

## Small Area Estimation of Unemployment Rate for the Economically Active Population Survey<sup>1)</sup>

Young-Won Kim<sup>2)</sup> · Ran Jo<sup>3)</sup>

### Abstract

In the Korean Economically Active Population Survey(EAPS), the sample sizes for small areas are typically too small to provide reliable estimators because the EAPS has been designed to produce unemployment statistics for large areas such as Metropolitan Cities and Province. In this study, we consider the synthetic and composite estimators for the unemployment rate of small areas, and apply them to real data on Choongbook province which is from the Korean EAPS of December 2000. The mean square errors of these estimators were estimated by the Jackknife method, and the efficiencies of small area estimators were evaluated in terms of the relative standard errors and the relative root mean square errors. As a result, the composite estimator is much more efficient than other estimators and it turns out that the composite estimator can produce the reliable estimates of the unemployment rate of small areas under the current EAPS system.

**Keywords:** Composite estimator, Jackknife estimation, Korean economically active population survey, Small area estimation, Unemployment rate

### 1. 서 론

경제활동인구조사를 비롯한 대부분의 정부통계를 생산하기 위한 표본설계는 전국

---

1) This Research was supported by the Sookmyung Women's University Research Grants 2001

2) First Author : Professor, Department of Statistics, Sookmyung Women's University, Seoul, Korea

E-mail: ywkim@sookmyung.ac.kr

3) Graduate Student, Department of Statistics, Sookmyung Women's University, Seoul, Korea

단위나 광역시 또는 도 단위와 같은 대영역의 통계를 생산할 목적으로 설계되었기 때문에, 시군구 등과 같은 각 소지역에 대한 통계를 추정하기에는 표본 조사구수가 극히 적어지게 된다. 하지만 현실적인 비용을 고려할 때 전국의 시군구 등과 같은 소지역에 대한 신뢰성 있는 통계를 얻기 위해 새로운 표본조사를 실시하는 것은 불가능하다고 할 수 있다. 따라서 주어진 여건을 감안할 때 기존의 표본설계에서 조사된 자료를 가지고 일정 수준의 정확성을 갖춘 시군구 단위의 통계를 생산할 수 있는 추정기법에 대한 연구가 요구되고 있다. 이와 같이 표본설계 당시에 통계 산출 대상으로 감안하기 못한 좀더 작은 지역에 대한 신뢰성 있는 통계를 산출하기 위한 방법으로 소지역 추정법이 사용되고 있다. 소지역 추정법이란 표본크기가 작은 소지역(small area)이나 성별, 연령, 교육수준, 소득수준 등과 같은 변수의 특성으로 분류된 소영역(small domain)에 대한 통계를 생산하는데 이용되고 있는 추정 방법이다.

특히 지방자치제의 실시 등에 따라 다양한 행정부처에서 지역별 정부정책 수립을 위해 시군구 등 소지역 실업률에 대한 정확한 통계를 필요로 하고 있다. 그러나 현재 우리나라에서 실시되는 경제활동인구조사의 표본설계는 관심영역이 특광역시 및 도별 실업자 통계산출을 목적으로 하기 때문에 시군구와 같은 소지역에 대한 통계 산출은 표본설계에서 고려하고 있지 않다. 따라서 현재 사용되고 있는 표본설계를 이용하여 소지역 통계를 직접 생산할 경우 시군구 지역에 대해서는 표본 조사구의 수가 너무 작게 배정되어 있기 때문에 단순히 조사 자료만을 활용하는 경우 신뢰할만한 소지역 단위의 통계생산은 현실적으로 불가능하다.

본 논문에서는 2000년 12월 경제활동인구조사 자료를 이용하여 적절한 규모라고 판단되는 충청북도지역을 연구 대상으로 충청북도내의 10개의 시군을 실업률 추정대상 소지역으로 선정하였다. 실제 충청북도는 3개의 시와 8개의 군으로 구성되어 있는데 그중에 진천군은 표본조사구가 1개만이 배정되어 있어 연구대상에서 제외한 것이다. 본 연구에서는 소지역 추정기법을 적용하여 이들 10개 시군구의 실업률을 산출하고 그 효율성을 검증하는 것을 목적으로 한다.

현재 외국에서는 분석대상에 따라 다양한 형태의 소지역 추정기법들에 대한 심층적인 연구들이 진행되고 있다. 특히 미국, 캐나다, 영국 등의 경우 정부기관과 전문 학자들과의 공동연구를 통해 소지역 추정기법에 대한 연구가 활발히 추진되고 있다. Rao(2003)은 다양한 형태의 소지역 추정방법에 대한 최근까지의 연구결과들을 체계적으로 정리하고 있다.

우리나라의 경우 성나영, 김영원(2000)이 도소매업 사업체 조사를 바탕으로 소지역 추정기법의 도입 가능성을 검토한 이후 박종태, 이상은(2001)은 경제활동인구조사를 토대로 경기도 시군구의 실업자총계 추정문제를 제한적인 모의실험을 통해 다루었다. 특히 최근에 들어 우리나라 정부기관에서도 소지역 추정기법 도입의 필요성을 절감하고 관련 학자들과 연계하여 연구를 추진하고 있는데, 현재까지 우리나라에서 진행된 대표적인 연구로는 이계오, 류제복, 김영원(2001)이 노동부와 협조하여 수행한 소지역 실업자 총계 추정 가능성 검토 및 전산프로그램 개발을 위한 기초연구가 있었으며 정연수, 이계오, 이우일(2003) 및 Chung, Lee, Kim(2003)이 통계청의 도움을 받아 수행한 충청북도 시군구 실업자 총계 추정에 대한 연구가 있다. 정연수 등(2003) 및 Chung 등(2003)의 연구는 실제 통계청에서 경제활동인구조사를 기반으로 시군구 공식 실업통계 산출을 염두에 두고 추진된 본격적인 소지역추정기법의 개발을 목적으로 한 것이었다.

현재 우리나라에서 실업통계와 관련된 소지역 추정 연구에서 실업자 총계 추정에 대한 연구는 활발하나 아직까지 주요 관심대상이 될 수 있는 실업률에 대한 소지역 추정은 국내에서 연구되어지지 않고 있다. 우리나라 실업통계 소지역 추정 문제와 관련된 대표적인 연구인 Chung 등(2003)에서는 결과적으로 시군구에 대한 실업자 총계와 경제활동인구 총계를 별도의 소지역 추정방법을 적용하여 산출하고, 이들의 비(ratio)를 구하여 소지역 실업률을 산출할 수는 있지만 실업률 추정 자체를 목적으로 하는 경우 이는 효율적인 방법이라 볼 수 없다. 이런 접근방식은 우리나라의 관련 분야 연구자들이 실업통계와 관련된 소지역 통계기법이 활발히 연구되고 있는 캐나다의 실업률 소지역추정기법(Dew 등(1982), Rao(2003) 참조)에 많은 영향을 받은 것에 기인한 것으로 판단된다. 하지만 지역별 실업자 수가 아니라 실업률 자체가 주 관심대상인 경우 좀더 효율적이고 안정적인 소지역 실업률 산출을 위해서는 소지역별 실업률을 직접 활용하는 소지역 추정기법이 또 다른 대안이 될 수 있다. 특히 Ambler(2002)를 참고로 하면 영국 통계국(ONS)에서는 지역별 실업률을 직접 이용하기 위해 이항분포를 기반으로 한 소지역 추정기법을 연구 개발하고 있다는 것을 확인할 수 있다.

따라서 본 논문에서는 소지역별 실업자수를 추정하여 실업률을 간접적으로 산출하는 대신 직접 소지역추정기법으로 시군구 실업률을 산출하는 방법을 제시하고 그 효율성을 검증하고자 한다. 이를 위해 2000년 12월 기준 경제활동인구조사 및 상주인구 추계자료를 바탕으로 직접추정법, 합성추정법, 복합추정법 등 3가지 추정법을 적용하여 충청북도 내의 10개 시군 실업률을 추정한다. 아울러 추정방법들의 효율성을 비교하기 위해 잭나이프 방법을 이용한 평균제곱오차(MSE) 추정방법을 제시하고 이를 통해 추정량들의 상대표준오차(RSE: Relative Standard Error)와 상대평균제곱오차제곱근(RRMSE: Relative Root Mean Square Error)을 비교 분석한다.

## 2. 경제활동 인구조사 개요 및 실업률 산출방법

우선 독자의 이해를 돕기 위해 연구대상이 되는 경제활동인구조사의 개요 및 실업률 산출방법을 정리하고자 한다. 경제활동인구연보(2001)에 따르면 경제활동인구조사는 경제개발 5개년 계획의 수립과 그 수행을 위하여 1962년 8월부터 표본조사방법에 의하여 분기별로 실시하여 왔다. 그 동안 인구주택총조사 결과에 따라 일곱 차례의 표본개편이 있었으며, 1980년부터는 다양한 자료생산을 위하여 조사항목을 확대하였고, 1982년 7월부터는 고용구조의 변화 추이를 신속히 파악하기 위하여 조사주기를 월별로 조정하였다. 경제활동인구조사는 취업, 실업, 노동력 등과 같은 인구의 경제적 특성을 조사하여 노동공급, 고용구조, 노동시간 및 인력의 활용정도를 파악함으로써 정부의 고용정책 입안 및 평가에 필요한 자료를 제공하는데 목적을 둔 조사이다. 조사대상 기간은 매월 15일이 포함된 1주간(일요일~토요일)으로 하며, 실제조사는 조사대상 기간 다음 주에 실시한다. 조사 항목으로는 가구원 성명, 생년월일, 취업여부, 1주간 및 1개월간의 구직 여부, 취업가능성 여부, 취업의사 유무 등 30개 항목을 조사하고 조사 결과는 지역 고용정책 수립의 기초 자료로 활용하기 위하여 매월 6개 시도별 자료가 함께 공표되고 있으며 고용상황에 대한 심층 분석을 위하여 구직 단념자 통계, 구직기간별 실업자 통계 등 여러 가지 보조지표를 작성하여 공표하고 있다.

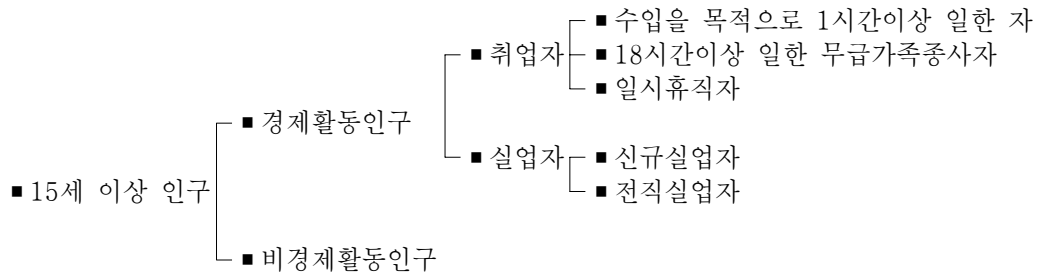
경제활동인구조사에 발표하는 실업률은 다음과 같은 방법으로 산출된다. 고용통계에서 내리는 실업률 정의는 실업자가 경제활동인구에서 차지하는 비율을 말한다. 즉,

$$\text{실업률(\%)} = \frac{\text{실업자}}{\text{경제활동인구}} \times 100$$

여기서 경제활동인구(Economically active population)란 만 15세 이상 인구 중 취업자와 실업자를 말한다. 취업자와 실업자에 대한 구분 기준은 아래와 같다. 아래의 분류기준은 ILO(국제노동기구) 및 세계 각국이 공통적으로 적용하고 있는 기준이며, 통계청도 이러한 국제기준에 따라 통계를 작성하고 있다.

- 취업자
  - 조사대상 주간 중 수입을 목적으로 1시간 이상 일한 자.
  - 자기에게 직접적으로는 이득이나 수입이 오지 않더라도 자기가구에서 경영하는 농장이나 사업체의 수입을 높이는 데 도운 가족종사자로서 주당 18시간이상 일한 자(무급 가족 종사자).
  - 직장 또는 사업체를 가지고 있으나 조사대상 주간 중 일시적인 병, 일기불순, 휴가 또는 연가, 노동쟁의 등의 이유로 일하지 못한 일시 휴직자.
- 실업자
  - 조사대상 주간 중 수입 있는 일에 전혀 종사하지 못한 자로서, 적극적으로 구직활동을 하고, 즉시 취업이 가능한 자.
  - 30일 이내에 새로운 직장에 들어갈 것이 확실한 취업 대기자는 구직활동 여부에 관계없이 실업자로 분류.

이런 실업자 분류기준을 간단히 요약하면 [그림 2.1]과 같다.



[그림 2-1] 고용통계에서의 분류기준

이런 방법으로 얻어진 경제활동 관련 통계는 대지역 단위의 통계를 생산할 목적으로 설계되어진 표본자료이므로 이러한 표본자료만을 가지고 지방자치행정의 기본 단위인 시군구와 같은 소지역 통계를 산출한다면 해당하는 소지역별 조사구 수의 불균형적 분포로 추정량의 변동이 커지게 되고 따라서 신뢰성 있는 소지역 통계를 산출할 수 없게 된다. 더구나 경제활동 관련 통계는 개인뿐만 아니라 사회 또는 정치적으로 주요 관심사가 되고 있기 때문에 어느 정도 신뢰도를 만족하는 시군구의 실업통계를 생산할 수 있는 방안에 대한 연구는 매우 중요하다.

### 3. 경제활동인구조사에서 소지역 실업률 추정방법

#### 3.1 직접추정량

소지역  $i$ 에 대한 실업률을 추정하기 위한 직접추정량은 소지역에 배정된 표본조사구에서 얻어진 자료만을 이용하여 추정된다. 경제활동인구조사 체계에서의 소지역  $i$ 의 실업률에 대한 직접추정량  $\hat{r}_i^d$ 은 다음과 같다.

$$\hat{r}_i^d = \frac{\sum_{s=1}^2 \sum_{k=1}^{n_i} {}_s\hat{Y}_{ik}}{\sum_{s=1}^2 \sum_{k=1}^{n_i} {}_s\hat{X}_{ik}} = \frac{\sum_{s=1}^2 \sum_{k=1}^{n_i} {}_sM_{is}Y_{ik}}{\sum_{s=1}^2 \sum_{k=1}^{n_i} {}_sM_{is}X_{ik}}, \quad i = 1, \dots, I \quad (3.1)$$

여기서  $s$ 는 성별(남, 여)를 나타내고,  $n_i$ 는 경제활동인구조사에서의 소지역  $i$ 에 할당된 표본조사구 수,  ${}_sY_{ik}$ 는 경제활동인구조사에서 소지역  $i$ 의  $k$ 번째 표본조사구에서 성별 15세 이상의 실업자 수,  ${}_sX_{ik}$ 는 경제활동인구조사에서 소지역  $i$ 의  $k$ 번째 표본조사구에서 성별 15세 이상의 경제활동인구를 나타내며, 승수  ${}_sM_i$ 는 행정자료를 이용해 산출된 경제활동인구조사에서 사용되는 가중값이다. 여기서  ${}_sM_i = {}_sW_i^* / {}_sW_i$ 으로  ${}_sW_i^*$ 는 소지역  $i$ 에 속한 성별 15세 이상의 상주추계인구를 나타내며  ${}_sW_i$ 는 경제활동인구조사에서 집계된 성별 15세 이상의 상주조사인구를 나타낸다.

한편, 소지역  $i$ 에 대한 직접 추정량  $\hat{r}_i^d$ 은 비추정량(ratio estimator)으로 이에 대한 분산은 다음과 같은 잭나이프 분산추정공식을 적용하여 추정한다.

$$\widehat{Var}_J(\hat{r}_i^d) = \frac{n_i - 1}{n_i} \sum_{k=1}^{n_i} (\hat{r}_i^d(k) - \hat{r}_i^d)^2 \quad (3.2)$$

여기서  $\hat{r}_i^d(k)$ 는 소지역  $i$ 에서 조사구  $k$ 를 빼고 구한 직접추정량을 나타낸다.

#### 3.2 합성추정량

전체 연구대상 지역을 유사한 지역으로 묶기 위해 우선 중복지역을 서로 특성이 유사한 두개의 부차영역(시지역과 군지역)으로 분류하여 각각 합성추정량을 산출한다. 한편 각 부차영역의 전체 조사대상자를 성별(남, 여)\*연령대별(15~34세, 35세 이상)로 구분한 4개의 범주로 구분하여 각각 실업률을 추정하고, 이를 다시 각 시군 단위 행정구역들에 대한 범주별 상주추정인구를 보조정보로 이용하여 합성추정량을 산출한다. 결과적으로 소지역  $i$ 의 실업률에 대한 합성추정량  $\hat{r}_i^s$ 은 식(3.3)과 같다.

$$\hat{r}_i^s = \sum_{j=1}^4 \frac{P_{ij}}{P_i} \hat{r}_j \quad , \quad i = 1, \dots, I \quad (3.3)$$

여기서  $\hat{r}_j$  ( $j = 1, 2, 3, 4$ )는 경찰인구조사에서 각 부차영역(시지역, 군지역)에서의 각 범주(성별(남, 여)×연령(15-34세, 36세 이상))별 실업률을 나타내고,  $P_i$ 은 소지역  $i$ 에서 15세 이상 추계상주인구를  $P_{ij}$ 는 소지역  $i$ 에서 범주  $j$ 에 속하는 추계상주인구를 나타낸다.

한편 합성추정량  $\hat{r}_i^s$ 의 분산은 다음과 같은 잭나이프 분산추정공식을 적용하여 추정한다.

$$\widehat{Var}_J(\hat{r}_i^s) = \frac{n_i - 1}{n_i} \sum_{k=1}^{n_i} (\hat{r}_i^s(k) - \hat{r}_i^s)^2 \quad (3.4)$$

여기서  $\hat{r}_i^s(k)$ 는 소지역  $i$ 에서 조사구  $k$ 를 빼고 구한 합성추정량을 나타낸다.

### 3.3 복합추정량

직접추정량의 불안정성과 합성추정량의 편향을 서로 보완할 수 있는 방법으로 두 추정량의 가중평균으로 표현되는 복합추정량은 다음과 같이 표현된다.

$$\hat{r}_i^c = w_{i(opt)} \hat{r}_i^d + (1 - w_{i(opt)}) \hat{r}_i^s \quad (3.5)$$

여기서  $\hat{r}_i^d$ 는 소지역  $i$ 의 직접추정량이고  $\hat{r}_i^s$ 은 소지역  $i$ 의 합성추정량이다.  $w_i$ 는 0과 1사이의 값으로 복합추정량의 MSE를 최소로 하는 다음과 같은 최적가중값을 사용한다.

$$w_{i(opt)} = \frac{MSE_J(\hat{r}_i^s)_J}{MSE_J(\hat{r}_i^s)_J + \widehat{Var}_J(\hat{r}_i^d)_J}, \quad i = 1, \dots, I \quad (3.6)$$

한편 복합추정량  $\hat{r}_i^c$ 의 분산은 다음과 같은 잭나이프 분산추정공식을 적용하여 추정한다.

$$\widehat{Var}_J(\hat{r}_i^c) = \frac{n_i - 1}{n_i} \sum_{k=1}^{n_i} [\hat{r}_i^c(k) - \hat{r}_i^c]^2 \quad (3.7)$$

여기서  $\hat{r}_i^c(k)$ 는 다음과 같이 소지역  $i$ 에서 조사구  $k$ 를 빼고 구한 복합추정량을 나타낸다.

$$\hat{r}_i^c(k) = w_{i(opt)} \hat{r}_i + (1 - w_{i(opt)}) \hat{r}_i^s(k)$$

### 3.3 잭나이프 방법을 이용한 평균제곱오차(MSE) 추정

합성추정량  $\hat{r}_i^s$  또는 복합추정량  $\hat{r}_i^c$ 의 경우 편향(bias)이 있는 추정량들이기 때문에 이들 추정량의 정확도를 나타내는 측도로 평균제곱오차(MSE)를 고려하는 것이 필요하다. 이들 추정량은 상당히 복잡한 형태의 추정량이기 때문에 이에 대한 MSE를 직접 추정하는 데는 상당한 어려움이 따른다. 따라서 경제활동인구조사 자료에 대해 대영역 내의 표본조사구들을 서로 독립이고 동일한 임의 분포를 갖는 확률표본으로

가정한 상태에서 정연수 등(2003)이 제시한 잭나이프 추정방법을 이용하여 MSE를 추정한다. 한편 이와 관련하여 Jiang 등(2002)은 소지역 추정법에 효과적으로 적용될 수 있는 근사적으로 불편인 잭나이프 MSE 추정방법을 제시하고 있다.

정연수 등(2003)이 제시한 방법에 따라 식(3.3)에 제시된 합성추정량  $\hat{r}_i^s$ 의 MSE를 구하는 절차를 정리하면 다음과 같다.

① 각 소지역에서 표본 조사구를 하나씩 제외하고, 다른 지역의 원 자료는 그대로 유지하여 실업자 총수  $Y$ 와 경제활동인구수  $X$ 에 대한 총  $n = \sum_{i=1}^J n_i$ 개의 반복표본을 생성한다.

② 소지역  $i$ 에서 하나의 조사구가 제외된 경우 다른 소지역의 승수는 그대로 유지하고, 해당 소지역내의 나머지  $n_i - 1$ 개의 표본조사구에 대한 승수는 다음과 같이 보정한다.

$${}_sM_i^{adj} = \frac{n_i}{n_i - 1} {}_sM_i$$

③ 나머지  $n - 1$ 개의 표본 조사구를 이용하여 잭나이프 합성추정값  $\hat{r}_i^s(k)$ 를 반복적으로 계산한다. 결과적으로 소지역  $i$ 에서는  $n_i$ 개의 서로 다른 합성추정값을 얻게 된다.

소지역  $i$ 의 실업률에 대한 합성추정량의 잭나이프 MSE 추정식은 다음과 같이 표현된다.

$$MSE_J(\hat{r}_i^s) = Var_J(\hat{r}_i^s) + [Bias_J(\hat{r}_i^s)]^2$$

여기서  $Var_J(\hat{r}_i^s)$ 와  $Bias_J(\hat{r}_i^s)$ 은 다음과 같이 구해진다.

$$Var_J(\hat{r}_i^s) = \frac{n_i - 1}{n_i} \sum_{k=1}^{n_i} (\hat{r}_i^s(k) - \hat{r}_i^s)^2$$

$$Bias_J(\hat{r}_i^s) = (n_i - 1) \left[ \frac{1}{n_i} \sum_{k=1}^{n_i} \hat{r}_i^s(k) - \hat{r}_i^s \right]$$

한편 복합추정량  $\hat{r}_i^c$ 의 MSE도 합성추정량의 경우에 적용한 것과 유사한 과정을 통해 잭나이프 방법으로 추정한다. 아울러 직접추정량의 경우도 실업률의 경우 편향이 발생하는 비(ration) 추정량이 사용되고 있기 때문에 제시된 잭나이프 방법을 적용하여 MSE를 추정하여 추정량들의 효율성을 비교한다.

#### 4. 효율성비교

충청북도 각 시·군에 대한 실업률 및 MSE 추정 결과는 [표 4.1]과 같다. 실업률에 대한 MSE 추정값들을 살펴보면, 모든 소지역에 대해 복합추정량의 경우 MSE가 가장 작고 안정적임을 알 수 있다. 특히 다른 지역에 비해 실업률 직접추정에 있어 많

은 차이를 보이는 소지역 3과 6의 경우 복합추정에서는 이런 차이가 상당부분 축소되고 있음을 볼 수 있다.

또한 각 소지역에 대한 3가지 추정량의 상대오차를 비교하기 위해 RSE(Relative Standard Error)와 RRMSE(Relative Root Mean Squared Error)를 계산하여 효율성을 평가한다. 소지역  $i$ 에 대한 실업률을  $\hat{r}_i$ 라고 할 때, RSE와 RRMSE는 식(3.8)과 식(3.9) 같이 계산된 것이다. 만약  $\hat{r}_i$ 가 불편추정량이면 RSE와 RRMSE는 같아진다.

[표 4.1] 충청지역의 시군에 대한 실업률 및 MSE 추정값

소지역 번호	직접추정량		합성추정량		복합추정량		표본 조사구
	$\hat{r}_i^d$	$MSE_J$	$\hat{r}_i^s$	$MSE_J$	$\hat{r}_i^c$	$MSE_J$	
1	0.033427	0.000079	0.034227	0.000041	0.033844	0.000022	22
2	0.034281	0.000181	0.031743	0.000027	0.032145	0.000022	11
3	0.009195	0.000114	0.030958	0.000020	0.026523	0.000017	4
4	0.034236	0.001036	0.035349	0.000022	0.035326	0.000019	2
5	0.023945	0.000573	0.039977	0.000025	0.039272	0.000022	3
6	0.054701	0.000687	0.040378	0.000031	0.041034	0.000028	3
7	0.028310	0.000402	0.038526	0.000027	0.037837	0.000023	5
8	0.032669	0.000318	0.041015	0.000053	0.039711	0.000040	6
9	0.031924	0.000273	0.042515	0.000030	0.041325	0.000025	5
10	0.032042	0.000817	0.038253	0.000021	0.038092	0.000018	2

$$RSE(\hat{r}_i) = \frac{\sqrt{\widehat{Var}(\hat{r}_i)}}{\hat{r}_i} \times 100 \quad (3.8)$$

$$RRMSE(\hat{r}_i) = \frac{\sqrt{\widehat{MSE}(\hat{r}_i)}}{\hat{r}_i} \times 100 \quad (3.9)$$

충청북도지역의 10개의 시·군에 대한 실업률 추정량들의 RSE와 RRMSE를 구해보면 [표 4.2]와 같다. 여기서도 직접추정량에 비해 합성추정량과 복합추정량의 경우 RSE와 RRMSE가 모두 대폭 감소하고 있다는 것을 볼 수 있으며, 직접추정량의 경우 RSE와 RRMSE가 각각 평균적으로 64.3%, 69.01%로 매우 크게 나와 실제 공식 통계로 활용하기에는 부적합하지만 복합추정량의 경우 각각 6.55%, 13.32%로 어느 정도 신뢰할 수 있는 추정결과를 산출하고 있다는 것을 알 수 있다.

한편 [표 4.2]의 결과를 동일한 자료를 이용해 실업자 총계 추정을 목적으로 하고 있는 정연수 등(2003)의 논문에 제시된 [표 5.2]와 비교해 보면, 정연수 등(2003)의 연구에서는 복합추정량의 RRMSE가 소지역별로 상당한 변동을 보이고 있는 데 비해 [표 4.2]에서는 RRMSE에 있어 소지역별로 변동이 별로 발생하고 있지 않다는 차이점



을 볼 수 있다. 두 연구에서 이런 차이가 발생한 이유는 본 연구의 경우 연구변수로 실업자 총수 대신 비(ratio) 형식의 실업률을 사용함에 따라 추정결과가 상대적으로 안정적으로 나타나기 때문인 것으로 추측된다. 이와 관련된 사항에 대해서는 좀더 심층적인 분석이 필요하다고 생각된다.

결과적으로 현재의 경제활동인구조사의 표본설계 하에서 실업률 추정을 목적으로 하는 경우 제시된 복합추정법을 활용하면 공식 통계로 활용하는 데 큰 무리가 없는 상당한 수준의 안정적이고 신뢰성 있는 소지역 실업률 통계를 생산할 수 있다는 점을 확인할 수 있다.

[표 4.2] 충청지역의 시군에 대한 실업률 추정량에 대한 RSE 및 RRMSE

소지역 번호	직접추정량		합성추정량		복합추정량	
	RSE	RRMSE	RSE	RRMSE	RSE	RRMSE
1	19.88	26.54	13.54	18.64	7.13	13.97
2	35.27	39.22	9.06	16.52	7.53	14.72
3	96.96	116.12	3.40	14.57	3.16	15.65
4	92.45	94.03	4.72	13.27	4.62	12.40
5	96.90	99.96	5.88	12.45	5.72	11.83
6	46.68	47.90	8.60	13.85	8.08	12.78
7	67.67	70.78	7.02	13.37	6.66	12.64
8	51.57	54.62	14.08	17.68	12.27	15.98
9	48.37	51.75	7.78	12.92	7.11	12.14
10	87.29	89.20	3.27	11.92	3.20	11.14
Average	64.30	69.01	7.74	14.52	6.55	13.32

## 5. 결론

본 연구에서는 2000년 12월 기준으로 경제활동인구조사 자료를 바탕으로 충청북도의 10개의 시·군 소지역 실업률에 대한 직접추정값, 합성추정값, 복합추정값을 산출하는 동시에 잭나이프 방법을 이용하여 평균제곱오차와 분산추정을 하고, RSE와 RRMSE를 기준으로 3가지 추정량의 효율성을 비교 분석하였다. 그 결과 직접 추정량의 경우 시군구에 배정된 조사구가 워낙 적기 때문에 신뢰할 수 있는 실업률을 추정하는 데 한계를 갖고 있지만 제시된 복합추정량을 사용하면 직접추정량에 비해 그 효율성을 대폭 향상시킬 수 있다는 것을 알 수 있었다. 따라서 이런 소지역추정기법을 통해 현재 경제활동인구조사 체계를 그대로 유지하더라도 공식 통계로 사용하는 데 있어 크게 문제가 되지 않을 정도의 상당한 수준의 신뢰성을 갖는 시군구 실업률 산출이 가능하다는 점을 확인할 수 있었다.

아울러 경제활동인구조사를 바탕으로 실업 관련 소지역 통계를 생산하는 데 있어서 추정대상이 시군구 실업자 총수인 경우 정연수 등(2003) 또는 Chung 등(2003)에 제시된 방법이 효과적인 방법일 수 있지만 추정대상이 시군구의 실업률인 경우 또 다른

대안으로 본 연구에서 제시된 방법이 충분히 검토될 필요가 있다고 판단된다.

한편 본 논문에서는 경제활동인구조사에서 시군구 실업률 산출을 위해 합성추정량과 복합추정량을 제시하고 이에 대한 효율성을 검토하고 있지만, 향후 보다 효율적인 시군구 실업률 산출을 위한 소지역추정기법 개발을 위해서는 실업률과 같은 이항 자료를 적합 시키는 데 매우 유용한 모형으로 알려져 있는 GLM(generalized linear model) 또는 HGLM(hierarchical GLM)을 기반으로 한 소지역 추정기법의 적용 가능성에 대해 보다 심층적인 연구가 필요하다고 판단된다.

### 참고문헌

1. 박종태, 이상은 (2001). 소지역 추정법에 관한 비교연구, *Journal of the Korea Data & Information Science Society*, 12, 47-55.
2. 성나영, 김영원 (2000). 소지역 통계 생산을 위한 추정방법, *Journal of the Korea Data & Information Science Society*, 11, 111-126.
3. 이계오, 류제복, 김영원(2001). <소지역 실업률 추정기법 및 전산프로그램개발>, 노동부 학술연구용역사업 보고서.
4. 정연수, 이계오, 이우일 (2003). 시군구 실업자 총계 추정을 위한 설계기반 간접추정법, <응용통계연구>, 16, 1-14.
5. 통계청(2002). <2001 경제활동인구연보>, 통계청.
6. Ambler, R., Caplan, D., Chambers, R., Kovacevic, M., and Wang, S.(2001). Combining unemployment benefits data and LFS data to estimate ILO unemployment for small area: An application of a modified Fay-Herriot method, *Proceedings of the International Statistical institute*, Seoul. 128-138.
7. Chung, Y-S., Lee, K. and Kim, B-C.(2003). Adjustment of Unemployment Estimates Based on Small Area Estimation in Korea, *Survey Methodology*, 29, 45-52.
8. Derw, D., Singh, M. P and Choudhry, G. H. (1982). Evaluation of Small Area Estimation Techniques for the Canadian Labour Force Survey, *Survey Methodology*, 8, 17-47.
9. Jiang, J., Lahiri, P., Wan, S., (2002). A Unified Jackknife Theory for Empirical Best Prediction with M-Estimation, *The Annals of Statistics*, 30, 1782-1810.
10. Rao, J. N. K. (2003). *Small Area Estimation*, Wiley, New York.

[ 2003년 10월 접수, 2003년 12월 채택 ]