

소지역 추정법을 이용한 효율적인 지역 실업률 추정

김수택¹

¹경상대학교 정보통계학과

(2011년 11월 접수, 2011년 11월 채택)

요약

지방자치제가 정착되면서 시·군·구와 같은 소지역 단위의 실업률에 대한 통계는 실업대책이나 고용지원 사업을 위한 예산 편성과 같은 지역 노동시장정책의 수립은 물론 정책 집행 후 그에 대한 사후평가를 하는데 있어서도 중요한 기준을 제공해주기 때문에 지자체 기관장들의 관심의 대상이 되고 있는 실정이다. 그러나 경제적, 시간적으로 많은 제약이 따르는 기초자치단체에서 지속적으로 실업률 통계를 생산하는 데에는 자료의 신뢰성, 연속성 및 시의성 면에서 많은 문제를 노정하고 있다. 본 연구에서는 소지역 실업률 추정치에 대한 신뢰성(변동계수 25% 이하)을 확보하면서 통계생산에 소요되는 조사비용을 최소화할 수 있는 효율적인 추정법 및 최소 표본조사구 수를 모의실험을 통하여 제시하고자 한다.

주요어: 소지역 추정법, 실업률, 모의실험, 변동계수.

1. 서론

지방자치제가 정착되면서 지역정책수립의 기초 자료인 지역통계의 중요성이 그 어느 때보다 증대되고 있다. 따라서 지방자치단체에서도 지역통계를 생산하거나 개발을 계획하고 있지만 지방자치단체의 열악한 통계조직, 통계자료의 활용부진, 예산부족 등으로 지역통계의 생산은 미미한 실정이다. 그러나 인구집단의 특성에 관한 통계는 국가나 광역시 및 도 차원뿐만 아니라 시·군·구 수준에서도 관심이 되고 있다. 특히, 시·군·구와 같은 소지역 단위에서 실업률에 대한 통계는 실업대책이나 고용지원 사업을 위한 예산 편성과 같은 지역 노동시장정책의 수립은 물론 정책 집행 후 그에 대한 사후평가를 하는데 있어서도 중요한 기준을 제공해주기 때문에 지자체 기관장들의 관심의 대상이 되고 있는 실정이다. 그렇지만 통계조사만으로 늘어나는 지역통계의 수요를 충족시키는 데에는 경제적, 시간적으로 많은 제약이 따르는데 이에 대한 대안으로 소지역 추정법을 활용할 수 있다.

미국, 영국, 일본 등과 같은 통계 선진국에서는 국가통계 생산에 소지역 추정법을 적극적으로 활용하고 있으며 학문적으로도 Falorsi 등 (1994), Farrell (2000), Rao (2003), Fabrizi 등 (2007), Ugate 등 (2007)과 같이 꾸준히 연구가 지속되고 있다. 국내에서의 소지역 추정에 관한 연구는 지난 2000년을 전후해서 이계오 (2000), 박종태 (1998), 박종태 (2001), 박종태와 이상은 (2001), 김영원과 성나영 (2000), 전경배 (2001) 등에 의하여 처음 소개되기 시작하였다. 그 후의 연구로는 김달호와 김남희 (2002), 양영춘 등 (2003), 정연수 등 (2003), 김재두 등 (2005) 등이 있다.

소지역 추정에 사용되는 추정량으로는 설계기반 추출법(design-based sampling)의 표본추출확률을 이용한 직접추정량(direct estimator)과 목표 소지역의 특성과 유사한 인근 소지역에 관한 통계정보(보

본 연구는 2009년도 경상대학교 학술진흥사업 연구비에 의하여 수행되었음(RPP-2009-000).

¹(660-701) 경남 진주시 진주대로 501, 경상대학교 정보통계학과, 교수. E-mail: kstioa@gnu.ac.kr

표 1.1. 2009년 경제활동인구조사의 표본 조사구 수

시군	거제	거창	고성	김해	남해	마산	밀양	사천	산청	양산
센서스 조사구수	1029	345	288	2169	294	2166	606	581	201	1166
경찰조사 표본조사구 수	6	2	1	11	2	12	6	4	2	6
시군	의령	진주	진해	창원	창녕	통영	하동	함안	함양	합천
센서스 조사구수	157	1694	712	2398	352	688	263	323	222	274
경찰조사 표본조사구 수	2	11	4	17	3	3	2	2	2	2

조정보)를 이용하는 간접추정량(indirect estimator)이 있다. 간접추정량은 설계기반(design-based)과 모형기반(model-based) 간접추정량으로 구분할 수 있다. 설계기반 간접추정량으로는 합성추정량과 복합추정량이 있으며 대표적인 모형기반 추정량으로는 경험적 최량 선형 비편향 추정량(Empirical Best Linear Unbiased Predictor; EBLUP), 경험적 베이저안 추정량(Empirical Bayesian Estimator; EB) 및 계층적 베이저안 추정량(Hierarchical Bayesian Estimator; HB) 등을 들 수 있다.

본 연구에서는 통계청에서 매월 조사하여 발표하고 있는 경제활동인구조사와 매 5년마다 시행하고 있는 인구 및 주택센서스 자료를 이용하여 우리나라 지방자치단체가 처한 현실적인 상황에서 실업률 등과 같은 소지역 노동 통계를 생산해 낼 수 있는 방안을 모색하고자 한다. 통계청에서 2009년 경제활동인구조사를 수행하기 위하여 선정한 표본 조사구들 중에서 경남 지역 시·군에 배정된 조사구 수를 나타내고 있는 표 1.1을 보면 센서스 조사구 수에 비하여 시·군에 배정된 조사구 수가 매우 작음을 볼 수 있다. 이와 같이 크기가 작은 표본 자료로부터 실업률을 직접 추정하게 되면 추정 오차가 너무 커서 통계의 신뢰성을 보장할 수 없다. 따라서 경상남도에서는 2007년 9월 통계청의 지도하에 10개 각 시에서 평균 50개의 조사구를 선정하고 경제활동인구조사를 실시하여 소지역(시)의 실업률을 추정하고자 하였다. 통계청의 경제활동인구조사에서 각 소지역에 배정된 표본조사구 수보다 훨씬 많은 50여개에 달하는 조사구를 조사하여 직접추정법으로 실업률을 추정하였지만 통계청에서 요구하는 목표정도에 도달하지 못하는 결과를 초래하였다. 따라서 본 연구에서는 시·군과 같은 소지역 실업률 추정에서 추정치의 신뢰성(변동계수 기준)을 확보하는 동시에 조사에 소요되는 비용을 최소화하여 지속적으로 소지역 실업률 통계를 생산할 수 있도록 하는 방안을 모색하고자 한다. 이를 위하여 일정 크기 이상의 표본을 필요로 하는 직접추정법 대신 보조정보를 이용하는 간접추정법을 활용하는 방안을 검토하고 모의실험을 통하여 최소 표본 조사구 수를 결정하고자 한다.

2. 소지역 추정법

2.1. 직접추정법

경제활동인구조사(경찰조사) 자료로부터 소지역 i 의 총 실업자 수를 추정할 수 있는 직접추정량 \hat{Y}_i 는

$$\hat{Y}_i^{dir} = \sum_{s=1}^2 \sum_{h=1}^{n_i} \frac{s\hat{X}_i}{sX_i} sY_{ih},$$

여기서 s 는 성을 나타내는 지표, n_i 는 경찰조사에서 소지역 i 의 조사구 수, sY_{ih} 는 경찰조사의 소지역 i 내 h 번째 표본조사구의 성별 실업자 수를 각각 나타낸다. 그리고 $s\hat{X}_i$ 는 소지역 i 의 거주인구에 대한 추정치이며, sX_i 는 경찰 표본조사로부터 도출한 표본조사 거주인구를 각각 나타낸다. 위와 같이 주어지는 직접추정량 \hat{Y}_i 는 근사적 비편향 추정량이다. 우리나라 통계청에서는 \hat{Y}_i 의 분산을

$$\text{Var}(\hat{Y}_i) = \sum_{s=1}^2 sM_i^2 \left(\zeta_i \sum_{h=1}^{n_i} sU_{ih}^2 \right) + 2 {}_1M_{i2}M_i \left(\zeta_i \sum_{h=1}^{n_i} {}_1U_{ih2}U_{ih} \right)$$

으로 추정한다. 여기서,

$$\begin{aligned} {}_sU_{ih} &= d({}_sY_{ih}) - {}_s\rho_i \cdot d({}_sX_{ih}), \\ d({}_sY_{ih}) &= {}_sY_{ih} - {}_sY_{i,h+1}, \quad d({}_sX_{ih}) = {}_sX_{ih} - {}_sX_{i,h+1}, \\ {}_s\rho_i &= \frac{{}_sY_i}{{}_sX_i}, \quad \xi_i = \frac{(1-f_i)n_i}{2(n_i-1)}, \quad f_i = \frac{n_i}{10N_i} \end{aligned}$$

이고, N_i 는 소지역 i 의 추출틀에 속하는 조사구의 수를 나타낸다. 직접추정량은 근사적 비편향 추정량이지만 소지역에 할당된 표본의 크기가 작은 경우에는 추정량의 분산이 커져서 신뢰성이 떨어지게 된다.

2.2. 합성추정법

소지역 i 에 대한 합성추정량 \hat{Y}_i^S 는 직접조사 자료 및 소지역 i 와 특성이 유사한 지역들의 정보를 추정에 이용하는 간접적인 설계기반 추정량이다.

I 개의 시·군 단위 소지역이 있는 대지역을 특성이 동질적인 L 개의 하위 지역그룹으로 분할하고 각 그룹들을 다시 J 개의 성별-연령대별 범주로 분할하면 $I = \sum_{k=1}^L I_k$ 이고, k 번째 그룹은 I_k 개의 동질적인 소지역(시·군)들로 구성된다. 따라서 각 소지역은 L 개의 하위 지역그룹 중 어느 하나에 속하게 되고, 소지역이 속하는 하위 지역그룹에 대한 보조정보를 추정에 이용한다. 따라서 소지역 i 에 대한 합성추정량 \hat{Y}_i^S 는 직접추정량에 비하여 그 추정오차가 현저하게 줄어들 수 있지만 해당 소지역이 속해 있는 하위 지역그룹 내의 소지역들에 대한 정보가 해당 소지역의 것과 동질적이지 않으면 편향이 커질 수 있다.

N_i 를 표집틀에서 소지역 i 에 속하는 총 조사구 수, n_i 를 경찰조사에서 소지역 i 에 할당된 표본조사구 수, ${}_jP_{i,2005}^C$ 를 2005년 센서스에서 추계된 (i, j) 범주의 상주인구, ${}_jP_{i,2005}^R$ 를 (i, j) 범주의 2005년 주민등록인구, ${}_jP_{i,month}^R$ 를 (i, j) 범주에 대한 경찰조사 월의 주민등록인구, ${}_j\hat{X}_i$ 를 (i, j) 범주의 상주인구에 대한 직접추정치, ${}_jY_{ih}$ 를 (i, j) 범주의 h 번째 표본 조사구에서 조사된 실업자 수라 하면 I_k 개의 소지역을 포함하는 하위 지역그룹 내 소지역 i 의 총 실업자 수 Y_i 에 대한 합성추정량 \hat{Y}_i^S 는

$$\hat{Y}_i^S = \sum_{j=1}^J \frac{{}_j\hat{P}_i}{{}_j\hat{X}_i} {}_j\hat{Y}_{dir}, \quad i = 1, \dots, I_k$$

와 같은 식으로 주어진다. 여기서

$${}_j\hat{P}_i = \frac{{}_jP_{i,2005}^C}{{}_jP_{i,2005}^R} {}_jP_{i,month}^R, \quad {}_j\hat{X}_i = \sum_{i=1}^{I_k} {}_j\hat{X}_i, \quad {}_j\hat{Y}_{dir} = \sum_{i=1}^{I_1} \sum_{h=1}^{n_i} \frac{{}_j\hat{X}_i}{{}_jX_i} {}_jY_{ih}$$

이며, ${}_j\hat{P}_i$ 는 행정보고 자료를 이용하여 추계된 소지역 i 의 j 번째 성별-연령대별 범주에 대한 상주 추계인구, ${}_j\hat{X}_i$ 는 경제활동인구조사 자료로부터 추계한 소지역 i 의 j 번째 범주(셀)에 대한 상주추정인구, ${}_j\hat{Y}_{dir}$ 는 경제활동인구조사 자료로부터 추계한 j 번째 범주의 총 실업자 수에 대한 직접추정치를 나타낸다.

합성추정량 \hat{Y}_i^S 의 정밀도에 대한 측도로는 다음과 같은 평균제곱오차

$$MSE(\hat{Y}_i^S) = \text{Var}(\hat{Y}_i^S) + [\text{Bias}(\hat{Y}_i^S)]^2$$

를 고려한다. Ghosh와 Rao (1994)는 \hat{Y}_i 가 Y_i 의 직접추정량일 때 $\text{Cov}(\hat{Y}_i, \hat{Y}_i^S) = 0$ 이라는 가정 하에서

$$\text{mse}(\hat{Y}_i^S) \approx (\hat{Y}_i^S - \hat{Y}_i)^2 - \widehat{\text{Var}}(\hat{Y}_i)$$

표 3.1. 모의실험에서 모집단으로 이용된 소지역별 표본조사구 수

시군구분(1)	집락번호(2)	센서스 조사구수(3)	2007년 표본조사구수(4)	2007년 조정된 표본조사구수(5)	2006 경찰조사 표본조사구수(6)
01(거제)	1	1029	52	49	6
02(김해)	2	2169	52	49	10
03(마산)	3	2166	57	47	13
04(밀양)	2	606	51	39	4
05(사천)	4	581	54	40	4
06(양산)	2	1166	55	50	6
07(진주)	4	1694	50	49	10
08(진해)	3	712	39	37	4
09(창원)	3	2679	47	47	16
10(통영)	1	688	51	47	4

주: (1)은 경남의 10개 시를, (2)는 그룹화된 집락번호를, (3)은 인구 및 주택센서스에서 각 지역에 배정된 조사구 수를, (4)는 2007년 경남에서 실시한 경제활동인구조사에서 각 지역에 배정된 표본조사구 수를, (5)는 (4)의 각 지역별 표본조사구 중에서 자료가 불확실한 일부 표본을 제외한 표본조사구 수를 그리고 (6)은 2006년 통계청 경제 활동인구조사에서 각 지역에 배정된 표본조사구 수를 각각 나타낸다.

가 \hat{Y}_i^S 의 편향에 대한 근사적 비편향 추정량임을 보였다. 그러나 이 추정량은 표본조사구의 수가 충분히 크지 않을 때는 불안정한 경향을 보인다. 보다 안정적이고 지역의 특성을 충분히 반영할 수 있는 추정법으로 Jackknife 추정법을 고려한다.

2.3. 복합추정법

비편향 추정량인 직접추정량 \hat{Y}_i 을 보완하는 방안으로 편향이 아주 작거나 거의 없지만 분산이 큰 직접 추정량 \hat{Y}_i^{dir} 와 분산은 작지만 편향이 있는 합성추정량 \hat{Y}_i^S 의 가중평균을 복합추정량이라 하는데, i 지역의 복합추정량 \hat{Y}_i^C 는 다음과 같다.

$$\hat{Y}_i^C = \omega_i \hat{Y}_i + (1 - \omega_i) \hat{Y}_i^S, \quad i = 1, \dots, I_k,$$

여기서 ω_i 는 0과 1 사이의 값을 갖는 가중치이다. 추정량의 평균제곱오차를 최소화 하는 최적 가중치 $\omega_{i(opt)}$ 는

$$\hat{\omega}_{i(opt)} = \frac{\text{mse}_J(\hat{Y}_i^S)}{\text{mse}_J(\hat{Y}_i^S) + \widehat{\text{Var}}(\hat{Y}_i)}$$

의 식으로 추정한다 (Rao, 2003). 복합추정량의 역시 보다 안정적인 값을 제공하는 Jackknife 추정법으로 추정한다 (이계오, 2002).

3. 모의실험

3.1. 모의실험 방법

모의실험에서 표본조사구를 추출하기 위하여 경남에서 통계청의 협조를 얻어 2007년 9월에 실시한 경제활동인구조사 자료를 이용하였다. 표 3.1을 보면 2007년 경남의 자체적인 경찰조사에서 조사한 표본조사구 수는 2006년도 통계청 경찰조사의 표본조사구 수에 비하여 3배에서 많게는 13배까지 크게 늘어

표 3.2. 제 1집락: 거제, 통영

거제시 표본수	실업률(%)			변동계수(CV: %)		
	직접	합성	복합	직접	합성	복합
4	2.03	2.87	2.13	78.33	48.64	23.51
5	1.90	2.71	1.99	73.41	48.27	21.68
6	1.76	2.57	1.84	70.01	47.28	19.85
7	1.68	2.48	1.77	65.88	45.61	18.32
8	1.64	2.45	1.73	63.70	44.67	17.41
9	1.63	2.48	1.71	60.66	43.53	16.06
10	1.61	2.41	1.68	57.45	42.56	15.39
통영시 표본수	실업률(%)			변동계수(CV: %)		
	직접	합성	복합	직접	합성	복합
4	2.66	2.37	2.18	67.76	37.07	21.12
5	2.53	2.26	2.08	63.66	38.44	20.33
6	2.51	2.25	2.08	58.46	36.35	17.82
7	2.45	2.20	2.02	54.35	35.71	16.25
8	2.42	2.23	2.02	50.58	34.62	14.83
9	2.42	2.24	2.04	47.61	32.82	13.31
10	2.42	2.23	2.02	45.45	33.17	13.17

났지만 직접추정법으로 구한 소지역 실업률 추정치가 CV 25%이하이어야 한다는 목표 정도를 만족하지 못한 것으로 알려져 있다. 따라서 표본조사구 수를 크게 늘리지 않으면서 직접추정법에 비하여 더 효율적인 추정법으로 실업률을 추정하는 방안을 고려할 필요가 있다. 이를 위하여 추정방법에 따른 소지역 실업률에 대한 추정치가 통계청에서 요구하는 목표 정도, 즉, CV 25%이하를 만족하기 위한 최소 표본조사구 수를 도출하기 위하여 모의실험을 실시하였다.

모의실험에서 간접추정법에 의하여 실업률을 추정하기 위해서는 먼저 10개의 시들을 그 특성이 유사한 시들을 묶어서 그룹화 해야 한다. 각 소지역들에 대한 그룹화는 실업률과 관련 있는 지리적 및 인구학적 특성 변수들을 조사하여 군집분석 등의 방법으로 이루어져야 하는데 본 연구에서 지리적 위치와 직접추정법으로 추정한 각 시의 실업률의 유사성을 고려하여 10개의 시를 4개의 집락으로 분류하였다. 다음으로 각 소지역(시)의 표본조사구 내 인구를 연령(30세 미만, 30세 이상)과 성별(남자, 여자)에 따른 4개의 범주로 구분한다. 단, 4개 범주 중의 어느 하나라도 인구수가 0인 표본조사구는 표본 추출 대상 조사구에서 제외하였다.

이제, 추정에 사용될 표본조사구에 관한 자료는 조정된 표본조사구에서 $n = 4$ 개의 조사구를 랜덤하게 뽑고, 간접추정에 사용될 해당 지역과 동일한 집락으로 묶인 시에 대한 표본조사구 자료는 2006년 경제활동인구조사에서 배정된 표본조사구 수만큼 조정된 표본조사구에서 랜덤하게 뽑은 다음, 각 소지역의 실업률을 직접추정법, 합성추정법, 복합추정법 등의 세 가지 방법으로 추정하는 과정을 5000번 반복해 5000개 실업률의 평균을 구한다. 다음으로 n 을 4, 5, 6, 7, 8, 9, 10으로 바꾸어 위와 같은 과정을 반복한다.

3.2. 모의실험 결과

표 3.2~표 3.5는 각 집락 내 소지역별 실업률에 대한 추정치와 추정치에 대한 변동계수를 나타내고 있다. 이 표들을 보면 표본조사구 수가 증가함에 따라 세 가지 추정법에 의한 추정치들이 감소하는 경향을 보이고 있는데 이는 표본수가 커지면서 편향이 감소하기 때문이다. 그러나 표본조사구 수가 8개 이

표 3.3. 제 2집락: 김해, 밀양, 양산

김해시 표본수	실업률(%)			변동계수(CV: %)		
	직접	합성	복합	직접	합성	복합
4	2.65	4.08	2.82	70.51	39.57	15.81
5	2.53	4.02	2.71	66.85	39.31	14.67
6	2.41	3.87	2.58	62.93	37.70	12.79
7	2.37	3.88	2.55	58.13	36.60	11.85
8	2.38	3.90	2.54	54.07	35.18	10.94
9	2.38	3.84	2.52	51.10	34.23	10.14
10	2.40	3.88	2.55	47.60	32.88	9.48
밀양시 표본수	실업률(%)			변동계수(CV: %)		
	직접	합성	복합	직접	합성	복합
4	4.05	3.42	3.02	79.11	16.24	9.56
5	3.69	3.32	2.87	76.17	15.62	8.04
6	3.57	3.34	2.87	72.17	15.05	7.10
7	3.40	3.31	2.83	68.32	14.17	5.95
8	3.38	3.27	2.84	65.26	13.76	5.33
9	3.35	3.24	2.81	62.99	13.66	4.96
10	3.20	3.27	2.79	60.31	12.81	4.33
양산시 표본수	실업률(%)			변동계수(CV: %)		
	직접	합성	복합	직접	합성	복합
4	3.78	3.82	3.37	65.30	21.32	10.48
5	3.70	3.75	3.31	59.58	20.90	9.01
6	3.61	3.72	3.29	55.59	19.91	7.74
7	3.58	3.72	3.27	50.92	19.55	6.95
8	3.55	3.74	3.28	47.85	18.76	6.02
9	3.52	3.74	3.28	45.33	18.11	5.49
10	3.62	3.74	3.35	42.27	18.00	5.06

상이 되면 추정치들이 안정되고 있음을 볼 수 있다. 한편, 표 3.2를 보면 통영시의 표본이 4개로 주어지고 거제시 자체의 표본조사구 수가 4개일 경우 직접, 합성 및 복합 추정법에 의한 실업률 추정치는 2.03%, 2.87%, 2.13%인데 직접추정법에 의한 실업률이 가장 낮고 합성추정법에 의한 실업률이 가장 높게 나타났으며 그 차이는 0.84% 포인트이다. 그리고 직접, 합성 및 복합 추정치의 변동계수는 78.33%, 48.64%, 23.51%으로 추정방법에 따라 변동계수가 크게 변함을 볼 수 있다. 다음으로 집락 내 여타 지역의 표본수가 고정되어 있는 상태에서 해당 소지역의 표본수가 증가함에 따라 각 추정방법에 따른 실업률 추정치와 추정치의 변동계를 보면 거제시와 통영시 모두 표본수가 증가함에 따라 추정된 실업률 및 변동계수의 크기가 감소하는 것을 볼 수 있다. 여기서 주목할 만한 사항은 거제시의 경우 표본수가 대략 8개 수준에서 복합추정법에 따른 변동계수의 크기가 17%대를 보이고 있으며, 통영시의 경우에는 표본수가 6개 수준 정도에서 17%대의 변동계수 값을 보임으로써 표본수가 그리 크지 않더라도 집락 내 여타 지역의 표본정보를 활용할 경우 추정치에 대한 상당한 정도의 통계적 신뢰성을 확보할 수 있다는 점이다. 이러한 경향은 나머지 모든 집락 내의 소지역들에 있어서도 비슷한 결과를 보여주고 있다. 또한, 같은 집락 내에서 정보를 제공하는 소지역의 표본수가 클 경우 정보를 빌려오는 소지역의 변동계수 값이 매우 크게 감소하고 있다. 이는 표 3.4의 제 3집락에서 가장 크게 나타나고 있는데, 창원시와 마산시는 배정된 표본수가 각각 16개 및 13개로 경남도 내에서 가장 표본수가 많은 시에 해당한다. 이 경우 진해시를 보면 표본수가 4개일 때 변동계수가 3.6%에 불과하며, 표본수가 7개로 늘어남에 따라 변동계수가 1%대

표 3.4. 제 3집락: 마산, 진해, 창원

마산시 표본수	실업률(%)			변동계수(CV: %)		
	직접	합성	복합	직접	합성	복합
4	3.81	3.38	3.28	66.97	34.54	19.76
5	3.58	3.26	3.14	61.71	33.34	17.64
6	3.55	3.30	3.15	56.14	31.51	15.13
7	3.48	3.25	3.10	53.93	30.83	14.23
8	3.54	3.28	3.15	48.84	30.06	12.92
9	3.51	3.26	3.14	46.09	28.92	11.90
10	3.50	3.27	3.17	43.81	28.13	11.04
진해시 표본수	실업률(%)			변동계수(CV: %)		
	직접	합성	복합	직접	합성	복합
4	2.90	3.30	2.85	75.04	9.26	3.60
5	2.65	3.28	2.66	71.88	9.01	2.88
6	2.54	3.28	2.58	66.33	8.63	2.21
7	2.48	3.26	2.52	62.39	8.40	1.87
8	2.46	3.28	2.51	58.12	8.01	1.59
9	2.47	3.29	2.52	54.66	7.80	1.42
10	2.41	3.26	2.46	52.12	7.59	1.25
창원시 표본수	실업률(%)			변동계수(CV: %)		
	직접	합성	복합	직접	합성	복합
4	1.99	3.71	2.21	72.10	27.55	9.29
5	1.90	3.65	2.10	66.83	26.55	7.51
6	1.83	3.56	2.01	64.01	26.32	6.65
7	1.81	3.52	1.97	58.83	25.36	5.69
8	1.73	3.49	1.89	56.38	24.89	5.01
9	1.76	3.54	1.90	51.40	23.69	4.24
10	1.72	3.47	1.85	49.53	23.28	4.06

수준으로 급감하고 있다. 이러한 경향은 비록 그 크기가 상대적으로 작지만 표 3.3의 제 2집락에서 표본수가 작은 밀양시(표본수 4개)와 양산시(표본수 6개)가 표본수가 큰 김해시(표본수 10개)로부터, 그리고 표 3.5의 제 4집락에서 표본수가 작은 사천시(표본수 4개)가 표본수가 큰 진주시(표본수 16개)로부터 상당히 큰 영향을 받고 있는 것으로 나타났다. 이는 지역의 특성이 유사한 인근지역 표본정보를 활용할 경우 소지역 추정치의 신뢰도를 효율적으로 향상시킬 수 있음을 암시하고 있다.

4. 결론

본 연구에서는 소지역 노동정책 수립에 반드시 필요한 실업률의 효율적 추정방안에 대하여 살펴보았다. 여기서의 효율성이란 추정치의 변동계수를 기준으로 일정 수준의 신뢰성을 유지한다는 전제 하에서 추정에 따른 비용을 최소화할 수 있는 표본조사구의 수를 도출하는 것을 의미한다. 분석에서는 최근 경남도가 관내 10개 시에 대하여 실시한 경제활동인구조사의 표본조사구 정보를 이용하였으며, 주요 분석 방법은 모의실험을 통하여 직접추정, 합성추정 및 복합추정 등 각각의 추정방법에 따른 소지역 추정결과(실업률과 변동계수)를 비교하였다. 물론, 이러한 경제활동인구조사에서 각 소지역별로 약 50여개씩 설정한 모든 표본조사구에 관한 정보를 활용할 수 있다면 그 추정결과는 3가지 추정방법 모두에서 더 정확하게 나타날 것이다. 그러나 본 연구의 목적은 연 1회 조사에 그치는 실업률에 대한 최상의 추정치를

표 3.5. 제 4집락: 사천시 및 진주시

사천시 표본수	실업률(%)			변동계수(CV: %)		
	직접	합성	복합	직접	합성	복합
4	3.88	4.10	3.28	78.43	22.89	12.47
5	3.56	4.01	3.09	73.72	22.59	10.79
6	3.43	3.95	3.02	69.50	21.85	9.47
7	3.28	3.92	2.96	65.52	20.80	8.34
8	3.27	3.94	2.98	61.98	20.44	7.76
9	3.23	3.94	2.97	59.61	19.71	6.99
10	3.18	3.90	2.96	56.27	19.22	6.43
진주시 표본수	실업률(%)			변동계수(CV: %)		
	직접	합성	복합	직접	합성	복합
4	3.32	4.64	3.50	64.07	50.26	22.31
5	3.12	4.50	3.32	59.68	47.19	19.85
6	3.09	4.46	3.29	55.26	44.83	18.29
7	3.04	4.41	3.24	51.77	43.28	17.22
8	3.06	4.40	3.24	48.62	42.04	16.48
9	3.11	4.47	3.30	45.16	39.55	15.01
10	3.05	4.43	3.24	42.68	38.26	14.11

구하는 것이 아니라 각 시군 등 소지역들이 최소한의 조사비용을 투입하여 매월 단위의 소지역 통계를 지속적으로 산출해낼 수 있는 방안을 모색하려는 데 있다. 이러한 목적 하에서 수행한 에서 다음과 같은 의미 있는 결과를 얻을 수 있었다.

첫째, 집락 내에서 인근 소지역의 표본조사구 수가 클 경우 변동계수도 크게 감소하였다. 이와 같은 현상은 김해시, 밀양시 및 양산시를 묶은 제 2집락, 마산시, 진해시 및 창원시를 묶은 제 3집락, 그리고 사천시와 진주시를 묶은 제 4집락에서 나타났는데 그 중에서도 제 2집락의 밀양시가 집락 내 여타 지역으로부터 가장 큰 영향을 받은 것으로 나타났다.

둘째, 경제활동인구조사에서 표본 수가 매우 작은 시, 즉 진해시(4), 밀양시(4), 양산시(6), 사천시(4), 거제시(6), 통영시(4) 등에서 표본조사구 수를 10개 정도로 늘리면 각 집락 내 모든 지역의 변동계수가 크게 감소할 것으로 예상된다.

셋째, 모의실험 결과에 의하면 각 소지역에서 최소 10개 정도의 표본조사구를 확보하면 복합추정법에 따라 변동계수 15% 이하의 수준에서 소지역 단위 실업률을 추정할 수 있을 것으로 보인다.

넷째, 통계청의 경제활동인구조사 자료를 이용할 수 있을 경우, 해당 소지역에 대해서라도 표본조사구 수를 늘려줄 것을 요청할 필요가 있음을 발견하였다.

다섯째, 기초 지자체에서 자체 조사를 할 경우 각 소지역별로 최소한 10개의 표본조사구 수를 확보한다면 변동계수가 15% 이하인 통계청 요구 수준의 신뢰성을 확보할 수 있을 것으로 확인되었다.

끝으로, 소지역 통계의 개발과 관련해 같은 집락으로 묶이는 소지역들의 특성이 유사하지 않을 경우 간접추정량은 심각한 편향을 유발할 수 있으므로 최적 집락화 방안에 대한 연구가 이루어져야 할 것이다.

참고문헌

김달호, 김남희 (2002). 반복조사에서 소지역자료의 페이지안 분석, <응용통계연구>, 15, 119-128.

- 김영원, 성나영 (2000). 소지역 통계 생산을 위한 추정방법, *Journal of the Korean Data & Information Science Society*, **11**, 111-126.
- 김재두, 신기일, 이상은 (2005). 공간 시계열 모형을 이용한 소지역 추정, <응용통계연구>, **18**, 627-637.
- 박종태 (1998). 소지역추정, <평택대학교 논문집>, 11.
- 박종태 (2001). 국내 경제활동인구조사에 대한 소지역 추정의 경험적 비교, <평택대학교 자연과학연구 창간호>, 69-79.
- 박종태, 이상은 (2001). 소지역 추정법에 관한 비교연구, *Journal of the Korean Data & Information Science Society*, **12**, 47-55.
- 양영준, 이상은, 신민웅 (2003). Kalman Filter기법을 이용한 실업자 수의 소지역 추정, <응용통계연구>, **16**, 239-246.
- 이계오 (2000). 시·군·구 실업자 추정을 위한 소지역 추정법, <응용통계연구>, **13**, 276-86.
- 이계오 (2002). <소지역 추정법에 의한 시·군·구 실업통계 개발>, 통계청.
- 정연수, 이계오, 이우일 (2003). 시·군·구 실업자 총계 추정을 위한 설계기반 간접추정법, <응용통계연구>, **16**, 1-14.
- 전경배 (2001). 공식통계의 소지역 적용사례, <계간국민계정>, **1**, 22-37, 한국은행.
- Fabrizi, E., Ferrante, M. R. and Pacei, S. (2007). Small area estimation of average household income based on unit level models for panel data, *Survey Methodology*, **33**, 187-198.
- Falorsi, P. D., Falorsi, S. and Russo, A. (1994). Empirical comparison of small area estimation methods for the Italian Labour Force Survey, *Survey Methodology*, **20**, 171-176.
- Farrell, P. J. (2000). Bayesian Inference for small area proportions, *Sankhyā: The Indian Journal of Statistics, Series B*, **62**, 402-416.
- Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal, *Statistical Science*, **9**, 55-93.
- Rao, J. N. K. (2003). *Small Area Estimation*, Wiley, New York.
- Ugarte, L., Militino, A. F., Prado, C., Vincente, J. S. and Ayestaran, M. (2007). Small area estimation in official statistics: The labour force survey of the basque country, *Proceedings of the Section on the Small Area Estimation, IASS Satellite Conference*.

An Efficient Estimation of Local Area Unemployment Rate Based on Small Area Estimation

Soo Taek Kim¹

¹Department of Information & Statistics, Gyeongsang National University

(Received November 2011; accepted November 2011)

Abstract

Small area estimation has received significant attention in recent years due to a growing demand for reliable local area statistics. Traditional area-specific direct estimates based solely on sample survey data in the areas of interest do not provide adequate small area precision; however, design-based indirect local area estimators borrow strength from sample observations of related areas to increase the effective sample size. Design-based indirect estimation methods such as synthetic and composite estimators are considered to adjust local area unemployment rate estimates in the Korean Economically Active Population Survey. This study suggests an efficient alternative to minimize the cost to construct the unemployment rate of a local area through simulation under the condition that we can maintain a certain level of CV for the estimates. We obtained the results that the composite estimators using a sample size greater than 10 are more stable and significant at the level of CV 25% in our design scheme.

Keywords: Small area estimation, composite estimator, coefficient of variation(CV), simulation.

This study was supported by Gyeongsang National University in 2009.

¹Professor, Department of Information & Statistics, Gyeongsang National University, 501 Jinju-daero, Jinju Gyeongnam 660-701, Korea. E-mail: kstioa@gnu.ac.kr