같이 정의한다.

$$X_i = \begin{cases} 1, & i = 1(2, \dots, 6) \\ 0, & i \neq 1(2, \dots, 6) \end{cases}$$

$$Y_j = \begin{cases} 1, & j = 1(2, 3) \\ 0, & j \neq 1(2, 3) \end{cases}$$

여기서 $i = 1, \dots, 6$ 은 각각 초졸, 중졸, 고졸, 초대졸, 대졸, 대학원졸 이상이며, $j = 1, 2, 3$ 은 각각 미혼, 유배우, 무배우인 경우를 나타낸다.

다음은 교육상태가 경찰상태에 미치는 영향력을 검정한 결과이다. 검정 결과, 고용조사의 경우에는 교육상태의 모든 수준에서 경찰상태와 유의미한 연관성이 있는 것으로 나타났으며, 경찰조사의 경우에는 대졸인 경우를 제외한 나머지에서 유의미한 연관성이 확인되었다. 즉, 교육상태가 경찰상태에 영향력 있는 변수라고 할 수 있다.

<표 1-7> 교육상태가 경활상태에 미치는 영향

교육상태	경활 상태	고용조사				경활조사			
		표본수	평균	표준 편차	t	표본수	평균	표준 편차	t
초졸	취업자	218,405	0.236	0.425	46.37**	39,105	0.162	0.369	11.99**
	실업자	4,864	0.066	0.248		1,056	0.067	0.251	
중졸	취업자	218,405	0.124	0.329	6.83**	39,105	0.114	0.318	2.55*
	실업자	4,864	0.095	0.293		1,056	0.091	0.288	
고졸	취업자	218,405	0.358	0.479	-16.57**	39,105	0.392	0.488	-6.23**
	실업자	4,864	0.478	0.500		1,056	0.487	0.500	
초대졸	취업자	218,405	0.089	0.284	-11.65**	39,105	0.105	0.307	-4.8**
	실업자	4,864	0.149	0.356		1,056	0.160	0.367	
대졸	취업자	218,405	0.167	0.373	-5.62**	39,105	0.194	0.395	0.91
	실업자	4,864	0.120	0.400		1,056	0.183	0.387	NS
대학원졸	취업자	218,405	0.026	0.158	7.84**	39,105	0.033	0.178	5.8**
	실업자	4,864	0.013	0.112		1,056	0.012	0.110	

*p<0.05, **p<0.01

<표 1-8>은 고용조사 및 경활조사에서 혼인상태가 경활상태에 미치는 영향을 검정한 결과이다.

<표 1-8> 혼인상태가 경활상태에 미치는 영향

혼인 상태	경활 상태	고용조사				경활조사			
		표본수	평균	표준 편차	t	표본수	평균	표준 편차	t
미혼	취업자	218,405	0.150	0.357	-48.91**	39,105	0.173	0.379	-19.79**
	실업자	4,864	0.503	0.500		1,056	0.480	0.500	
유배우	취업자	218,405	0.749	0.433	48.26**	39,105	0.739	0.439	20**
	실업자	4,864	0.407	0.491		1,056	0.431	0.495	
기타	취업자	218,405	0.101	0.301	2.35*	39,105	0.088	0.283	-0.15
	실업자	4,864	0.091	0.288		1,056	0.089	0.285	NS

*p<0.05, **p<0.01



<표 1-7>에서와 같이 혼인상태의 각 수준을 더미변수로 변환하여 t검정을 시행하였으며, 검정 결과, 고용조사의 경우에는 혼인상태의 모든 수준에서 경활상태와 유의미한 연관성이 있는 것으로 나타났다. 경활조사의 경우에는 기타인 경우에서 유의미한 연관성을 찾지 못하였으나, 결과를 종합적으로 살펴보면 혼인상태가 경활상태에 영향력 있는 변수임을 알 수 있다.

<표 1-7>과 <표 1-8>의 결과를 바탕으로 교육상태와 혼인상태가 경활상태에 영향력이 있는 특성변수임을 알 수 있다. 따라서 가중치 보정 작업에 인구사회학적 특성 변수인 지역, 성, 연령과 함께 교육상태와 혼인상태 등을 벤치마킹 변수로 고려하도록 한다.

나. 보정 방안

고용조사의 지역별(16개 시도별) 경활상태를 보정하기 위한 가중치 보정 방법으로 가장 먼저 고려할 수 있는 방법은 고용조사에서 추정되는 만 15세 이상 인구에 대한 지역별, 성별, 연령별, 경활상태의 분포가 현재의 만 15세 이상 인구의 모집단 분포와 일치하도록 가중치를 조정하는 방법이다. 이때 만 15세 이상 인구의 지역별, 성별, 연령별 모집단 분포는 인구추계를 통하여 알 수 있으나, 경활상태에 대한 모집단 분포는 알 수 없는 상황이다. 이에 경활조사에서의 경활상태 분포를 준거표본으로 이용하여 가중치 보정 작업에 이용하도록 한다. 준거표본으로 이용하고자 하는 경활조사는 1957년 「노동력조사」라는 명칭으로 처음 시행된 이래 매일 조사가 이루어지고 있다. 따라서 조사의 전통과 지속성(시계열 유지), 그리고 우리나라의 대표적인 경제활동 관련 통계결과로써 고용조사의 경활상태에 대한 준거표본으로 경활조사의 경활상태를 이용하는 것이 바람직하다고 할 수 있다. 또한 고용조사와 경활조사에서 동시에 조사되고 있는 인구사회학적 특성 중 교육상태와 혼인상태는 두 통계의 경활상태에 유의미한 영향력을 미치는 것으로 확인되었으므로, 고용조사의 가중치 보정 작업에 포함시키도록 한다. 결과적으로 총 6개의 벤치마킹 변수(지역, 성, 연령, 경활상태, 교육상태, 혼인상태)를 이용하여 다음과 같은 가중치 보정 방안을 검토한다.

- 방안1) 지역, 성, 연령 및 경활상태 분포를 이용한 반복비례가중법
- 방안2) 지역, 성, 연령, 교육상태, 혼인상태 및 경활상태 분포를 이용한 반복비례가중법
- 방안3) 지역, 성, 연령 및 경활상태의 2차원 결합분포를 이용한 2차원 기준 반복비례가중법
- 방안4) 지역별 성, 연령, 교육상태 및 혼인상태를 이용한 반복비례가중법을 수행 후 경활상태를 이용한 비조정

사후층화(post-stratification)에 의한 가중치 보정 방안은 비는 칸(empty cell)이 다수 발

생하게 되어 보정에 좋지 않은 영향을 미치는 것으로 확인되었으므로, 보정 방법들의 비교에서는 제외하기로 한다. 각 가중치 보정 방법을 자세히 살펴보면, 방안1과 방안2는 립가중법 또는 레이킹비로 잘 알려진 반복비례가중법을 각각 4개 벤치마킹 변수(지역, 성, 연령, 경찰상태)와 6개 벤치마킹 변수(지역, 성, 연령, 경찰상태, 교육상태, 혼인상태)의 1차원 주변분포를 이용하여 진행한 것이며, 방안3은 4개 벤치마킹 변수(지역, 성, 연령, 경찰상태)의 2차원 결합분포(${}_4C_2 = 6$ 개)를 이용하여 반복비례가중법을 적용한 것이다. 이때 6개의 벤치마킹 변수를 이용할 경우, ${}_6C_2 = 15$ 개의 결합분포를 이용하여 반복비례가중법을 적용해야 하므로 수렴을 위한 시간이 너무 길어지게 된다. 본 연구는 실제로 고용조사의 경찰상태 분포와 경찰조사의 경찰상태 분포를 기준으로 보정하는 작업에 이용되는 것을 목적으로 하고 있으므로, 보정 작업에 지나치게 많은 시간이 소요되는 방법은 고려하지 않기로 한다. 마지막으로 고용조사 자료를 지역별로 분류하여 반복비례가중법을 수행한 후 최종적으로 경찰상태에 대한 비조정을 시행하는 방안을 방안4로 고려하였다.

다. 분석결과

다음 <표 1-9>는 4가지 방안을 이용하여 전국 단위로 조정한 결과이다.

<표 1-9> 가중치 비교(전국)

방안	경찰상태	평균값	표준편차	변동계수	합계*
고용조사	취업	108.945	101.027	92.732	22,867
	실업	151.796	109.147	71.903	706
	비경찰	117.212	97.963	83.578	16,153
방안1	취업	113.612	104.526	92.002	23,847
	실업	158.132	113.953	72.062	736
	비경찰	109.889	91.459	83.228	15,144
방안2	취업	113.612	109.611	96.478	23,847
	실업	158.132	124.155	78.513	736
	비경찰	109.889	96.095	87.447	15,144
방안3	취업	113.612	104.039	91.574	23,847
	실업	158.133	114.938	72.685	736
	비경찰	109.889	94.266	85.782	15,144
방안4	취업	113.612	109.824	96.666	23,847
	실업	158.132	122.696	77.591	736
	비경찰	109.889	97.644	88.857	15,144

* 단위: 천 명



보정 전 고용조사의 실업자는 706천 명이었으나 보정 후 736천 명으로 약 30천 명이 증가 했으며, 취업자는 22,867천 명에서 23,847천 명으로 약 980천 명 증가하였다. 이에 반해 비경활인구는 16,153천 명에서 15,144천 명으로 약 1,010천 명 줄어들었음을 확인할 수 있다. 즉, 고용조사에서의 경활인구(실업자 및 취업자)는 경활조사에 비해 과소 추정되어있으며, 비경활인구는 과대 추정되어 있는 상태임을 알 수 있다. 이와 같은 결과는 고용조사가 시군 단위의 소지역 추정을 목적으로 하기 때문에 군 단위에서의 표본 비중이 경활조사에 비해 크기 때문인 것으로 보인다.

<표 1-9>의 가중치 비교 결과 중 취업자의 경우를 살펴보면, 고용조사 원가중치에 비해 방안3을 이용한 보정 가중치의 변동계수가 92.732에서 91.574로 다소 줄어들었음을 알 수 있다. 방안1의 경우에도 보정 이후의 변동계수가 다소 줄어 4개의 벤치마킹 변수를 이용하는 경우 취업자의 보정 가중치가 원가중치에 비해 더 안정된다는 것을 알 수 있다. 이에 반해 벤치마킹 변수를 6개 이용하는 방안2와 방안4의 경우에는 변동계수가 각각 96.478과 96.666으로 다소 커지는 것을 알 수 있다. 실업자의 경우에는 방안1과 방안3의 변동계수가 보정 전 71.903에서 보정 후 72.062와 72.685로 다소 커졌으나 거의 변화하지 않음을 알 수 있으며, 방안2와 방안4의 경우에는 변동계수가 78.513과 77.591로 원가중치에 비해 다소 증가함을 알 수 있다. 마지막으로 비경활인구의 경우에서도 방안1의 변동계수가 83.228로 보정 전 83.578과 거의 유사함을 알 수 있으며, 방안3의 변동계수는 85.782로 다소 커진 것을 알 수 있다. 방안2와 방안4의 경우에는 변동계수가 각각 87.447과 88.857로 나타나 벤치마킹 변수의 수가 보정 가중치의 변동폭에 영향을 미치고 있음을 알 수 있다. 가중치의 변동이 커지는 원인으로서는 고려하는 벤치마킹 변수의 수와 준거표본 분포의 이용 등을 들 수 있다. 즉, 너무 많은 벤치마킹 변수를 고려하거나 신뢰할 수 있는 모집단이 아닌 다른 준거 분포를 이용할 경우 가중치의 변동이 커지게 되어 결국에는 추정의 효율성이 떨어질 우려가 있는 것이다. 따라서 벤치마킹 변수의 선택과 이용 가능한 준거 분포의 신뢰성이 가중치 보정에 매우 중요한 역할을 하게 된다.

가중치 비교 결과를 전반적으로 살펴보면, 방안3의 경활상태별 변동계수가 보정 전과 유사하거나 더 개선되는 것으로 나타났다. 그러나 보정 방안에 따른 변동성의 차이는 전반적으로 그다지 크지는 않은 것으로 보이며, 이와 같은 결과는 고용조사의 표본 크기가 커 가중치 보정 방법에 따른 영향을 크게 받지 않았기 때문인 것으로 보인다. 이에 지역별로 가중치 보정 방안을 적용하고 그 결과를 비교해보도록 한다.

〈표 1-10〉 기중치 비교(서울, 부산, 대구, 인천)

지역	방안	취업				실업				비경제활동			
		평균값	표준편차	변동계수	합계	평균값	표준편차	변동계수	합계	평균값	표준편차	변동계수	합계
서울	고용조사	274.755	67.540	24.582	4,790	276.870	64.953	23.460	188	266.810	65.479	24.541	3,296
	방안1	285.654	69.411	24.299	4,980	288.651	67.348	23.332	196	250.782	60.751	24.224	3,098
	방안2	284.627	99.261	34.874	4,962	293.297	104.512	35.634	199	251.976	83.626	33.188	3,113
	방안3	281.494	73.014	25.938	4,908	274.267	74.305	27.092	186	257.441	68.673	26.675	3,180
부산	방안4	281.494	102.786	36.514	4,908	274.267	104.213	37.997	186	257.441	90.130	35.010	3,180
	고용조사	164.263	70.511	42.926	1,509	166.981	67.774	40.588	53	161.018	64.057	39.782	1,334
	방안1	171.807	72.654	42.288	1,578	174.557	69.956	40.076	55	152.363	60.330	39.596	1,262
	방안2	171.323	85.709	50.028	1,574	174.905	88.769	50.753	55	152.886	69.391	45.387	1,267
대구	방안3	170.777	73.799	43.213	1,569	196.051	85.266	43.492	62	152.682	62.105	40.676	1,265
	방안4	170.777	82.631	48.385	1,569	196.049	96.574	49.260	62	152.682	67.634	44.297	1,265
	고용조사	224.521	91.033	40.545	1,108	234.177	92.379	39.448	41	218.003	85.752	39.335	875
	방안1	234.540	93.339	39.796	1,157	244.911	97.168	39.675	43	205.223	80.505	39.228	824
인천	방안2	233.559	107.909	46.202	1,153	246.022	110.641	44.972	43	206.380	94.459	45.769	829
	방안3	230.807	92.864	40.234	1,139	263.646	108.559	41.176	46	208.994	86.326	41.305	839
	방안4	230.808	111.799	48.438	1,139	263.644	123.912	47.000	46	208.994	98.101	46.939	839
	고용조사	194.340	114.408	58.870	1,239	228.716	105.968	46.331	51	199.172	102.959	51.694	869
인천	방안1	202.598	118.553	58.516	1,292	239.227	109.493	45.770	53	186.573	96.329	51.631	814
	방안2	202.599	131.610	64.961	1,292	236.660	125.554	53.053	52	186.701	106.438	57.010	815
	방안3	201.547	119.469	59.276	1,285	222.050	109.714	49.410	49	188.978	100.675	53.273	825
	방안4	201.547	130.572	64.785	1,285	222.049	117.672	52.994	49	188.978	108.312	57.315	825

<표 1-11> 기중치 비교(광주, 대전, 울산, 경기)

지역	방안	취업			실업			비경제활동					
		평균값	표준편차	변동계수	합계	평균값	표준편차	변동계수	합계	평균값	표준편차	변동계수	합계
광주	고용조사	216.243	71.449	33.041	624	224.056	95.589	42.663	24	205.586	75.830	36.885	488
	방안1	226.308	73.504	32.479	653	236.099	98.835	41.862	26	192.795	68.713	35.640	457
	방안2	222.607	94.372	42.394	642	237.980	117.811	49.505	26	197.209	84.730	42.965	468
	방안3	220.975	73.694	33.350	637	230.523	93.411	40.521	25	199.535	74.846	37.510	473
대전	방안4	220.975	94.153	42.608	637	230.521	112.626	48.857	25	199.535	87.592	43.898	473
	고용조사	223.297	70.085	31.387	689	243.046	81.516	33.539	22	210.730	64.211	30.471	499
	방안1	233.117	71.583	30.707	719	253.226	80.188	31.667	23	197.558	59.041	29.885	468
	방안2	231.476	92.960	40.160	714	248.163	110.473	44.516	22	199.886	70.690	35.365	474
울산	방안3	228.913	70.934	30.987	706	270.731	86.121	31.811	24	202.376	67.946	33.574	479
	방안4	228.913	90.755	39.646	706	270.731	118.270	43.685	24	202.376	74.594	36.859	479
	고용조사	164.762	49.839	30.249	519	181.497	59.643	32.862	24	164.298	51.781	31.516	345
	방안1	171.812	50.500	29.393	542	189.900	58.139	30.616	25	153.180	46.872	30.599	321
경기	방안2	170.697	64.489	37.780	538	186.155	75.195	40.394	24	155.089	56.950	36.721	325
	방안3	168.974	49.705	29.416	533	176.588	55.317	31.325	23	158.276	51.085	32.276	332
	방안4	168.974	65.221	38.598	533	176.590	75.051	42.500	23	158.275	60.868	38.457	332
	고용조사	149.788	110.050	73.471	5,320	169.999	115.480	67.930	154	146.315	104.763	71.601	3,770
경기	방안1	156.082	113.749	72.878	5,544	178.095	120.907	67.889	161	137.355	97.590	71.050	3,539
	방안2	156.292	121.115	77.493	5,551	178.237	128.668	72.190	161	137.060	101.546	74.089	3,531
	방안3	158.462	115.206	72.703	5,628	179.905	124.497	69.201	162	134.012	96.854	72.273	3,453
	방안4	158.462	121.508	76.680	5,628	179.905	129.712	72.100	162	134.012	99.965	74.594	3,453

〈표 1-12〉 가중치 비교(강원, 충북, 충남, 전북)

지역	방안	취업			실업			비경제활동					
		평균값	표준편차	변동계수	합계	평균값	표준편차	변동계수	합계	평균값	표준편차	변동계수	
강원	고용조사	41.423	33.214	80.182	663	55.044	40.982	74.453	11	44.131	34.099	77.267	495
	방안1	43.340	34.119	78.724	694	56.512	41.698	73.787	11	41.369	31.071	75.107	464
	방안2	43.571	35.501	81.478	697	56.534	44.751	79.157	11	41.038	31.712	77.276	460
	방안3	44.194	34.313	77.643	707	73.591	54.742	74.386	15	39.841	31.508	79.083	447
	방안4	44.194	35.686	80.749	707	73.592	57.101	77.591	15	39.841	30.589	76.778	447
충북	고용조사	57.535	58.924	102.414	708	70.107	68.120	97.166	18	62.080	64.064	103.197	478
	방안1	60.145	60.471	100.543	741	73.017	71.371	97.746	19	57.808	58.254	100.771	445
	방안2	60.541	60.512	99.953	745	73.419	74.365	101.289	19	57.161	57.380	100.383	440
	방안3	59.686	59.288	99.333	735	66.580	66.017	99.154	17	58.756	61.929	105.399	452
	방안4	59.686	61.931	103.761	735	66.581	70.365	105.684	17	58.756	60.605	103.146	452
충남	고용조사	56.262	43.238	76.851	959	72.046	48.329	67.080	24	57.182	41.727	72.972	586
	방안1	58.671	43.406	73.982	1,000	74.266	49.904	67.196	25	53.099	37.337	70.316	544
	방안2	59.180	44.652	75.451	1,009	72.728	53.347	73.352	25	52.303	38.118	72.879	536
	방안3	58.586	42.505	72.552	999	69.110	48.210	69.759	23	53.411	39.843	74.597	547
	방안4	58.586	44.585	76.102	999	69.111	49.440	71.537	23	53.411	39.619	74.179	547
전북	고용조사	57.718	60.429	104.698	804	77.814	64.797	83.273	18	70.835	61.136	86.307	593
	방안1	60.439	61.754	102.176	842	79.506	65.700	82.635	18	66.260	55.786	84.193	554
	방안2	60.762	62.032	102.091	846	77.720	65.525	84.309	18	65.770	56.407	85.763	550
	방안3	60.348	60.314	99.944	841	91.207	76.564	83.946	21	66.093	58.022	87.789	553
	방안4	60.348	63.161	104.661	841	91.206	76.454	83.826	21	66.093	57.965	87.702	553

〈표 1-13〉 가중치 비교(전남, 경북, 경남, 제주)

지역	방안	취업				실업				비경제활동			
		평균값	표준편차	변동계수	합계	평균값	표준편차	변동계수	합계	평균값	표준편차	변동계수	합계
전남	고용조사	41.807	28.042	67.075	894	56.012	29.560	52.774	15	48.344	30.710	63.523	514
	방안1	43.592	28.027	64.295	933	56.913	30.098	52.884	16	44.728	27.086	60.556	476
	방안2	43.854	28.717	65.483	938	55.619	31.685	56.968	15	44.234	28.729	64.948	470
	방안3	43.183	27.324	63.274	924	54.261	30.982	57.097	15	45.620	29.792	65.305	485
경북	방안4	43.183	28.719	66.505	924	54.262	30.757	56.682	15	45.620	29.667	65.030	485
	고용조사	55.522	48.619	87.568	1,321	77.715	52.313	67.314	27	63.149	47.222	74.778	812
	방안1	57.928	49.128	84.809	1,379	79.573	53.235	66.902	28	58.646	42.166	71.898	754
	방안2	58.309	50.544	86.683	1,388	78.098	56.157	71.907	27	57.980	43.216	74.535	746
경남	방안3	58.994	48.746	82.630	1,404	81.517	56.554	69.377	29	56.620	43.292	76.460	728
	방안4	58.994	53.009	89.856	1,404	81.516	60.454	74.162	29	56.620	43.644	77.082	728
	고용조사	71.802	54.250	75.555	1,430	96.434	56.252	58.332	31	76.910	50.955	66.253	1,064
	방안1	75.078	55.288	73.641	1,495	99.732	57.540	57.695	32	72.115	46.112	63.942	997
제주	방안2	75.291	57.644	76.562	1,499	97.953	62.583	63.891	32	71.851	48.228	67.122	994
	방안3	77.476	55.517	71.658	1,543	103.768	60.654	58.452	34	68.569	46.149	67.303	948
	방안4	77.476	60.893	78.596	1,543	103.767	68.868	66.367	34	68.569	47.040	68.603	948
	고용조사	98.679	45.605	46.215	289	103.228	46.535	45.080	5	101.150	42.654	42.169	136
제주	방안1	102.299	46.213	45.175	299	105.232	47.234	44.885	6	93.191	38.444	41.253	125
	방안2	101.714	50.806	49.950	298	105.028	54.136	51.544	6	94.473	43.590	46.140	127
	방안3	98.958	44.922	45.395	289	79.453	40.306	50.729	4	101.479	44.806	44.153	136
	방안4	98.958	53.971	54.540	289	79.455	43.698	54.997	4	101.480	52.939	52.167	136

<표 1-10>~<표 1-13>은 16개 지역별 가중치 비교 결과이다. 전국에서와 마찬가지로 지역별 가중치 비교 결과를 취업자와 실업자 그리고 비경활인구 각각에 대하여 살펴보면, 4개의 벤치마킹 변수를 이용하여 보정한 방안1과 방안3의 변동계수가 6개의 벤치마킹 변수를 이용하여 보정한 방안2와 방안4에 비하여 비교적 작게 나타남을 알 수 있다. 또한 방안1의 경우 취업자, 실업자, 비경활자수에서의 변동계수가 보정 전과 거의 유사하거나 더 작게 나타나는 지역이 많음을 알 수 있다. 이를 통하여 고용조사의 보정 전 가중치에 비하여 벤치마킹 변수를 이용한 보정 후 가중치가 안정되어 있음을 확인할 수 있다. 이에 반해 벤치마킹 변수의 수가 늘어날 경우(방안2, 방안4)나 벤치마킹 변수들의 2차원 결합분포를 이용하는 경우(방안3)에는 변동계수가 조금씩 증가함을 알 수 있다. 이와 같은 결과는 4개의 벤치마킹 변수에 추가적으로 더한 2개의 벤치마킹 변수의 분포가 인구추계가 아닌 준거표본에 준하는 것이기 때문이다. 즉, 벤치마킹 변수의 주변분포를 모집단이 아닌 준거표본을 이용하여 구하는 경우 추정량의 정도를 떨어뜨릴 위험이 있음을 인지해야 한다. 또 다른 이유로는 벤치마킹 변수의 수가 많아져 고려해야 하는 분포의 수가 늘어나거나, 1차원 분포가 아닌 2차원 분포를 이용하여 고려해야 하는 정보의 수가 더 많아졌기 때문인 것으로 볼 수 있다. 그러나 일반적으로 벤치마킹 변수를 이용하여 가중치 보정을 하는 경우에 가중치의 변동이 커져 추정의 효율성이 떨어질 우려가 있음을 고려할 때, 각 보정 방안에서의 보정 가중치의 변동이 그리 크게 증가하지 않는다는 것은 매우 고무적인 일이라 할 수 있다.

<표 1-14>는 16개 지역별 경활상태의 보정 결과를 경활조사에서의 경활상태와 비교한 것으로, 보정 방안에 따른 보정 결과의 편차를 살펴볼 수 있다. 본 연구에서의 궁극적인 목적은 벤치마킹 변수를 이용하여 고용조사의 경활상태를 보정하고, 그 값을 경활조사에서의 경활상태와 유사하도록 하여 두 통계의 일관성을 확보하는 것이다. 따라서 보정 방안에 따른 보정 결과의 편차를 살펴보고, 어떤 방안을 이용하는 경우에 경활조사에서의 지역별 경활상태 편차가 가장 작게 나타나는지 확인하도록 한다.

<표 1-14> 경활조사와의 차이(계속)

지역(표본수)	방안	취업자	실업자	비경활자
서울 (30,466)	고용조사	117498.5	-1764.73	-115738
	방안1	-72510.4	-9752.56	82262.59
	방안2	-54615.8	-12902.3	67517.77
	방안3	4.81	-0.47	0.29
	방안4	0	0	0



〈표 1-14〉 경찰조사와의 차이(계속)

지역(표본수)	방안	취업자	실업자	비경찰자
부산 (17,787)	고용조사	59840.65	9214.64	-69057.5
	방안1	-9456.87	6813.1	2643.87
	방안2	-5014.96	6702.82	-1687.73
	방안3	0.34	-0.47	-1.35
	방안4	0	0	0
대구 (9,126)	고용조사	31023.02	5156.87	-36181.6
	방안1	-18421.9	3278.44	15143.5
	방안2	-13581	3083.98	10497.07
	방안3	0.53	-0.21	-0.19
	방안4	0	0	0
인천 (10,964)	고용조사	45970.32	-1473.54	-44499.2
	방안1	-6699.85	-3796.46	10496.25
	방안2	-6708.98	-3228.99	9937.9
	방안3	0.63	-0.31	-0.08
	방안4	0	0	0
광주 (5,365)	고용조사	13646.24	704.78	-14352.3
	방안1	-15381.1	-607.96	15989.16
	방안2	-4705.53	-812.97	5518.56
	방안3	0.35	-0.22	-0.15
	방안4	0	0	0
대전 (5,543)	고용조사	17325.94	2463.94	-19791.1
	방안1	-12970.1	1557.87	11412.14
	방안2	-7906.52	2008.52	5897.93
	방안3	0.73	0.03	-0.37
	방안4	0	0	0
울산 (5,381)	고용조사	13276.04	-642.81	-12635
	방안1	-8945.36	-1743.5	10689.15
	방안2	-5431.8	-1252.94	6685.03
	방안3	0.1	0.35	0.28
	방안4	0	0	0

〈표 1-14〉 경찰조사와의 차이(계속)

지역(표본수)	방안	취업자	실업자	비경찰자
경기 (62,188)	고용조사	308063.9	8944.64	-317014
	방안1	84506.16	1634.43	-86140.9
	방안2	77044.12	1506.08	-78550.5
	방안3	2.42	-0.34	2.07
	방안4	0	0	0
강원 (27,425)	고용조사	44346.46	3765.38	-48114.4
	방안1	13665.94	3467.31	-17132.9
	방안2	9958	3462.86	-13420.5
	방안3	-0.12	0.26	-0.78
	방안4	0	0	0
충북 (20,270)	고용조사	26491.41	-906.13	-25588.2
	방안1	-5645.99	-1653.98	7299.95
	방안2	-10521.9	-1757.31	12279.19
	방안3	-0.62	0.31	0.08
	방안4	0	0	0
충남 (27,626)	고용조사	39613.68	-992.22	-38622.2
	방안1	-1449.06	-1742.46	3191.67
	방안2	-10122.1	-1222.55	11344.79
	방안3	-0.95	0.44	-0.11
	방안4	0	0	0
전북 (22,523)	고용조사	36636.23	3040	-39680
	방안1	-1257.16	2655.85	-1398.95
	방안2	-5762.08	3061.36	2700.48
	방안3	-0.99	-0.34	-0.05
	방안4	0	0	0
전남 (32,303)	고용조사	29437.83	-482.95	-28954.5
	방안1	-8756.98	-731.5	9488.56
	방안2	-14361.7	-374.4	14736.19
	방안3	-2.47	0.43	-0.43
	방안4	0	0	0



〈표 1-14〉 경활조사와의 차이

지역(표본수)	방안	취업자	실업자	비경활자
경북 (37,010)	고용조사	82624.94	1338.04	-83963.2
	방안1	25364.28	684.02	-26048.3
	방안2	16286.45	1203.23	-17489.7
	방안3	-1.55	-0.49	0.38
	방안4	0	0	0
경남 (34,063)	고용조사	112970.9	2368.65	-115341
	방안1	47732.41	1303.5	-49036.2
	방안2	43504.73	1877.96	-45382.9
	방안3	-1.19	-0.18	-0.14
	방안4	0	0	0
제주 (4,322)	고용조사	816.28	-1259.94	442.7
	방안1	-9773.53	-1366.17	11140.02
	방안2	-8060.65	-1355.36	9416.33
	방안3	-0.55	0.15	0.27
	방안4	0	0	0

결과를 살펴보면, 방안4를 이용하여 보정한 후의 경활상태가 경활조사에서의 추정 결과와 정확하게 일치함을 알 수 있다. 이는 자료를 지역별로 분리하여 보정한 후에 최종적으로 경활상태에 대하여 비조정함으로써 두 통계의 경활상태를 각 지역별로 정확하게 일치시키는 방법을 이용했기 때문이다. 벤치마킹 변수들의 2차원 분포를 이용한 방안3의 경우에는 지역별 경활상태라는 2차원 분포에서의 편차가 매우 작게 나타남을 확인할 수 있으며, 방안1과 방안2의 경우에는 각 벤치마킹 변수들의 1차원 주변 분포만을 이용하였으므로 지역별 경활상태에 대한 편차를 줄이는 데 한계가 있음을 확인할 수 있다.

이상 살펴본 것과 같이 각 지역에서 두 통계의 경활상태에 일관성을 유지하기 위한 보정 방법 중 방안3과 방안4의 결과가 가장 효율적임을 확인할 수 있다. 이 외에 성별, 연령, 교육상태, 혼인상태별 가중치 비교 결과를 살펴본 결과(부록), 성별과 연령에서는 방안3의 보정 결과와 경활조사에서의 추정결과 사이에 편차가 거의 나타나지 않았으며, 교육상태와 혼인상태에서의 편차는 비교적 큰 것으로 나타났다. 이는 지역에서의 결과와 마찬가지로 성과 연령에 대한 2차원 결합분포를 이용하였기 때문에 성과 연령에 대한 경활상태의 편차가 거의 없어졌기 때문이다. 이에 반해 교육상태와 혼인상태는 방안3에서의 벤치마킹 변수로 고려되지 않았기 때문에 비교적 큰 편차가 발생한 것임을 알 수 있다. 방안2와 방안4의 경우에는 전반적으로 큰 편차가 발생하고 있으나, 교육상태와

혼인상태에서의 편차는 방안1과 방안3에 비해 다소 작게 나타나는 것을 확인할 수 있다. 이는 방안1과 방안3에는 교육상태와 혼인상태가 벤치마킹 변수로 이용되지 않았기 때문에 나타난 현상이다.

종합적으로 살펴보면, 방안3의 지역, 성, 연령별 경황상태의 변동계수가 보정 전과 비교하여 유사한 것으로 나타났으며, 경황조사 결과와의 편차도 거의 없는 것으로 나타났다. 이에 반해 방안3에 벤치마킹 변수로 이용하지 않은 교육상태와 혼인상태에 대해서는 비교적 큰 편차를 가지고 있음을 확인할 수 있었다.

<표 1-15>는 각 벤치마킹 변수에 대한 경황상태의 편차를 정리한 결과이다.

<표 1-15> 각 벤치마킹변수와 경황상태와의 2차원 분포 편차

		지역	성	연령	교육상태	혼인상태
경황상태	방안1	◎	◎	◎	◎	◎
	방안2	◎	○	○	○	○
	방안3	○	○	○	○	○
	방안4	-	○	○	○	○

◎>○>○>○

분석 결과들을 통하여 알 수 있듯이 교육상태 및 혼인상태를 포함한 방안3을 이용할 경우, 6개의 벤치마킹 변수를 이용한 4가지 보정 방안 중 편차가 가장 작은 보정 방안을 얻을 수 있을 것이다. 이 경우 총 15개의 2차원 분포를 이용하게 되며, 본 연구에서 고려하지 않은 다른 벤치마킹 변수가 더 추가되면, 고려해야 하는 2차원 분포의 수가 $n C_2$ (n : 벤치마킹 변수의 수)개로 크게 증가하게 된다.

3. 취업자 특성 및 분석 결과

본 절에서는 시도/산업별, 시도/직업별 취업자 부분을 추가 보정하는 작업을 진행하도록 한다. 이 때 이용되는 벤치마킹 변수는 경황조사를 준거표본으로 하며, 취업자들의 산업 및 직업 분류는 대분류를 이용하도록 한다.²⁾ 각각의 대분류는 다음과 같다.

2) 경황조사와 고용조사에서는 산업 및 직업의 중분류까지 공표하고 있으나, 준거표본인 경황조사 내 중분류에서의 분포에 빈 셀이 많아 보정에 어려움이 있다. 따라서 두 통계의 취업자에 대한 산업 및 직업 분류는 대분류를 이용하여 보정하도록 한다.



〈표 1-16〉 산업대분류

분류 기호	산업대분류
A	농업, 임업 및 어업
B	광업
C	제조업
D	전기, 가스, 증기 및 수도사업
E	하수·폐기물 처리, 원료재생 및 환경복원업
F	건설업
G	도매 및 소매업
H	운수업
I	숙박 및 음식점업
J	출판, 영상, 방송통신 및 정보서비스업
K	금융 및 보험업
L	부동산업 및 임대업
M	전문, 과학 및 기술 서비스업
N	사업시설관리 및 사업지원 서비스업
O	공공행정, 국방 및 사회보장 행정
P	교육 서비스업
Q	보건업 및 사회복지 서비스업
R	예술, 스포츠 및 여가관련 서비스업
S	협회 및 단체, 수리 및 기타 개인 서비스업
T	가구내 고용활동 및 달리 분류되지 않은 자가소비 생산활동
U	국제 및 외국기관

〈표 1-17〉 직업대분류

분류 번호	직업분류
1	관리자
2	전문가 및 관련 종사자
3	사무 종사자
4	서비스 종사자
5	판매 종사자
6	농림어업 숙련 종사자
7	기능원 및 관련 기능 종사자
8	장치, 기계조작 및 조립종사자
9	단순노무 종사자

가. 보정 방안

고용조사의 지역별(16개 시도별) 경황상태에 대한 가중치 보정 후 취업자들을 대상으로 산업 및 직업분류를 추가 벤치마킹 변수로 고려하여 가중치 보정 작업이 이루어진다. 앞 절에서 진행한 인구특성에 따른 보정 과정에서 혼인상태는 경황조사와 고용조사의 공표 형태가 겹치지 않아 벤치마킹 변수에서 제외하도록 하였다.

앞서 소개한 바와 같이 만 15세 이상 인구의 지역별, 성별, 연령별 모집단 분포는 인구추계를 통하여 알 수 있으나, 교육상태와 경황상태에 대한 모집단 분포는 알 수 없는 상황이다. 하물며 시도/산업별 또는 시도/직업별 취업자의 분포는 말할 것도 없다. 이에 경황조사 결과 분포를 준거표본으로 이용하여 가중치 보정 작업에 이용하도록 한다. 준거표본으로 이용하고자 하는 경황조사는 표본수가 고용조사에 비해 작아 표집오차는 크지만 고용조사에 비해 비표집오차가 작다. 따라서 준거표본으로 경황조사를 이용함으로써 비표집오차를 보정하는 작업은 바람직하다고 할 수 있다.

취업자를 대상으로 하는 추가적인 가중치 보정 작업은 다음과 같은 단계로 이루어진다.

- 1단계) 지역, 성, 연령, 교육상태 및 경황조사의 경황상태의 2차원 결합분포를 이용한 2차원 기준 반복비례가중법
- 2단계) 1단계의 가중치 보정 이후 “취업자” vs “취업자 외”로 자료 분리
- 3단계) 분리된 취업자 자료를 이용하여, 지역, 성, 연령, 교육상태, 산업, 직업별 2차원 결합분포를 이용한 2차원 기준 반복비례가중법
- 4단계) 3단계의 가중치 보정 이후 2단계에서의 취업자 외 자료와 통합
- 5단계) 통합된 자료를 이용하여 보정 결과 확인(각 범주별 오차 확인)

4.2절에서는 벤치마킹 변수에 혼인상태가 포함되어 있었으므로, 해당 변수를 제외하고 1단계 작업부터 보정 작업을 진행하였다. 3단계에서 지역, 산업, 직업만 고려하지 않고 성, 연령, 교육상태를 함께 고려하는 이유는 기존의 보정 작업을 통해 보정된 결과를 유지하게 위해서이다. 만약 시도/산업 및 시도/직업별 취업자 분포를 맞추기 위하여 이 3개 벤치마킹 변수만을 고려하여 가중치 보정 작업을 진행할 경우에는 이전의 성, 연령, 교육상태에 대한 분포에 큰 오차가 발생할 수 있다. 따라서 취업자 자료를 따로 분리하여 추가적인 가중치 보정 작업을 진행하는 경우에도 이전에 고려했던 벤치마킹 변수를 지속적으로 이용하도록 한다.

1단계에서는 5개 벤치마킹 변수(지역, 성, 연령, 교육상태, 경황상태)의 2차원 결합분포(${}_5C_2 = 10$ 개)를 이용하여 반복비례가중법을 적용하게 되며, 3단계에서는 6개 벤치마킹 변수(지역, 성, 연령, 교육상태, 산업대분류, 직업대분류)의 2차원 결합분포(${}_6C_2 = 15$



개)를 이용하게 된다. 이 때 연령별 산업대분류와 산업대분류별 직업대분류에 대한 2차원 결합분포에는 다수의 빈 셀이 포함되어 있어 그대로 이용하는데 무리가 있다. 허명희 외(2005)는 2차원반복비례가중법의 적용 시 빈 셀에 대한 처리 문제를 해결하기 위한 방안으로 빈 셀을 포함한 일부 변수에 대해 기존의 방법을 적용하거나, 빈 셀은 모집단수도 매우 작은 경우가 대부분이므로 해당되는 셀의 손실을 감내하는 것도 현실적인 방법이라고 제안하였다. 본 연구에서는 전자의 방법을 선택하여 적용하였다.

다. 분석결과

<표 1-18>과 <표 1-19>는 취업자를 대상으로 추가 보정한 시도/산업별 및 시도/직업별 취업자수이다.

결과를 살펴보면, 2차원 반복비례가중법을 이용하여 추가 보정한 후의 시도/산업 및 시도/직업별 분포가 경찰조사에서의 추정결과와 거의 정확하게 일치함을 알 수 있다. 이는 총 13개(연령별 산업, 산업별 직업 분포 제외)의 2차원 결합분포 내 각 벤치마킹 변수들의 정보가 반복적으로 적용되어 그만큼 편차를 줄였기 때문이다. 이상 살펴본 것과 같이 시도 단위에서 두 통계의 경찰상태 분포에 일관성을 유지하기 위한 보정 방법을 적용하고, 취업자의 특성에 대해 추가 보정하는 경우 2차원 반복비례가중법 이용이 오차를 줄이는데 효과가 있음을 확인할 수 있다. 이 외에 성별, 연령, 교육상태별 가중치 비교 결과(부록), 고용조사의 보정 후 결과와 경찰조사에서의 추정결과 사이에 편차가 거의 나타나지 않음을 확인할 수 있다. 즉, 2차원 반복비례가중법을 2번 연속 이용하여 편차가 매우 작은 보정 결과를 얻을 수 있었다.

본 연구에서는 두 번의 반복비례가중법에서 각각 10개와 13개의 2차원 분포를 이용하였으며, 만약 고려하지 않은 다른 벤치마킹 변수가 더 추가되면 고려해야하는 2차원 분포의 수는 ${}_n C_2(n)$: 벤치마킹 변수의 수)개로 증가하게 된다. 또한 벤치마킹 변수가 많을수록 가중치의 변동성이 커지는 경우가 많으므로 얼마나 많은 벤치마킹 변수를 활용하여 보정 작업을 수행할 것인지 먼저 결정해야 안정된 가중치를 얻을 수 있게 된다. 만약 두 통계의 모든 범주들에 대하여 보정 작업을 하게 된다면, 가중치의 변동성이 크게 확대되고 보정 작업에 들어가는 시간이 크게 늘어나 정확성과 효율성 그리고 시의성도 크게 악화될 가능성이 있다. 따라서 두 유사 통계에 대한 보정 작업을 수행할 경우에는 어떤 통계를 준거표본으로 이용할 것인지 결정하고, 어느 범주까지를 보정 범위 안에 넣을 것인지 사전에 결정해야할 필요가 있다.

〈표 1-18〉 두 조사에서의 시도/산업별 취업자수 비교(경활조사-고용조사)

	서울	부산	대구	인천	광주	대전	울산	경기	강원	충북	충남	전북	전남	경북	경남	제주
농림어업	-0.00896	0.05568	-0.03885	-0.15987	-0.19945	0.09279	6.77387	-1.73615	-0.19199	-0.26819	-1.13008	-0.86814	-0.94121	-1.19495	-0.26429	-0.45292
광업	0.00009	0.00009	0.00018	-0.00005	0.0001	-0.00008	-	0.00187	0.00281	0.00094	0.00158	0.00053	0.00387	0.00062	0.00027	0.00091
제조업	0.26903	0.09987	0.09438	0.16681	0.05012	0.02586	-2.35397	0.78963	0.03441	0.10378	0.28924	0.14843	0.14274	0.32106	0.19141	0.02155
전기&수도	0.00417	0.00043	0.00053	0.00043	0.00312	0.00091	-0.057	0.00408	0.00195	0.00164	0.00782	0.00368	0.00699	0.00771	0.00074	0.00067
하수&환경	0.00016	0.00042	0.00024	0.00209	0.00113	-0.00002	-0.00592	0.00781	0.00164	0.00151	0.00947	0.00544	0.00327	0.00587	0.00195	0.00111
건설	0.07242	0.01467	0.01963	0.02425	0.01945	0.00609	-0.49737	0.17645	0.02636	0.02881	0.08644	0.10149	0.12909	0.11353	0.02865	0.04564
도소매	-0.06186	-0.0324	-0.00869	-0.00964	0.02031	-0.02173	-0.72054	0.1896	0.00905	0.02867	0.14001	0.1394	0.1666	0.14868	0.01112	0.06563
운수	0.01858	0.00825	0.00706	0.01864	0.01202	-0.0013	-0.29234	0.13971	0.01136	0.0144	0.06987	0.04902	0.05082	0.06796	0.01164	0.02685
숙박	-0.0428	-0.01333	-0.01266	0.00473	0.01145	-0.01454	-0.54121	0.09561	0.00244	0.01633	0.08147	0.0729	0.09344	0.08879	0.00737	0.05204
&음식점	-0.09095	-0.01207	-0.00618	-0.00648	-0.00196	-0.00945	-0.06572	-0.01631	-0.00131	-0.00032	0.00808	0.00532	0.00974	0.01133	-0.00381	0.00674
출판&방송	-0.07945	-0.01884	-0.00808	-0.00833	-0.00396	-0.01398	-0.17533	-0.00314	-0.00322	-0.00014	0.02645	0.02016	0.01931	0.02216	-0.0137	0.01035
금융&보험	-0.0033	-0.00235	-0.00239	0.00197	0.00252	-0.00217	-0.11441	0.03772	0.00228	0.00204	0.02178	0.00535	0.01266	0.01486	0.00168	0.00712
과학기술	-0.05608	-0.00851	-0.00557	-0.00537	0.00085	-0.01445	-0.14275	0.01341	0.00045	0.00164	0.01082	0.01113	0.01297	0.02006	-0.00586	0.00951
시설관리	0.1033	0.03241	0.02036	0.0328	0.01623	0.01719	-0.26835	0.19247	0.01212	0.01701	0.07964	0.02382	0.04306	0.06024	0.03831	0.0267
행정	-0.00741	-0.00512	-0.00192	0.00008	0.00735	-0.00631	-0.18103	0.03186	0.00343	0.00699	0.0462	0.04328	0.05042	0.06044	0.00483	0.02619
교육	-0.08879	-0.03739	-0.02293	-0.0124	0.00217	-0.02509	-0.55578	0.02963	-0.0055	0.00548	0.06422	0.07678	0.06661	0.06496	-0.01667	0.0315
보건&복지	-0.04817	-0.01947	-0.00929	-0.004	0.00025	-0.01294	-0.26013	0.01527	-0.00119	0.00344	0.02792	0.03662	0.04057	0.04267	-0.00742	0.02411
예술	-0.04337	-0.01055	-0.01579	-0.0085	-0.00263	-0.00589	-0.12717	-0.03039	-0.0053	0.00004	0.01311	0.01109	0.01623	0.0106	-0.0058	0.01584
&스포츠																
협회&기타	0.03033	0.00449	0.00618	0.00996	0.01703	-0.00591	-0.38473	0.13053	0.01456	0.01563	0.06761	0.05553	0.08511	0.06921	0.01317	0.03033
서비스	0.0768	0.01288	0.00874	0.00493	0.00297	0.00557	-0.00265	0.05095	0.00324	0.00174	0.00899	0.0106	0.00507	0.0032	0.01082	0.00143
외국기관	-0.00277	-	-0.00143	-	-	-	-	-0.00437	-	-	-	-	-	0.00018	-	-



〈표 1-19〉 두 조사에서의 시도/직업별 취업자수 비교(경활조사-고용조사)

	서울	부산	대구	인천	광주	대전	울산	경기	강원	충북	충남	전북	전남	경북	경남	제주
관리자	-0.00821	-0.00039	0.00293	0.00201	-0.00198	0.00054	0.00015	0.0103	-0.00221	-0.00017	0.00157	-0.00238	-0.00171	-0.00399	-0.00054	0.00407
전문가	-0.03276	0.01663	-0.00335	-0.00877	-0.00315	-0.00441	0.00202	0.02979	0.00038	0.00019	0.00477	-0.00244	-0.00268	0.00045	0.00515	-0.00182
사무	-0.01613	0.01165	0.00045	0.01594	0.00027	-0.00155	0.0116	0.02434	-0.01328	0.00065	-0.01404	-0.00744	-0.00177	-0.01335	0.01195	-0.00929
서비스	0.00077	0.00164	-0.00035	0.00629	-0.00289	0.0002	0.00217	0.01428	-0.00515	-0.00031	-0.00598	-0.0019	-0.00228	0.00056	-0.00149	-0.00556
판매	0.00247	0.00538	0.0026	0.00514	-0.00168	0.00161	0.00239	-0.00606	-0.00132	-0.00428	-0.00486	0.00043	-0.00474	0.00016	0.00241	0.00035
농림어업	-0.01478	-0.00365	-0.00362	-0.00028	-0.00416	-0.00291	-0.00424	-0.01028	-0.00558	0.00039	0.00714	-0.00145	0.01942	0.00291	0.01186	0.00924
기능원	0.01502	0.00192	-0.00267	0.00103	-0.00157	0.00018	0.00227	-0.01149	-0.0042	0.00028	-0.00001	0.00355	-0.00021	-0.00267	0.00072	-0.00215
장차기계	-0.02393	-0.0104	0.00034	0.00714	-0.00358	-0.00278	0.00528	-0.01384	-0.003	0.00901	0.01078	0.01534	0.00316	-0.00027	0.00656	0.00018
단순노무	0.11854	0.04637	0.02717	0.02355	-0.01993	0.02367	0.00583	0.0792	-0.04805	-0.02399	-0.06872	-0.05129	0.00818	-0.04461	-0.03219	-0.04372

제5절 결론 및 논의

본 연구에서는 통계청에서 생산하고 있는 대표적인 경제활동 관련 두 통계의 추정결과를 일관성 있게 제공하기 위한 보정 방안에 대하여 살펴보았다. 이 중 경활조사를 준거표본으로 이용하였으며, 고용조사에서의 경활상태를 경활조사에서의 추정결과와 유사하도록 보정하였다. 이 과정에서 표본의 크기가 절대적으로 큰 고용조사를 상대적으로 표본수가 작은 경활조사를 기준으로 보정하는 것에 대한 논의가 있었으나, 경활조사의 역사와 시계열성 유지 그리고 우리나라를 대표하는 경제활동 관련 통계로서의 역할 등을 고려할 때, 신생 통계인 고용통계의 준거표본으로 이용하는데 무리가 없다는 결론을 얻었다. 또한 고용조사의 공표 시기는 동일 시점의 경활조사 결과가 이미 공표된 시기이므로, 이미 공표된 경활조사 결과(시도 단위)에 준하는 고용조사 결과를 발표하는 것이 이용자의 혼란을 최소화할 수 있는 방법이라고 할 수 있으며, 고용조사에 비해 비표집오차가 상대적으로 작은 경활조사 결과를 이용하여 비표집오차를 보정하는 작업은 바람직하다고 할 수 있다. 이에 지역, 성, 연령, 경활상태, 교육상태, 혼인상태 등 6개의 벤치마킹 변수를 이용하여 고용조사의 가중치를 보정하였다.

고용조사의 가중치 보정 작업을 위하여 4가지 보정 방안이 고려되었으며, 각 보정 방안을 수행한 결과, 전국 단위에서는 어떤 보정 방안을 사용하여도 상당히 신뢰할 수 있는 수준의 추정결과를 얻을 수 있음을 확인하였다. 그 중 방안3을 이용한 보정 가중치의 정도는 보정 전 가중치에 비해 유사하거나 향상되는 것으로 나타났다. 또한 지역별 경활상태에 대한 2차원 분포에서의 편차를 확인한 결과, 방안4에서는 편차가 발생하지 않았으며, 방안3에서는 매우 작은 편차만이 발생함을 알 수 있었다. 이에 반해 방안1과 방안2의 경우에는 상대적으로 큰 편차를 가지고 있음을 확인하였다.

앞에서 논의된 가중치 보정 방안의 검토 결과를 살펴보면 다음과 같다.

1. 각 벤치마킹 변수의 주변분포만을 이용하는 경우(방안1, 방안2), 4개의 벤치마킹 변수를 이용한 방안1이 6개의 벤치마킹 변수를 이용한 방안2에 비하여 가중치의 변동을 상당 부분 줄일 수 있다는 측면에서 효율적이나, 2차원 분포에서의 편차를 고려할 때는 방안2의 편차가 방안1에 비하여 더 작게 나타남을 알 수 있다.

2. 벤치마킹 변수의 2차원 결합분포를 이용하는 경우(방안3), 고려 대상이 되는 벤치마킹 변수(지역, 성, 연령)에 대해서는 경활상태와의 2차원 분포에서의 편차가 다른 방안에 비하여 매우 작게 나타남을 알 수 있다. 그러나 2차원 반복비례가중법은 고려해야 하는 벤치마킹 변수의 수가 늘어날 때, 고려 대상이 되는 2차원 결합분포의 수도 급격하게 증가한다는 단점을 가지고 있다.



3. 지역별로 반복비례가중법을 수행하고 최종적으로 경활상태에 대한 비조정을 하는 경우(방안4), 지역에서의 편차는 발생하지 않으며 고려 대상이 되는 각 벤치마킹 변수와 경활상태와의 편차도 비교적 작은 것으로 나타났다. 그러나 방안3과 비교할 때 성, 연령의 경활상태에 대한 2차원 분포에서의 편차는 상대적으로 크게 나타나는 것을 확인할 수 있었다.

4. 보정 가중치 비교 결과를 종합적으로 살펴보면, 고용조사와 경활조사의 지역별 경활상태에 대한 일관성을 유지한다는 측면에서 방안3이 가장 효과적인 것으로 나타났다. 그러나 교육상태와 혼인상태에 대한 분포를 고려하지 않을 경우 해당 분포에 대한 편차가 크게 나타나는 것을 확인할 수 있었다.

따라서 고용조사에서의 경활상태가 경활조사에서의 경활상태와 일관성을 유지하기 위한 가중치 보정에 교육상태와 혼인상태를 벤치마킹 변수로 추가한 방안3을 사용하는 것이 가장 효과적인 가중치 보정 방법이라고 결론내릴 수 있다.

이어 취업자 표본을 분리하여 시도/산업별, 시도/직업별 분포를 맞추기 위한 추가 보정 작업을 수행하였다. 이 때 해당 보정 작업은 벤치마킹 변수들 중 혼인상태를 제외하고 진행되었으며, 인구 특성에서와 마찬가지로 2차원 반복비례가중법이 이용되었다. 보정 작업을 수행한 결과, 각 벤치마킹 변수들의 범주에서는 경활조사 결과와의 오차가 매우 작아 상당히 신뢰할 수 있는 수준의 추정결과를 얻을 수 있음을 확인하였다.

참고문헌

- 허명희, 윤영아, 김규성 (2005). 2차원 기준 반복비례가중법 연구, 통계연구, 10(1), 1-22.
- Deville, J.C.. and Sarndal, C.E. (1992). Calibration estimators in survey sampling, *Journal of the American Statistical Association*, 87, 376-382.
- Deville, J.C., Sarndal, C.E. and Sautory, O. (1993). Generalized raking procedures in survey sampling, *Journal of the American Statistical Association*, 88, 1013-1020.
- Kim, J.K. and Park, M. (2009). Calibration estimation in survey sampling, submitted.
- Kim, J.K. and Rao, J.N.K. (2009). Design-based inference from two independent complex survey samples, Submitted.
- Renssen, R.H. and Nieuwenbroek, N.J. (1997). Aligning estimates for common variables in two or more sample surveys, *Journal of the American Statistical Association*, 92, p368-374.
- Wu, C. and Sitter, R.R. (1999). A model-based approach to using complete auxiliary information from survey data, *Journal of the American Statistical Association*, 96. 185-193
- Sarndal, C.E. and Lundstrom, S. (2004). *Estimation in surveys with nonresponse*, Wiley.
- Zieschang, K.E. (1990). Sample weighting methods and estimation of totals in the consumer expenditure survey, *Journal of the American Statistical Association*, 85, 986-1001.



<부록 I>

1. 성별 가중치 비교

성별 (표본수)	방안	취업				실업				비경제활동			
		평균값	표준편차	변동계수	합계*	평균값	표준편차	변동계수	합계*	평균값	표준편차	변동계수	합계*
남성 (165,644)	보정 전	82.517	101.458	122.954	13,668	2.683	24.849	926.323	444	31.887	73.965	231.959	5,282
	방안1	85.219	104.403	122.512	14,116	2.776	25.731	926.884	460	29.091	67.331	231.446	4,819
	방안2	84.481	107.872	127.688	13,994	2.829	27.036	955.601	469	29.776	70.979	238.375	4,932
	방안3	83.300	101.867	122.290	13,798	2.909	26.964	926.802	482	30.878	72.605	235.137	5,115
여성 (186,718)	방안4	85.526	109.384	127.896	14,167	2.870	27.303	951.409	475	30.028	72.135	240.231	4,974
	보정 전	49.266	84.949	172.430	9,199	1.402	17.740	1265.170	262	58.224	89.878	154.366	10,871
	방안1	52.115	89.238	171.235	9,731	1.477	18.710	1266.770	276	55.300	85.183	154.038	10,325
	방안2	52.769	92.612	175.503	9,853	1.430	18.692	1307.200	267	54.692	86.495	158.148	10,212
*친 명	방안3	53.817	91.762	170.506	10,049	1.359	17.254	1269.850	254	53.715	84.047	156.468	10,030
	방안4	51.843	90.981	175.494	9,680	1.394	18.095	1298.130	260	54.469	86.989	159.703	10,170

2.1 연령별 기증치 비교(계속)

연령별 (표본수)	방안	취업				실업				비경제활동			
		평균값	표준편차	변동계수	합계*	평균값	표준편차	변동계수	합계*	평균값	표준편차	변동계수	합계*
15~19 (24,393)	보정 전	6.177	36.268	587.165	157	0.494	10.022	2030.030	12	126.396	96.982	76.729	3,083
	방안1	5.997	35.271	588.174	146	0.489	9.940	2034.010	12	108.251	83.497	77.133	2,641
	방안2	4.862	29.095	598.442	119	0.404	8.447	2089.110	10	109.609	90.540	82.603	2,674
	방안3	2.438	14.330	587.643	59	0.217	4.387	2021.910	5	112.220	84.265	75.089	2,737
20~24 (15,900)	보정 전	4.411	26.007	589.530	108	0.384	7.949	2070.630	9	104.410	83.835	80.294	2,547
	방안1	69.653	112.568	161.613	1,107	5.539	35.757	645.586	88	92.616	117.612	126.990	1,473
	방안2	67.298	109.051	162.041	1,070	5.438	35.201	647.327	86	78.949	100.602	127.427	1,255
	방안3	70.708	117.846	166.666	1,124	5.519	36.855	667.742	88	75.407	98.813	131.040	1,199
25~29 (23,963)	방안3	55.525	90.936	163.775	883	4.449	28.584	642.428	71	91.660	117.315	127.989	1,457
	방안4	68.552	113.717	165.884	1,090	5.379	35.470	659.360	86	75.983	99.580	131.056	1,208
	보정 전	109.555	123.697	112.908	2,625	7.299	42.014	575.637	175	45.711	93.259	204.016	1,095
	방안1	112.463	127.299	113.192	2,695	7.610	43.868	576.486	182	41.660	85.105	204.287	998
*친 명	방안2	114.111	136.445	119.572	2,734	7.445	44.008	591.093	178	40.093	84.105	209.774	961
	방안3	105.663	119.360	112.963	2,532	7.578	43.401	572.701	182	48.408	100.096	206.778	1,160
	방안4	114.202	134.917	118.139	2,737	7.451	43.442	583.035	179	41.524	87.148	209.874	995

*친 명



2.2 연령별 기중치 비교(계속)

연령별 (표본수)	방안	취업				실업				비경제활동			
		평균값	표준편차	변동계수	합계*	평균값	표준편차	변동계수	합계*	평균값	표준편차	변동계수	합계*
30~34 (26,811)	보정 전	99.040	114.352	115.460	2,655	4.094	31.667	773.452	110	40.938	83.541	204.067	1,098
	방안1	97.200	112.516	115.757	2,606	4.090	31.656	774.081	110	36.093	73.788	204.440	968
	방안2	97.464	119.734	122.850	2,613	4.390	35.169	801.182	118	35.625	75.482	211.882	955
	방안3	97.075	112.583	115.976	2,603	4.326	33.455	773.296	116	36.077	74.747	207.187	967
35~39 (34,192)	방안4	97.847	120.196	122.841	2,623	4.421	35.465	802.246	118	36.390	78.278	215.110	976
	보정 전	93.741	102.050	108.863	3,205	2.631	23.781	903.730	90	31.513	72.788	230.978	1,077
	방안1	96.619	105.518	109.210	3,304	2.765	25.009	904.637	95	29.205	67.599	231.462	999
	방안2	96.347	110.511	114.702	3,295	2.958	27.796	939.714	101	29.298	69.749	238.070	1,002
40~44 (33,435)	방안3	95.435	103.881	108.850	3,263	2.861	25.832	903.018	98	30.306	70.587	232.914	1,036
	방안4	97.272	111.011	114.125	3,326	2.975	27.881	937.334	102	30.026	72.050	239.959	1,027
	보정 전	95.342	100.300	105.201	3,188	1.868	19.467	1042.350	62	26.130	66.939	256.173	874
	방안1	98.061	103.479	105.525	3,279	1.960	20.464	1044.050	66	24.132	61.947	256.699	807
*친 명	방안2	97.304	107.203	110.173	3,253	1.993	21.479	1077.780	67	24.834	65.094	262.114	830
	방안3	97.034	102.269	105.395	3,244	2.112	21.992	1041.050	71	24.985	64.669	258.832	835
	방안4	98.556	108.796	110.390	3,295	2.016	21.803	1081.320	67	25.610	68.125	266.010	856

*친 명

2.3 연령별 기증치 비교(계속)

연령별 (표본수)	방안	취업				실업				비경제활동			
		평균값	표준편차	변동계수	합계*	평균값	표준편차	변동계수	합계*	평균값	표준편차	변동계수	합계*
45~49 (36,767)	보정 전	88.426	97.105	109.815	3,251	1.613	18.357	1137.770	59	25.523	66.111	259.023	938
	방안1	90.331	99.564	110.222	3,321	1.685	19.205	1139.700	61	23.394	60.747	259.664	860
	방안2	89.334	101.444	113.556	3,285	1.615	19.134	1184.960	59	24.459	65.020	265.833	899
	방안3	91.525	100.152	109.426	3,365	1.819	20.498	1126.720	67	22.064	57.334	259.853	811
50~54 (32,581)	방안4	90.381	102.454	113.357	3,323	1.623	19.073	1175.190	60	25.164	67.555	268.463	925
	보정 전	79.652	95.941	120.450	2,595	1.445	18.041	1248.410	47	29.505	71.736	243.135	961
	방안1	87.643	105.927	120.862	2,855	1.620	20.264	1250.680	53	29.142	71.022	243.707	949
	방안2	86.696	107.157	123.601	2,825	1.509	19.226	1274.380	49	30.165	75.356	249.812	983
55~59 (25,510)	방안3	90.738	110.838	122.152	2,956	1.777	22.220	1250.130	58	25.855	63.792	246.734	842
	방안4	87.200	109.429	125.491	2,841	1.527	19.593	1283.180	50	31.054	78.980	254.333	1,012
	보정 전	63.606	87.096	136.932	1,623	1.267	16.610	1310.610	32	35.420	76.787	216.788	904
	방안1	69.286	95.221	137.433	1,767	1.406	18.458	1313.140	36	34.591	75.181	217.343	882
*친 평	방안2	68.843	95.887	139.283	1,756	1.322	17.843	1349.340	34	35.094	77.809	221.718	895
	방안3	73.119	101.100	138.268	1,865	1.471	19.263	1309.420	38	30.669	67.174	219.026	782
	방안4	68.629	96.132	140.075	1,751	1.326	17.876	1348.190	334	35.818	80.472	224.670	914

*친 평



2.4 연령별 기준치 비교

연령별 (표본수)	방안	취업				실업				비경제활동			
		평균값	표준편차	변동계수	합계*	평균값	표준편차	변동계수	합계*	평균값	표준편차	변동계수	합계*
60~64 (23,882)	보정 전	44.789	69.942	156.159	1,070	0.857	12.763	1488.980	20	40.758	77.293	189.636	973
	방안1	49.389	77.416	156.746	1,180	0.962	14.361	1492.170	23	40.226	76.456	190.069	961
	방안2	49.629	78.505	158.184	1,185	0.874	13.061	1494.120	21	40.100	77.918	194.309	958
	방안3	52.033	82.465	158.485	1,243	0.862	12.904	1496.370	21	37.708	72.680	192.747	901
≥65 (74,928)	방안4	48.789	77.615	159.083	1,165	0.872	13.023	1493.420	21	40.544	79.758	196.722	968
	보정 전	18.643	41.500	222.604	1,397	0.130	5.057	3882.060	10	49.072	76.316	155.521	3,677
	방안1	21.668	48.482	223.749	1,624	0.155	6.029	3891.830	12	51.040	79.695	156.144	3,824
	방안2	22.129	50.054	226.189	1,658	0.147	5.865	3991.020	11	50.564	81.071	160.333	3,789
*친 평	방안3	24.463	56.005	228.936	1,833	0.142	5.525	3888.510	11	48.235	77.151	159.948	3,614
	방안4	21.194	47.790	225.485	1,588	0.140	5.552	3972.430	11	49.605	80.345	161.971	3,717

*친 평

3.1 교육상태별 가중치 비교(계속)

교육상태 (표본수)	방안	취업			실업			비경제활동					
		평균값	표준편차	변동계수	합계*	평균값	표준편차	변동계수	합계*	평균값	표준편차	변동계수	합계*
≤초졸 (94,008)	보정 전	25.944	48.695	187.690	2,439	0.303	7.368	2431.900	28	36.557	66.194	181.071	3,437
	방안1	28.935	54.139	187.109	2,720	0.336	8.211	2443.090	32	37.387	67.983	181.835	3,515
	방안2	32.093	62.694	195.355	3,017	0.371	9.260	2497.980	35	39.198	72.881	185.933	3,685
	방안3	31.552	59.857	189.707	2,966	0.335	8.189	2445.490	31	34.798	64.577	185.580	3,271
중졸 (43,495)	방안4	31.455	62.115	197.475	2,957	0.373	9.330	2499.780	35	38.800	73.093	188.386	3,647
	보정 전	50.244	76.624	152.504	2,185	1.265	16.385	1295.720	55	45.967	82.064	178.529	1,999
	방안1	54.150	82.856	153.013	2,355	1.367	17.738	1297.640	59	43.888	78.843	179.645	1,909
	방안2	74.615	117.694	157.735	3,245	1.834	24.254	1322.790	80	58.944	107.107	181.711	2,564
고졸 (118,935)	방안3	56.381	87.189	154.642	2,452	1.410	18.319	1299.630	61	42.186	76.134	180.472	1,835
	방안4	74.445	116.682	156.735	3,238	1.844	24.083	1305.950	80	58.579	106.715	182.172	2,548
	보정 전	70.993	96.171	135.464	8,444	2.399	22.902	954.826	285	49.835	86.433	173.438	5,927
	방안1	73.730	100.148	135.830	8,769	2.502	23.938	956.581	298	45.443	79.252	174.399	5,405
*친 명	방안2	80.799	114.109	141.225	9,610	2.799	27.543	984.066	333	46.439	82.379	177.393	5,523
	방안3	73.481	100.265	136.451	8,739	2.526	24.274	960.861	300	45.557	79.465	174.431	5,418



3.2 교육상태별 기준치 비교

교육상태 (표본수)	방안	취업				실업				비경제활동			
		평균값	표준편차	변동계수	합계*	평균값	표준편차	변동계수	합계*	평균값	표준편차	변동계수	합계*
초대졸 (29,142)	방안4	81.152	115.362	142.156	9,652	2.814	27.613	981.268	335	46.352	83.579	180.312	5,513
	보정 전	99.680	111.211	111.568	2,905	4.399	32.369	735.791	128	40.052	83.811	209.257	1,167
	방안1	101.037	112.954	111.794	2,944	4.504	33.226	737.649	131	35.942	75.371	209.700	1,047
	방안2	90.778	106.316	117.116	2,645	4.092	30.695	750.030	119	30.703	66.337	216.061	895
대졸 (59,554)	방안3	96.683	108.290	112.005	2,818	4.364	32.296	739.987	127	38.074	80.826	212.284	1,110
	방안4	91.527	106.309	116.150	2,667	4.128	30.713	743.973	120	31.348	68.131	217.336	914
	보정 전	99.594	116.567	117.042	5,931	3.318	28.428	856.680	198	56.939	100.987	177.361	3,391
	방안1	101.817	119.314	117.185	6,064	3.420	29.335	857.729	204	51.271	91.173	177.826	3,0531
대학원졸 (7,228)	방안2	76.299	93.166	122.106	4,544	2.668	23.376	876.217	159	38.631	70.881	183.483	2,301
	방안3	98.805	116.178	117.583	5,884	3.407	29.312	860.276	203	55.093	99.072	179.826	3,281
	방안4	76.201	91.681	120.316	4,538	2.610	22.708	870.097	155	39.267	72.008	183.382	2,338
	보정 전	133.252	120.260	90.250	963	1.602	19.605	1223.890	12	32.114	83.472	259.923	232
*친 명	방안1	137.557	124.205	90.294	994	1.661	20.311	1222.990	12	29.764	77.271	259.608	215
	방안2	108.638	102.957	94.771	785	1.384	17.182	1241.400	10	24.500	65.706	268.184	177
	방안3	136.564	123.334	90.312	987	1.699	20.730	1220.430	12	31.720	83.484	263.192	229
	방안4	109.944	104.337	94.900	795	1.373	16.786	1222.650	10	25.450	68.876	270.630	184

*친 명

4. 혼인상태별 가중치 비교

혼인상태 (표본수)	방안	취업				실업				비경제활동			
		평균값	표준편차	변동계수	합계*	평균값	표준편차	변동계수	합계*	평균값	표준편차	변동계수	합계*
미혼 (72,710)	보정 전	68.414	109.095	159.463	4,974	5.236	35.243	673.114	381	77.118	104.388	135.362	5,607
	방안1	68.739	109.818	159.761	4,998	5.340	36.024	674.616	388	66.704	90.661	135.916	4,850
	방안2	77.693	128.044	164.807	5,649	5.693	39.458	693.133	414	68.516	96.648	141.059	4,982
	방안3	63.313	102.232	161.472	4,603	5.169	35.127	679.538	376	72.875	100.170	137.456	5,299
유배우 (229,541)	방안4	77.410	127.264	164.403	5,628	5.683	39.163	689.145	413	67.667	95.805	141.583	4,920
	보정 전	70.121	92.978	132.596	16,096	1.161	15.829	1363.290	267	35.399	75.106	212.168	8,126
	방안1	73.604	97.249	132.124	16,8956	1.235	16.849	1364.520	283	34.057	72.325	212.361	7,818
	방안2	71.347	95.662	134.080	16,3776	1.159	16.208	1398.400	266	34.167	74.123	216.940	7,843
기타 (50,111)	방안3	74.707	98.512	131.864	17,1486	1.279	17.465	1365.080	294	32.830	70.493	214.720	7,536
	방안4	71.547	96.453	134.811	16,423	1.165	16.293	1398.780	267	34.611	76.076	219.804	7,945
	보정 전	35.863	70.691	197.116	1,797	1.176	15.733	1337.580	59	48.305	76.315	157.986	2,421
	방안1	38.984	76.273	195.652	1,954	1.276	17.125	1342.380	64	49.424	78.206	158.234	2,477
*친 명	방안2	36.332	72.711	200.131	1,821	1.111	15.337	1380.700	56	46.291	74.725	161.424	2,320
	방안3	41.806	81.590	195.162	2,095	1.319	17.664	1339.060	66	46.091	74.561	161.770	2,310
	방안4	35.829	71.032	198.253	1,795	1.099	14.936	1359.070	55	45.491	73.501	161.572	2,280

*친 명



<부록 II> *

1. 전국 경활상태 비교(경활조사-고용조사)

경활상태	차이
취업자	1.319E-06
실업자	2E-09
비경활	8.74E-07

2. 시도별 경활상태 비교(경활조사-고용조사)

시도	취업자	실업자	비경활
서울	0.04099	0	0
부산	0.06915	0	0
대구	0.0235	0	0
인천	0.05205	0	0
광주	-0.03867	0	0
대전	0.01455	0	0
울산	0.02746	0	0
경기	0.11623	0	0
강원	-0.08242	0	0
충북	-0.01823	0	0
충남	-0.06933	0	0
전북	-0.04758	0	0
전남	0.01737	0	0
경북	-0.06079	0	0
경남	0.00442	0	0
제주	-0.0487	0	0

* 2차원 반복비례가중법을 이용한 최종 비교 결과이다. 2차원 반복비례가중법을 이용하였으므로 2차원 분포 (eg. 성/경활상태 또는 연령/경활상태 등)에서는 경활조사와 고용조사 사이에 차이가 거의 발생하지 않음을 알 수 있다.

3. 성별 가중치 비교(경찰조사-고용조사)

성별	취업자	실업자	비경찰
남성	-1.40574	0	0
여성	1.40574	0	0

4. 연령별 가중치 비교(경찰조사-고용조사)

연령	취업자	실업자	비경찰
15-19	-7.28E-12	-2.18E-11	-3.1E-08
20-24	5E-09	8.73E-11	1.7E-08
25-29	-4.4E-08	1E-09	9E-09
30-34	-5.8E-08	1.46E-10	2.2E-08
35-39	-4.6E-08	-8.73E-11	1.2E-08
40-44	5.1E-08	7.28E-11	1.6E-08
45-49	7.3E-08	-2.91E-11	1E-08
50-54	-4E-09	-1.38E-10	1.5E-08
55-59	-4.3E-08	-1.38E-10	6E-09
60-64	-3.6E-08	-9.09E-11	1E-09
65-	3.3E-08	-5.82E-11	-1.18E-07

5. 교육상태별 가중치 비교(경찰조사-고용조사)

교육상태	취업자	실업자	비경찰
초졸이하	0.068232	0	0
중졸	0.059328	0	0
고졸	-0.085911	0	0
초대졸	-0.060552	0	0
대졸이상	0.018903	0	0