

제3장

고용통계의 소지역별 추정 방법

권 순 필

제1절 서론

1. 연구 배경

경제활동인구조사는 우리나라 국민의 경제활동, 즉 국민의 취업, 실업 등과 같은 경제적 특성을 조사하여 거시경제 분석과 인력자원의 개발정책 수립에 필요한 기초자료인 노동공급, 고용구조, 가용노동시간 및 인력자원의 활용 정도를 제공하고, 정부의 고용정책 입안 및 평가에 필요한 기초자료를 제공하는 것을 목적으로 하는 통계청의 가구부문 경상통계 조사이다.

경제활동인구조사는 1962년 8월부터 경제기획원 통계국에서 표본조사방법으로 실시되어 왔으며, 2007년 현재 전국의 1,629개 표본조사구내 적격가구를 대상으로 매월 조사를 수행하고 있다. 표본가구수는 3만 3천 가구이며, 표본가구 내에 상주하는 자 중 만 15세 이상인 자가 직접 조사대상이 된다. 통계청에서는 전국과 지역(시·도)별로 성별·연령계층별·교육정도별 등의 범주에 대한 경제활동인구와 비경제활동인구 및 관련 비율(경제활동참가율, 고용률, 실업률) 등의 조사 결과를 온라인, 간행물, 통계DB(KOSIS) 등의 매체를 통하여 공표하고 있다.

1990년대 중반 이후에 우리나라는 지방자치제도의 부활과 외환위기를 겪는 큰 변화를 맞게 되었다. 이는 종래의 중앙정부 주도의 정책 수립,

수행을 벗어나서 지방정부의 필요에 의한 지역개발계획이 증가하고 주민들의 행정수요도 다원화됨으로써 다양한 지역통계 자료가 필요하게 되었음을 의미한다. 또한 실업이 사회적 문제가 되고 직업과 일자리에 대한 공급 문제가 중앙과 지방정부 모두의 당면한 문제가 되는 상황에서, 지방정부 주도의 실업대책을 수립하기 위한 지역의 통계가 필요하게 되었음도 의미한다. 그러나 경제활동인구조사는 전국과 광역시·도 등 규모가 큰 지역에 대해서만 그 조사 결과를 추정·공표하고 있어 시군구 단위의 지방자치단체에서 동 조사 결과를 직접 사용하기는 어렵다. 그럼에도 여타의 객관적인 관련 통계가 없어, 각 지방자치단체에서는 도 평균을 참고하여 단순히 지역별로 분류·집계한 자료나 행정자료(인구수 등)를 이용하여 정책을 수행하고 있는 실정이다.¹⁾

이러한 문제를 해결하기 위해 창원시 등 일부 지방자치단체에서는 경제활동인구조사 등 필요 통계에 대한 조사를 직접 수행하고 있으며, 그 외의 지역에서도 지역통계를 생산하고자 하는 노력이 증가하고 있다. 그러나 지방자치단체에서 지역통계에 대한 수요를 직접 충족시키기 위해서는 전문인력의 확보, 표본의 확대 등 인력과 예산이 추가로 필요한데, 이 재원을 마련하는 일이 쉽지 않다. 그것은 통계청 차원에서 지역통계를 작성하기 위해 표본규모를 크게 늘리는 것에도 마찬가지로 적용되는 어려움이다.

2. 연구 목적 및 필요성

이와 관련하여 직접 현장 조사를 하는 표본수를 늘리기보다는 통계적 추정 기법을 활용함으로써 최소 비용으로 지역통계 생산이 가능하도록 하는 방법론이 제시되었는데, 그에 대한 대안으로 등장한 것이 소지역 추정(small area estimation)기법이다. 소지역 추정이란, 표본수가 극히 적어 우리가 일반적으로 사용하는 방법으로는 신뢰성 있는 추정이 불가

1) 지방정부뿐만 아니라 중앙정부에서도 지역통계의 부실로 정책 집행에 오류를 범하는 경우가 있다. 감사원 확인 결과 보건복지부는 장애인시설 확충 계획을 수립하면서 지역별 저소득 중증장애인 비율 대신 전국적인 저소득 중증장애인 비율을 사용하여 필요 시설수 및 인원수를 과소하게 책정했다(한겨레 신문, 2007.7.19).

능한 경우에 사용하는 통계분석기법이다. 소지역의 의미에 대해서는 여러 가지 정의²⁾가 가능하겠지만, 본 연구에서는 소지역을 전국이나 광역 시·도에 대칭되는 지리적으로 작은 국가 행정구역인 시·군·구로 정의한다.

현재 통계청에서 시행하고 있는 경제활동인구조사는 16개 시·도 지역을 기준으로 표본설계 되어 있어 소지역에 해당하는 표본크기는 매우 작아지게 된다. 이 경우 각 지역에 해당하는 표본만을 사용하는 직접추정량은 소지역에 대해서는 신뢰성 있는 추정값을 제공하기 힘들다. 이에 반해 간접추정량은 부족한 표본크기를 보완하는 방법으로 유사한 지역의 정보 또는 그 지역의 과거 정보를 이용하거나 센서스 또는 다른 통계조사의 관련 정보를 이용하여 보다 신뢰성 있는 추정값을 제공하게 된다. 과거에 비해 발달된 정보처리기술로 행정자료 등 보조정보의 저장 및 획득이 용이해진 점도 소지역 추정에 대한 연구를 활발하게 한 계기가 되었다.

이러한 이유로, 국내에서도 소지역 추정 기법에 대한 연구 활동이 활발하다. 90년대 중반부터는 이상은·진영(1996), 박종태·이상은(2001), 이계오(2002), 통계청(2006) 등에 의한 경제활동인구조사 자료를 이용한 고용통계, 특히 실업자수 추정에 관한 간접추정량에 대한 연구가 많이 이루어졌으며, 2000년대 이후로는 이상은(2003), 김재두·신기일·이상은(2005), 통계청(2006) 등에 의한 모형기반추정량에 대한 연구 활동이 활발하였다. 김달호·김재광(2004)은 연구의 범위를 넓혀 가계조사를 이용한 지역별 가구 소득과 주요 소비 지출에 대한 소지역 추정 연구도 수행한 바 있다.

소지역 추정의 경우 추정량이 갖는 큰 변동성과 지방정부의 정책 집행에 미치는 영향력 때문에 추정량의 선택, 공표 등에 신중을 기해야 하며 심도 있는 검토가 이루어져야 한다. 그러나 고용통계에서의 소지역 추정 결과를 도출한 선행 연구를 검토해 본 결과, 연구의 결과가 동일하지 않은 경향을 나타내고 있어 해당 통계를 생산하고 이용하는 데 혼선

2) 소지역(small area)이란 시, 군 또는 구와 같이 지리적으로 작은 지역을 나타낼 뿐 아니라 소영역(small domain), 즉 특정 연령이나 성별, 인종의 그룹과 같은 작은 부차모 집단을 나타낼 수도 있다.

이 나타나고 있다. 따라서 이에 대한 검토 및 실증 분석이 향후 고용통계의 소지역 통계에 대한 연구를 진행시키기 전에 반드시 선행되어야 함을 알 수 있다.

본 연구에서는 지금까지 제시된 소지역 추정법의 종류와 국내외 통계기관의 선행 연구 사례 등을 살펴보고, 소지역 추정량들에 대한 검토 및 재확인 단계로서 경제활동인구조사 자료를 이용한 실증적 비교 분석 연구를 수행하였다. 비교 분석을 위해서 사용된 추정량은 직접, 합성, 복합추정량으로 설계기반 및 간접추정량을 대표하는 추정량들이다.

제2절 선행 연구

1. 소지역 추정량³⁾

소지역 추정량은 <표 3-1>에서 보는 바와 같이 크게 설계기반추정량(design-based estimator), 간접추정량(indirect estimator), 모형기반추정량(model-based estimator)의 3종류로 나뉜다.

설계기반추정량은 직접추정량으로서 조사된 자료 그 자체만을 이용하거나 또는 가능한 경우 그 지역의 행정자료나 센서스 자료를 보조정

<표 3-1> 소지역 추정량의 종류

추정량	종 류
설계기반	직접추정량(direct estimator) - 단순추정량, 사후층화추정량, 비추정량, 회귀추정량 등
간 접	합성추정량(synthetic estimator) 복합추정량(composite estimator) 등
모형기반	경험적최량선형비편향추정량(EBLIP) 경험적베이지안추정량(EB) 계층적베이지안추정량(HB) 등

3) 소지역 추정법에 관한 자세한 내용은 Rao(2003)를 참조.

보로 이용하는 추정량이다. 적정 표본수가 확보되는 대지역에서는 편향이 없는 이상적인 추정량이지만 해당 소지역에 배정된 표본의 크기가 작은 경우에는 추정량의 분산이 커져서 신뢰성이 떨어지게 된다.

간접추정량은 해당 지역의 조사자료뿐만 아니라 해당 지역을 포함하고 있는 더 큰 지역의 조사자료를 소지역 추정 과정에 이용하여 소지역 추정의 신뢰성을 확보하는 방법이다. 합성추정량은 소지역 추정시 소지역을 포함하는 대지역의 정보를 함께 이용하는 방법으로, 소지역과 대지역의 특성구조가 유사하다는 가정 하에서 이용된다. 합성추정량의 분산은 직접추정량의 분산에 비해 작지만, 전제된 가정이 성립하지 않을 경우에는 심각한 편향이 발생할 수 있다. 복합추정량은 직접추정량의 불안정성과 합성추정량의 잠재적 편향 가능성을 보완하기 위해 두 추정량의 가중평균을 취하는 형태를 갖는다. 그러나 이러한 간접추정량은 지역간에 발생하는 지역간 변동을 설명하지 못한다는 단점이 있다.

모형기반추정량은 보조정보들에 의해 설명되지 않는 지역간 변동을 포함하는 모형을 이용하는 추정량이다. 모형 내에 표본오차, 지역간 변동, 시계열 등을 모두 고려할 수 있다는 장점이 있고, 다양한 보조정보를 사용하여 추정량의 정도를 높일 수 있다. 경험적베이지안추정량(EB)과 계층적베이지안추정량(HB)은 실제로 가장 많이 쓰이는 방법으로, 베이지안추정량을 구하기 위해서는 복잡한 계산이 선행되어야 했으나 근래에는 컴퓨터와 다양한 계산 기법의 개발로 비교적 간단히 추정량을 구할 수 있다.

어떤 추정량이 가장 좋은 소지역 추정량인지는 주어진 상황-이용 가능한 보조정보의 양과 질 등-에 따라 다르지만, 선행 연구들을 살펴보면 일반적으로 모형기반추정량, 간접추정량, 직접추정량의 순으로 더 작은 오차를 가진다는 경험적인 추정 결과가 알려져 있고, 최근의 연구들⁴⁾도 모형기반추정법을 이용한 소지역 추정에 많은 초점을 맞추고 있다.

4) 최근의 연구 동향은 IASS의 위성 컨퍼런스인 SAE(<http://www.dipstat.ec.unipi.it/SAE2007/>)를 참고.

2. 연구 사례

가. 국내⁵⁾

우리나라에서도 이미 소지역추정법에 의한 소지역 고용통계의 개선 방안에 대한 연구와 논의에 대한 정보가 상당한 정도 축적되어 왔다. 통계청 연구 사례를 살펴보면, 이상은·진영(1996)은 「통계분석연구」지에 실린 “변동계수를 이용한 소지역 통계의 안정성 검토”를 통해 소지역에 대한 추정의 경우에는 직접추정량보다는 합성과 복합추정량이 보다 나은 CV값을 가진다는 것을 보인 바 있다. 그러나 P.D.Falorsi, S.Falorsi & A.Russo(1994)에 따르면, 합성추정법을 사용하는 경우의 추정량은 편의가 큰 것으로 나타나기 때문에 작은 값의 CV로만 추정방법을 선택하는 데에는 깊은 주의가 필요함을 동시에 언급하였다. 따라서 추정량의 선택 시에는 편의와 CV 모두를 고려하여야 한다. 이 연구에 계속하여 박종태·이상은(2001)은 경험적베이지안추정량(EB)에 대한 연구를 추가하였는데, 결과에 따르면 모형에 기반한 경험적베이지안추정량(EB)이 가장 작은 편의와 평균제곱오차를 가진다.

이계오(2002)는 「소지역 추정법에 의한 시·군·구 실업통계 개발」 보고서를 통하여 시·군·구 및 광역시·도 단위 세부영역(성별·연령대별, 성별·교육정도별)에 대한 직접, 합성, 복합추정량을 제안하였다. 이 연구에서도 직접추정량보다는 합성과 복합추정량이 보다 나은 CV값을 가진다는 경험적인 연구 결과를 도출하였다.

통계청(2006)에서는 기존의 연구에 모형기반추정량에 대한 연구를 추가하여 “소지역 실업통계 작성을 위한 표본규모 검토”라는 연구를 수행하였다. 이 연구는 2005년 경제활동인구조사 자료를 이용하여 소지역의 분기별 실업자수와 실업률에 대한 직접, 합성, 복합추정량, 모형을 이용한 계층적베이지안추정량(HB), 경험적베이지안추정량(EB) 결과를 구했는데, 연구 결과는 <표 3-2>에서 보는 바와 같이 소지역별 실업률을 위해서는 직접추정량보다는 복합추정량이나 계층적베이지안추정량(HB)

5) 통계청뿐만 아니라 학계에도 많은 연구 결과가 있으나, 본 연구에서는 통계청 내부의 연구를 중심으로 연구 사례를 살펴본다.

〈표 3-2〉 충청북도의 시/군단위 실업자 총계 추정결과(2005년 4/4분기)

(단위 : 명, %)

	직접추정법			복합추정법			HB추정법			EB추정법		
	추정값	CV	실업률	추정값	CV	실업률	추정값	CV	실업률	추정값	CV	실업률
청주시	6,631	15.0	2.1	7,076	5.9	2.3	5,975	8.6	1.9	6,455	14.6	2.1
충주시	3,7108	13.9	2.8	2,297	3.6	2.1	2,604	9.8	2.4	2,877	10.3	2.6
제천시	1,936	38.7	2.6	1,532	5.5	2.1	1,531	23.1	2.1	1,865	36.9	2.5
청원군	225	79.0	0.3	996	1.7	1.5	268	55.9	0.4	228	74.5	0.3

을 사용하는 것이 보다 안정된 추정량을 제공할 수 있을 것이라 결론지었다.

우리나라 실정에 맞는 고용통계의 소지역 추정법을 연구하는 데 위와 같은 선행연구 결과는 큰 도움이 된다. 그런데 위 4개의 연구를 검토해 본 결과, 연구 결과가 <표 3-3>에서 보는 바와 같이 동일하지 않은 경향을 보이고 있음이 나타났다.

〈표 3-3〉 고용통계의 소지역별 추정 연구 비교

연 구	결 과 (CV)
이상은·진영(1996)	합성 < 복합 < 직접
박종태·이상은(2001)	EB < 복합 < 합성 < 직접
이계오(2002)	복합 < 합성 < 직접
통계청(2006)	복합 < HB < EB < 직접

각 연구는 모두 경제활동인구조사 자료를 이용하여 수행되었다. 사용된 자료의 시점, 해당 지역 및 인근 지역 정보 등의 이용에는 차이가 있을 수 있지만, 각 연구들은 기본적으로 경제활동인구조사 결과와 인구추계조사 결과만을 이용한 연구들이다. 따라서 앞에서 기술한 시점

등 이용 정보의 차이는 전체적인 연구 결과의 흐름에 큰 영향을 미치지 않는다고 보고 결과를 보다 단순화시켜 살펴볼 필요가 있다. 연구 결과에 대한 실증적인 검토는 공식통계로서의 위치를 갖게 될 소지역 추정량에 대한 객관적 검증의 계기가 될 것으로 보인다. 따라서 향후 고용통계의 소지역 추정에 대한 연구를 진행시키기 전에 기본적으로 관련 연구에 대한 검토 및 실증 분석이 반드시 선행되어야 함을 알 수 있다.

본 연구에서는 이 부분에 대한 검토와 정리의 필요성 때문에 경제활동인구조사 자료를 이용하여 소지역 추정량들에 대한 실증적 비교 분석 연구를 수행한다. 비교 분석을 위해서 사용된 추정량은 직접, 합성, 복합 추정량으로 설계기반 및 간접추정량을 대표하는 추정량들이다. 또한 이러한 연구를 통하여 추정 프로그램에 대한 이해를 높여 향후 프로그램의 개선, 보완을 수월히 할 수 있게 되는 계기를 마련한다.

나. 국외

1) 미국

미국 노동통계국(Bureau of Labour Statistics: BLS)에서는 소지역 실업 통계 프로그램(Local Area Unemployment Statistics Program: LAUS)에 따라 전국적으로 약 7,000개 소지역⁶⁾에 대하여 취업, 실업에 관한 정보를 제공하고 있는데, 이는 센서스국(Census Bureau)에서 수집한 경상인구조사(Current Population Survey) 자료를 바탕으로 소지역에 관한 자료를 생산하는 것이다. BLS는 1950년대 이래로 소책자 추정방법을 사용하여 직접추정법을 적용하지 않는 지역에 대해서도 소지역 고용통계를 꾸준히 제공하고 있다. 그러나 예산 절감의 일환으로 CPS 표본 규모를 줄였기 때문에 1996년 1월부터는 시계열 모형에 의한 모형기반추정량을 제공하고 있다. 추정방법은 CPS 표본 자료가 확률적으로 변화하는 노동력 시계열인 signal 성분과 noise 성분의 합으로 표현되는 모형 추정에 근거한다. 표본 설계 정보에 따라 월별 CPS 노동력 추정치는 시계열 모형에

6) 7,000개 지역은 모든 주, DC, Puerto Rico, 대도시(metropolitan areas), 작은 노동시장 지역(small labor market areas), 카운티, 뉴잉글랜드에 있는 도시들, 인구 25,000명 이상의 도시를 말함.

실업보험자료(Unemployment Insurance Data: UI)와 경상고용통계자료(Current Employment Statistics Data: CES)를 결합하여 계산한다. 즉, 이전에 추정된 방법과는 달리 좀더 체계적인 방법으로 보조자료와 과거와 현재의 표본자료를 모두 이용하여 표본크기가 적음으로써 발생하는 CPS 추정치의 큰 분산값을 줄이고자 하는 것이 기본적인 생각이다.

물론 이와 같은 모형기반추정법을 실제 추정에 적용하기 위해서는 많은 노력이 필요하지만, BLS는 상당한 시간과 노력을 들여 실업보험자료에 대한 데이터베이스를 정비하고 모든 자료를 표준화하여 향후 소지역 추정법의 개발과 개선을 성공적으로 이끌 수 있었다.

2) 캐나다

캐나다의 경우, 영국 Economist지가 선정하는 세계 우수통계청에 캐나다 통계청이 선정된 바 있으며, 타국 통계청에 전문인력을 빈번하게 파견할 정도로 국제적으로 높은 인지도를 유지하고 있는 통계 선진국이다. 그만큼 캐나다는 소지역 추정에 대해서도 가장 활발하고 다양한 연구 활동을 하는 나라이다. 간접추정량 이외에도 다양한 모형(회귀모형, 비선형모형, 시계열모형 등)을 적용하는 연구를 캐나다 통계청과 학계가 공동으로 연구한다.

캐나다에서는 53개의 소지역으로 구분된 각 고용보험(Employment Insurance: EI) 지구의 실업률을 근거로 하여 실업보험수당 수혜자와 수혜기간을 결정하고 있다. 그러나 거의 모든 가계조사의 기초자료로 사용되는 노동력조사(Labour Force Survey: LFS)는 일부 EI 지역에서 너무 작은 표본의 크기를 가지고 있어 의미있는 추정량을 산출할 수가 없다. 그래서 캐나다 통계청에서는 LFS의 표본설계를 이원화함으로써 각 EI 지역의 실업률을 추정하고 있다. 즉, 표본설계 첫 단계에서는 유효한 전국·주별 통계의 산출에 필요한 적정수의 표본을 배분하고, 두 번째 단계에서는 나머지 표본을 소수의 인구를 가진 EI 지역에 집중 배분⁷⁾함으

7) 캐나다의 LFS 표본가구수는 약 5만 4천 가구이며, 이는 인구수 대비 표본수를 비교해 보면 우리나라의 약 2배가 넘는 수준이다. 이를 통해 가장 작은 소지역 발표 단위인 Economic Region에 대해서도 실업자의 3개월 평균 추정 결과에 대한 직접추정량의 CV가 25%를 넘지 않게 표본을 배정하도록 설계한다.

로써 대지역과 소지역에서 각각 의미있는 추정량을 끌어내는 설계기반 추정법을 제공하고 있다. 그리고 특별히 필요한 경우에만 간접추정량을 함께 사용하고 있다.⁸⁾

그러나 이러한 LFS의 주요 목표가 아닌 통계의 산출이 요구될 때는 시간·공간상의 보조정보를 이용하는 간접추정법을 이용하여 소지역 통계를 산출하고 있다. 예를 들어, 캐나다 통계청은 캘거리시의 155개 구역별 가계지출액에 대한 통계를 산출하기 위해 소득수준, 가계구성원 수, 연령별 인구수에 대한 종합적인 통계를 제공하는 센서스 자료를 보조정보로 이용하여 회귀 추정량을 계산하고, 또한 캘거리시와 비슷한 인구수를 가진 에드먼턴시에 배분된 가계지출조사 표본을 캘거리시의 표본으로 가정하여 충분한 표본수를 확보함으로써 각 구역별 가계지출액을 추정하였다.

3) 호주

호주 통계청(Australian Bureau of Statistics: ABS)에서는 1970년대부터 소지역 추정 연구를 시작하였다. 연구 초기에는 합성추정량, 비추정량 등에 초점을 맞추었고, 1980~90년대로 들어서면서 의료, 장애, 노동력, 가구 소비 등 다양한 분야에 대한 적용을 시도하였다. 이후 소지역 추정 결과에 대한 품질 측정에까지 연구 범위를 확대하였고, 2005년에는 소지역 추정량의 생산, 사용, 품질 평가, 타당성 등에 대해 기술한 “*Small Area Estimation Practice Manual ABS(2005)*”를 발간하기도 하였다.

ABS에서는 소지역 추정량을 생산하기 위하여 주로 다음의 네 가지 선택 사항을 이용한 회귀모형 구조를 이용하고 있다.

- ① 선형 혹은 일반화된 선형모형을 사용할지 여부
- ② 모형은 일변량 혹은 다변량인지 여부
- ③ 임의효과를 포함할지 여부
- ④ 모형을 단위/개인 수준으로 할지 지역 수준으로 할지 여부

모형의 주요 보조정보는 인구센서스 자료이다.

8) 일례로, 노스웨스트주(Northwest Territories)의 경우 역사, 지리, 환경적인 요인 때문에 적은 표본수를 배정하여 LFS 조사를 실시하고 있는데, 추정량의 신뢰성을 향상시키기 위하여 직접추정량 대신 3개월 이동평균을 근간으로 추정하고 있다.

이러한 단계를 거쳐 생산한 소지역 추정량은 실제 적용에 앞서 그 품질에 대한 측정 문제가 남게 된다. 그래서 ABS에서는 다음과 같은 세 개의 단계를 거쳐 소지역 추정량에 대한 품질을 측정한다.

- ① 사용된 소지역 모형의 적합과 특정화(specification) 부분의 통계적 적합성을 평가
- ② Brown et al.(2001)이 제안한 품질 진단 과정을 적용
- ③ 소지역 추정량의 그럴듯함(plausibility), 일치성, 표면적 타당성 등에 대하여 사용자, 학교, 정부 기관들로부터 ‘전문가 의견’을 듣고 그 타당성을 확보

소지역 추정 연구를 해 온 이래로 호주 ABS에서 나온 품질 관련 이슈들을 살펴보면, 첫째, 보조자료가 기대한 만큼 관심 자료와 강하게 관련되지 않는 경우가 발생하고 있어 모형에 대한 신뢰도와 정확성에 대한 우려를 발생시킨다. 둘째, 모형의 성능을 평가하기 위한 표준화된 진단 방법이 존재하지 않는다. 셋째, 앞에서 언급한 품질 측정 문제에서 거쳐야 하는 세 단계를 잘 조화시켜야 한다. 마지막으로, 최종 결과에 대한 사용자 이해를 높여야 한다.

이러한 이유로 ABS에서는 소지역 추정을 공표하기 앞서 다음과 같은 사항을 숙지하고 고려할 것을 권유하고 있다.

- ① 좋은 모형을 가지는 소지역 추정량 생산을 위해서는 시간과 비용 소비가 수반됨.
- ② 소지역 추정은 한 묶음으로 대량생산할 수 없음.
- ③ 궁극적으로 모형에 들인 시간과 노력을 정당화시킬 만큼 모형이 잘 적용될지 여부가 명확하지 않음.
- ④ 충분한 시간을 주지 않고 빨리 결과를 공표하도록 하는 압박이 있음.
- ⑤ 센서스 자료를 보조정보로 이용하는 한, 센서스 조사 사이의 기간 동안에 생기는 변화를 모형에 어떻게 반영시킬 수 있는가에 대한 문제가 상존.
- ⑥ 품질 측정을 위한 상대제곱근평균제곱오차(RRMSE) 유도가 완전히 소프트웨어로 만들어지지 않음. 즉, 특수화된 코드를 사용해야 함.

- ⑦ 국가통계기관은 종종 대학과의 업무 협조를 고려해야 할 필요가 있음.

4) 영국

영국의 소지역 추정치는 기본적으로 Neighbourhood Statistics DB를 기반으로 생산된다. Neighbourhood Statistics DB는 빈곤층에 대한 공공의 정책 계획 수립에 대한 보조를 목적으로 만든 것으로, 대부분의 데이터는 행정자료나 센서스로부터 얻는다. 반면에 이런 방법으로 얻을 수 없는 중요 변수들은 표본조사로부터 추정해야 하는데, 이런 이유로 소지역 추정 연구가 활발히 이루어지게 되었다.

영국은 노동력조사에 대하여 공식통계(Official Statistics)를 생산하고 있다. 초기에는 합성추정량을 생산하였으나 현재는 분기별로 모형기반 추정량 중에서도 경험적인 최량선형비편향추정량(EBLUP)을 생산하여 발표하고 있으며, 보조정보로는 행정자료인 실업수당요구율을 이용하고 있다. 이러한 소지역 추정량의 공식통계 생산은 모형기반추정량이 지역 정부에 의해 국가통계 수준으로 용인될 수 있다는 중요한 의미를 가지고 있다.

영국 통계청에 의해 제공되는 다른 소지역 추정량은 가구조사 자료로부터 얻어낸 가구 평균 수입에 대한 추정량이다. 이 추정량의 결과는 약 4천 가구 정도를 포함하는 규모의 지역들에 대하여 1998/99년에 대한 추정량을 2001/02년에 발표하였다. 이것은 실험통계(Experimental Statistics)로서 아직 공식통계로 인정받고 있지는 않다. 그 외에도 지역간 상관관계를 포함하는 모형추정량에 대한 연구도 지속 중이다.

다른 국가들과 마찬가지로 영국 통계청 역시 사용자 그룹으로부터 소지역 추정에 대한 다음과 같은 몇 가지 문제점에 대한 해결을 요구받고 있다. 첫째, 소지역 추정량과 대지역의 직접추정량 사이의 일치성 문제를 해결해야 한다. 둘째, 반복 추정의 경우 시간이 흐름에 따라 발생할 수 있는 변화를 어떻게 평가할 것인가에 대한 문제를 해결해야 하고, 마지막으로 빈곤층에 대한 소지역 추정치 비율이나 분위수 등을 적용한 모형 검토를 요구받고 있다.

5) 일본

일본의 경우, 통계국이 소지역 추정에 주목하게 되면서 2003년부터 2005년에 외국의 적용사례 및 참고문헌을 기초로 소지역 추정 방법에 대해서 조사함과 동시에 고용통계지역추정연구회를 설치해서 관련 추정법을 검토하였다.

다양한 소지역 추정법을 노동력조사에 적용해 보고 그 추정 결과를 근거로 각 추정법에 대해서 일반성, 재현성, 간명성⁹⁾, 정부통계로서 적용성, 추계 결과의 적용성, 실용성¹⁰⁾의 6개 관점으로 종합적인 평가를 실시하였다. 그 결과 일본 노동력조사에 적용시키기 위해서 미국 노동통계국에서 채용하고 있는 시계열 회귀모형에 공간(지역)정보를 회귀항으로 넣은 모형을 이용하였다.

일본의 경우, 소지역 추정에 대한 연구 시작 시기는 다른 국가들에 비해 늦은 편이지만, 2006년에 모형을 이용한 소지역 추정 결과를 모형 추계치라는 이름으로 발표하는 등 공식 발표는 비교적 빠르게 이루어졌다. 발표 범위는 도도부현별로 사분기, 연평균 결과와 함께 약 10년 전까지 역산한 결과도 함께 공표하였다.

6) 그 외 국가

상기한 5개국 외에도 각국의 통계청에 의한 소지역 추정 적용 연구는 다양한 분야에서 다양한 방법으로 지속되고 있다. 스페인은 2004년부터 EUSTAT과 학계의 공동 연구를 통해 경제와 산업통계에서 소지역 추정량을 생산하였으며, 현재는 노동력조사 특히 실업자수에 대한 소지역 추정량을 얻기 위한 시뮬레이션 연구를 수행하고 있다. 네덜란드 역시 지난 수년간 노동력조사 분야에서 모형을 이용한 소지역 추정 연구를 지속하고 있다. 그 외에도 스웨덴, 이탈리아 등의 유럽 국가들은 노동력조사, 가계조사, 건강·의료, 빈곤, 농업 등의 분야에 대한 실험적인 연구를 수행하고 있으며, 그 추정 방법으로는 합성·복합추정량 등 간접추정량으로부터 베이지안 방법을 이용한 모형기반추정량까지 다양하다. 개발된 추정량의 성능에 대한 평가는 주로 총조사 자료 등을 이용한 시뮬레이션 연구를 통하여 수행한다.

9) 간명성 : 해석이 가능한지 여부.

10) 실용성 : 시의성 있는 결과를 얻을 수 있는지 여부.

제3절 분석 연구

통계청에서 매월 조사하고 있는 경제활동인구조사 중 충북지역의 2005년(1~12월) 자료를 이용하여 직접, 합성, 복합추정량을 구해 보고 각 추정량의 효율은 잭나이프 분산 추정법을 이용하여 비교 분석하였다.¹¹⁾

1. 추정량

가. 직접추정량

대영역 내에서 소지역 i 에 대한 실업자 총계를 추정하기 위한 직접 추정량에 대한 총계 추정공식은 다음과 같다.

$$\begin{aligned} \hat{Y}_{i.} &= \sum_{s=1}^2 \sum_{k=1}^{11} s_k \hat{Y}_{i.}, \quad i = 1, 2, \dots, I; \quad s = 1, 2; \quad k = 1, 2, \dots, 11 \\ &= \sum_{s=1}^2 \sum_{k=1}^{11} \sum_{h=1}^{n_i} s_k \hat{Y}_{ih}, \quad h = 1, 2, \dots, n_i \\ &= \sum_{s=1}^2 \sum_{k=1}^{11} \sum_{h=1}^{n_i} s_k M_{i sk} Y_{ih} \end{aligned}$$

여기서, s, k, h 는 각각 성별(남, 여), 연령대, 조사구를 나타내는 첨자,

n_i : 경제활동인구조사에서 소지역 i 의 표본조사구 수,

Y_{ih} : 성별 · 연령대별 · 소지역별로 경제활동인구조사에서 조사한 실업자 수,

$\hat{X}_{i.}$: 성별 · 연령대별 · 소지역별 15세 이상 상주 추계인구,

$X_{i.}$: 성별 · 연령대별 · 소지역별 경제활동인구조사의 15세 이상 조사인구,

$M_{i sk} = \hat{X}_{i.} / X_{i.}$: 승수

11) 추정량들의 효율성 비교를 위하여 잭나이프 분산 추정 방법을 이용한 것을 제외하면, 추정량의 형식과 방법은 이계오의 2002년 연구를 참조하였다.

나. 합성추정량

대영역 내에서 소지역 i 에 대한 실업자 총계를 추정하기 위한 합성 추정량에 대한 총계 추정공식은 다음과 같다. 특히 합성추정량은 관심 지역(i 지역)을 포함하는 대지역(충북)에서의 관심변수와 보조변수의 비인 Y/X 와 관심지역에서의 비인 Y_i/X_i 가 같다는 가정 아래 관심변수를 추정하게 된다.

$$\widehat{Y}_{i.}^S = \sum_{j=1}^4 \frac{{}_j\widehat{X}_{i.}}{\widehat{X}_{i.}} {}_j\widehat{Y}_{dir}, \quad j = 1, 2, 3, 4$$

여기서, ${}_j\widehat{X}_{i.} = \sum_{i=1}^{m_i} {}_j\widehat{X}_{i.}$, ${}_j\widehat{Y}_{dir} = \sum_{i=1}^{m_i} \sum_{h=1}^{n_i} {}_jM_{i.} {}_jY_{ih}$,

m_i : 소지역(i)이 포함된 시·군·구를 나타내는 첨자,

j : 성별(남, 여) 및 연령대별(15~29세, 30세 이상) 4개의 범주를 나타내는 첨자,

${}_jY_{ih}$: 범주별·소지역별로 경제활동인구조사에서 조사한 h 조사구의 실업자 수,

${}_j\widehat{X}_{i.}$: 범주별·소지역별 15세 이상 상주 추계인구,

${}_jX_{i.}$: 범주별·소지역별 경제활동인구조사의 15세 이상 조사인구,

${}_jM_{i.} = {}_j\widehat{X}_{i.} / {}_jX_{i.}$: 승수

합성추정량을 구하기 위한 4개 범주는, 2005년 경제활동인구조사에서의 실업률 변동 상황이 <표 3-4>에서 보는 바와 같이 성별과 29세 이하와 30세 이상의 연령대별로 의미있는 차이가 나타나고, 또한 2002년 이계오의 연구와 2006년 통계청 지역통계과 연구와의 객관적 비교 분석을 위하여 그 둘의 연구와 동일한 범주를 사용하도록 한다.

〈표 3-4〉 성별, 연령대별 실업률(2005년 전국)

(단위 : %)

성별	연령대별						
	계	15~19세	20~29세	30~39세	40~49세	50~59세	60세이상
남	4.0	13.2	9	3.5	2.7	3	1.8
여	3.4	11.8	6.5	3.1	2.3	1.8	0.7

다. 복합추정량

대영역 내에서 소지역 i 에 대한 실업자 총계를 추정하기 위한 복합 추정량에 대한 총계 추정공식은 다음과 같다.

$$\hat{Y}_{i.}^C = \hat{w}_{i(opt)} \hat{Y}_{i.} + (1 - \hat{w}_{i(opt)}) \hat{Y}_{i.}^S$$

여기서, $\hat{w}_{i(opt)} = \frac{MSE(\hat{Y}_{i.}^S)}{MSE(\hat{Y}_{i.}^S) + Var(\hat{Y}_{i.})}$ 이다.

라. 잭나이프 MSE 추정

각 추정량의 MSE(직접추정량의 경우 분산)를 계산하기 위하여 반복 표본추출법을 이용하는 잭나이프(jackknife) 추정법을 고려하였다. 잭나이프 추정법은 추정량의 형태에 국한되지 않아 계산 및 응용이 쉽고 추정 결과도 신뢰할 만하다는 장점이 있다.

각 추정량들의 MSE 추정을 위한 추정식의 일반적인 형태는 다음과 같다.

$$mse_J(\hat{Y}_{i.}) = \widehat{Var}_J(\hat{Y}_{i.}) + [\widehat{Bias}_J(\hat{Y}_{i.})]^2$$

여기서, $\widehat{Var}_J(\hat{Y}_{i.}) = \frac{n_i - 1}{n_i} \sum_{h=1}^{n_i} [\hat{Y}_{i.}(h) - \frac{1}{n_i} \sum_{h=1}^{n_i} \hat{Y}_{i.}(h)]^2$

$$\widehat{Bias}_J(\hat{Y}_{i.}) = (n_i - 1) \left[\frac{1}{n_i} \sum_{h=1}^{n_i} \hat{Y}_{i.}(h) - \hat{Y}_{i.} \right]$$

$\hat{Y}_{i.}(h)$: h 번째 조사구를 제외하고 계산된 추정량

합성, 복합추정량 계산시 위 식에서 $\hat{Y}_{i.}$ 대신 $\hat{Y}_{i.}^S, \hat{Y}_{i.}^C$ 를 대입하여 같은 방식으로 계산한다.

2. 실증분석

소지역 추정법의 실증 분석을 위해 충청북도의 11개 시, 군 중 인구 10만 이상인 중소도시(청주, 충주, 제천, 청원)의 2005년도 경제활동인구 조사 자료를 사용하였다. 우선 기본적인 충청북도의 추계 인구 분포를 살펴보면 <표 3-5>와 같고, 도 단위의 실업자수 산출을 위해 81개의 표본조사구를 대상으로 경제활동인구조사를 수행하고 있다.

<표 3-5> 충청북도 시/군 단위 추계 인구 및 표본조사구 수 분포
(2005년 7월 현재)

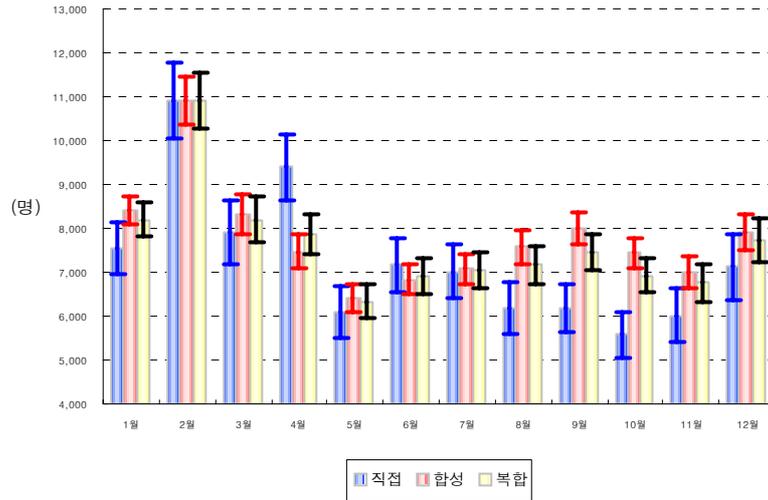
(단위 : 천 명, 개)

시군구	추계 인구			15세 이상 추계 인구			조사구 수
	계	남	여	계	남	여	
청주시	629	314	315	489	240	249	30
충주시	206	104	102	167	84	83	12
제천시	138	70	68	113	57	56	8
청원군	120	61	59	100	51	49	6
총 북	1,488	750	738	1,198	597	601	81

우선 청주 지역의 2005년도 월별 실업자수 추정 결과를 살펴보면 [그림 3-1]과 같다.

청주 지역은 2월에 가장 많은 실업자가 발생한 것으로 나타나고 있으며, 어느 한 추정량의 추정값이 많고 적은 것에 대한 뚜렷한 경향은 보이지 않고 있다. 오차 막대는 95% 신뢰수준에서 각 추정량의 신뢰구간을 표현해 본 것인데, 이 역시 직접추정량의 신뢰구간이 가장 넓고,

[그림 3-1] 청주 지역의 2005년도 월별 실업 동향



합성추정량의 신뢰구간이 가장 좁았다. 이는 표본이 적을 때 직접추정량이 가지는 신뢰성 문제를 단적으로 보여준다고 할 수 있다. 복합추정량의 경우 직접과 합성추정값 사이에 추정값이 주로 존재하였고 신뢰구간은 역시 직접추정량보다는 좁고 합성추정량보다는 넓은 경향을 보였는데, 이는 복합추정량의 형태가 직접추정량과 간접추정량의 가중평균 형태이므로 나타나는 결과로 해석할 수 있다. <표 3-6>을 보면 2005년도 각 월별 청주 지역의 추정량별 CV값을 알 수 있다.

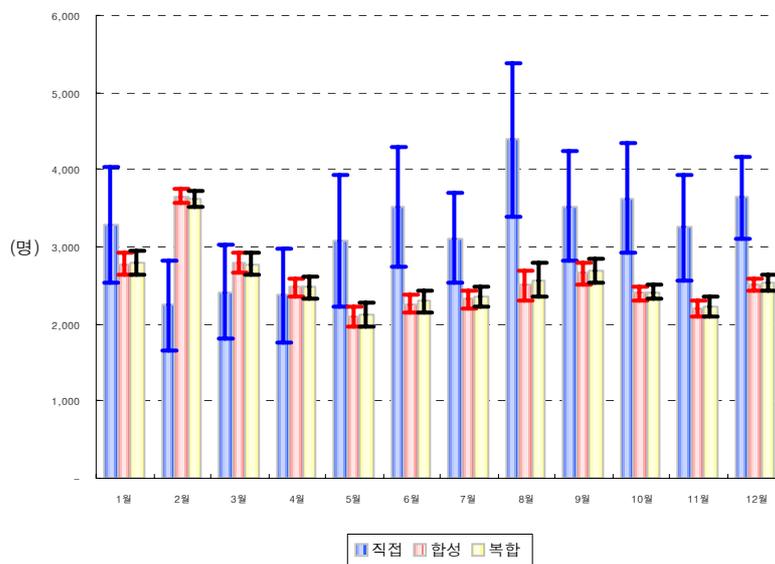
<표 3-6> 청주시 추정량의 월별 CV 변화

(단위 : %)

	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
직접	21.64	21.88	26.27	21.84	26.62	24.22	23.72	26.84	24.98	26.48	29.00	29.30
합성	11.05	14.13	15.34	14.52	14.50	14.25	13.80	14.18	12.35	12.26	14.66	15.11
복합	13.21	16.33	18.11	16.24	17.18	16.57	16.22	17.29	15.30	15.37	17.81	18.23

다음으로 충주 지역의 2005년도 월별 실업자수 추정 결과를 살펴보면 [그림 3-2]와 같다.

[그림 3-2] 충주 지역의 2005년도 월별 실업 동향



충주 지역은 8월에 가장 많은 실업자가 발생한 것으로 나타나고 있으며, 대체로 직접추정량이 가장 많은 실업자수 추정값을, 합성추정량이 적은 실업자수 추정값을 나타내었다. 그러나 합성추정값과 복합추정값의 차이는 직접추정값의 그것과 비교하여 크지 않은 것으로 나타났다. 오차 막대 역시 직접추정량의 신뢰구간이 가장 넓고, 합성추정량의 신뢰구간이 가장 좁았다. 충주 지역의 경우, 복합추정량의 월별 추정값과 신뢰구간의 추정 경향이 합성추정량과 유사하게 나타났는데, 이는 가중치 $\hat{w}_{i(opt)}$ 가 직접추정량보다 합성추정량에 보다 많은 가중치를 주어 나타난 현상으로 보인다. <표 3-7>을 보면 2005년도 각 월별 충주 지역의 추정량별 CV값을 알 수 있다.

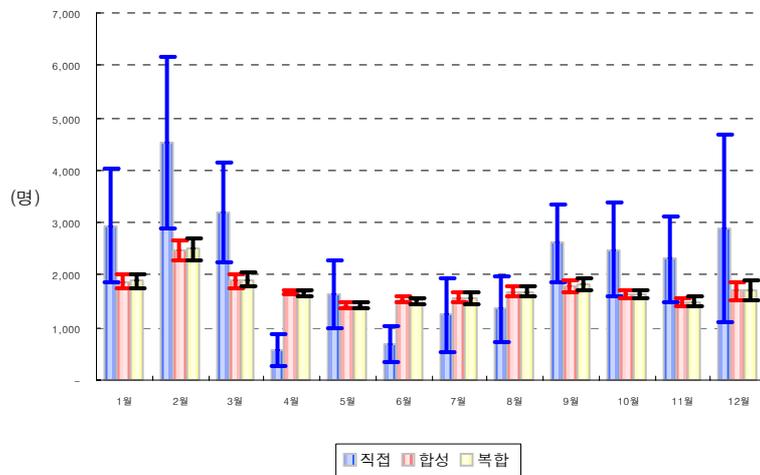
〈표 3-7〉 충주시 추정량의 월별 CV 변화

(단위 : %)

	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
직접	40.37	46.17	44.46	45.28	49.37	38.41	33.01	39.94	35.26	34.54	37.42	25.95
합성	8.99	4.78	8.02	8.59	11.09	9.37	8.68	13.98	9.31	6.55	8.71	6.18
복합	10.16	5.40	9.33	9.87	12.26	10.26	9.77	15.51	10.32	7.06	9.65	6.77

다음으로 제천 지역의 2005년도 월별 실업자수 추정 결과를 살펴보면 [그림 3-3]과 같다.

[그림 3-3] 제천 지역의 2005년도 월별 실업 동향



제천 지역은 2월에 가장 많은 실업자가 발생한 것으로 나타나고 있으며, 충주 지역과 마찬가지로 대체로 직접추정량이 가장 많은 실업자수 추정값을, 합성 추정량이 적은 실업자수 추정값을 나타내었다. 또한 합성추정값과 복합추정값의 차이는 직접추정값의 그것과 비교하여 크지 않은 것으로 나타났다. 오차 막대 역시 직접추정량의 신뢰구간이 가

장 넓고, 합성추정량의 신뢰구간이 가장 좁았다. 제천 지역의 경우도 복합추정량의 월별 추정값의 경향이 합성추정량과 유사하게 나타났는데, 이는 충주 지역과 같은 현상 때문인 것으로 보인다. <표 3-8>을 보면 2005년도 각 월별 제천 지역의 추정량별 CV값을 알 수 있다.

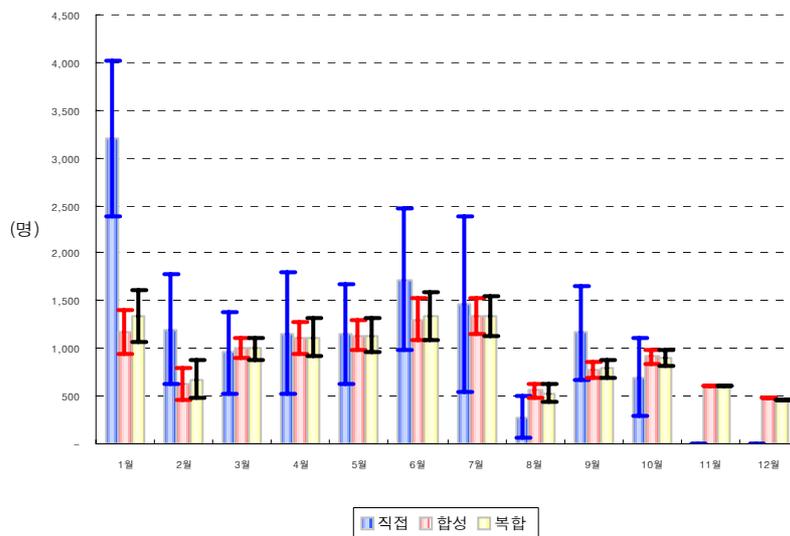
<표 3-8> 제천시 추정량의 월별 CV 변화

(단위 : %)

	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
직접	52.92	52.19	42.45	74.72	56.46	69.30	81.49	66.69	40.83	51.26	51.53	89.30
합성	9.27	11.14	9.58	3.80	6.00	4.66	8.69	6.98	8.32	6.99	7.80	14.08
복합	10.03	12.09	10.41	4.28	6.47	5.26	9.71	7.77	9.17	7.51	8.38	15.09

마지막으로 청원 지역의 2005년도 월별 실업자수 추정 결과를 살펴 보면 [그림 3-4]와 같다.

[그림 3-4] 청원 지역의 2005년도 월별 실업 동향



청원 지역은 1월에 가장 많은 실업자가 발생한 것으로 나타나고 있으며, 어느 한 추정량의 추정값이 많고 적은 것에 대한 뚜렷한 경향은 보이지 않고 있다. 그러나 오차 막대의 신뢰구간이 나타내는 경향은 앞에서 기술한 청주, 충주, 제천 지역과 유사하다.

청원 지역의 경우 특이한 현상은, 11월과 12월의 실업자수 추정 결과를 살펴보면 직접추정값은 0으로 계산된다는 것이다. 이는 표본조사구에서 단 한 명의 실업자도 조사되지 않았고 해당 지역의 조사 자료만 이용하는 직접추정법을 이용하였기 때문에 발생한 현상이다. 반면 합성과 복합추정값은 계산이 되었는데 이는 인근 지역의 보조정보를 이용했기 때문이다. 그러나 MSE값 계산 시에는 직접추정량뿐만 아니라 합성과 복합추정량의 MSE값도 추정할 수 없었다. 따라서 해당 2개월의 추정값은 어떤 추정량을 사용하더라도 의미있는 자료가 되지 못한다. 이는 작은 표본 규모로 인해 발생하는 문제로서 이러한 통계 추정값을 아무런 검증 없이 실제 적용한다면 잘못된 판단을 유도할 수 있는 대표적인 예가 될 것이다.

11월과 12월의 추정 결과를 제외하고 보면 청원 지역의 경우에도 직접추정량의 신뢰구간이 가장 넓고, 합성추정량의 신뢰구간이 가장 좁았다. <표 3-9>를 보면 2005년도 각 월별 청원 지역의 추정량별 CV값을 알 수 있다.

충북 지역의 각 시·군별로 2005년도 월별 실업 동향을 살펴보았다. 각 지역에 공통적으로 나타나는 현상은 복합추정량의 추정값과 CV가 모두 직접추정값과 합성추정값의 가중평균값으로 나타난다. 그리고 이 중에서 CV값으로만 소지역 추정량을 선택한다면 각 지역 모두 합성추정량의 CV가 가장 효율이 좋은 것으로 나타났기 때문에 합성추정량을 선택하여 소지역 추정을 할 수 있다. 그러나 합성추정법을 사용하는 경우의 추정량은 편의가 크기 때문에 작은 값의 CV로만 추정방법을 선택하는 데에는 깊은 주의가 필요하다(P.D.Falorsi, S.Falorsi & A.Russo, 1994).

〈표 3-9〉 청위구 추정량의 월별 CV 변화

(단위 : %)

	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
직접	31.62	59.19	56.26	69.23	57.32	53.92	78.63	100.36	52.77	73.18	-	-
합성	24.65	33.06	12.98	18.44	16.67	21.20	17.13	17.66	12.90	10.35	-	-
복합	25.47	36.29	14.96	21.49	19.64	24.26	19.67	22.44	13.94	11.78	-	-

제4절 결론

본 연구에서는 증가하는 소지역 통계 수요에 부응하고 통계적 추정 기법을 활용함으로써 최소 비용으로 지역통계 생산이 가능하도록 하고자 하는 목적으로 국내외에서 많은 관심을 받고 있는 소지역 추정법 중 직접, 합성, 복합추정량에 대한 실증적 분석 연구를 수행하였다. 이를 통해 통계청 내외에서 수행된 연구 결과들에 대한 체계적인 검토와 정리를 할 필요성이 있기 때문이다.

각 추정량에 대한 MSE 추정은 간단한 알고리즘을 통해서 계산할 수 있는 잭나이프 추정법으로 접근하였으며 3가지 추정량 모두에 동일한 계산 과정을 수행하였다. 그 결과 기존의 연구와는 다르게 CV 측면에서는 합성추정량이 직접추정량이나 복합추정량에 비해 가장 효율적이라는 것을 알 수 있다. 이는 이계오(2002)와 통계청(2006)의 연구와는 다른 경향을 나타낸 것인데, 상기한 두 연구가 직접추정량과 합성추정량의 가중평균 형태인 복합추정량 계산시 직접추정량 부분에 대한 변동성을 고려하지 않은 잭나이프 추정법을 적용한 데 기인한 것으로, MSE를 실제보다 과소하게 추정한 것으로 판단된다.

CV 측면에서는 합성추정량이 가장 효율적인 추정량이지만, 합성추정량의 경우 대지역과 소지역의 특성구조가 유사하다는 가정이 성립하지 않으면 심각한 편향이 발생할 수 있다. 따라서 직접추정량과 합성추정량의 이러한 단점을 보완한 복합추정량의 사용을 고려해 볼 수 있다.

복합추정량의 MSE는 직접추정량에 비해서 어느 정도 안정성을 보이기 때문이다. 그러나 추정량을 CV만으로 선택할 수는 없다. 편의에 대한 계산도 중요한 부분이기 때문이다. 문제는 현실적으로 소지역에 대한 편의를 계산할 수 있는 방법이 많지 않다는 데 있다. 본 연구에서도 이 부분에 대한 한계로 각 추정량의 편의를 계산하지 못했다. 그러나 현재 통계청의 지원으로 일부 시·군·구 지방자치단체에서 직접추정량 산출을 위한 지역통계를 생산하고 있기 때문에, 향후에는 이를 평가 기준이 되는 'gold standard'로 간주하고 가능한 여러 소지역 추정 방법을 실증적으로 비교해 보는 연구가 후속 수행된다면 각 추정량에 대한 보다 적극적인 검증 작업이 될 수 있을 것이라 생각한다. 또한 해외 각국의 사례를 살펴본 결과, 개발된 추정량의 성능에 대한 평가를 위하여 총조사 자료를 이용한 시뮬레이션 연구도 활발히 수행되고 있어 우리나라에서도 이를 소지역 추정량에 대한 검증의 한 도구로 사용해 볼 수 있을 것이다.

표본조사구의 크기가 큰 지역인 청주시에서는 세 추정량의 추정값이나 추정 정도가 어느 정도 동일한 경향을 갖는 것을 알 수 있었다. 따라서 일정 수준 이상의 표본수를 가지고 있는 시·군에 대한 실업 추정량은 불편성을 가지고 있는 직접추정량을 이용하여 산출할 수 있을 것이라 보이며, 이에 대한 후속 연구도 소지역 추정법에 대한 연구로서 계속 되어야 한다.

또한 행정구역 단위인 시·군·구를 중심으로 분석해 본 결과, 작은 표본 규모로 인해 표본오차와 추정치가 '0'의 값을 가지는 경우가 발생하여 추정에 영향을 미치는 경우가 있었다. 이는 자료 자체뿐만 아니라 추정치로서의 통계적 의미도 없는 것이고 실제 적용시 잘못된 판단을 유도할 수도 있다. 이러한 문제를 해결하기 위해서는 표본조사구 내의 관측값이 0이어도 해당 지역의 실업률을 계산할 수 있는 방법이 필요하다. 이러한 문제점에 대한 대안으로 인접지역의 정보를 보조정보로 하는 공간통계를 이용한 소지역 추정 모형도 고려할 수 있다(이상은, 2003).

본 연구에서는 이와 같이 실증 분석을 통한 추정량의 효율성을 비교해 보았다. 그러나 소지역 통계 생산을 위한 추정을 한다는 것은 추정식

을 개발하고 계산하는 기술적인 부분만을 의미하는 것은 아니다. 지방 정부에서 원하는 추정 결과의 범위(영역)는 어디까지인가, 시군구 이외에도 활용 가능한 소지역의 범위(생활권역별, 노동시장별)를 어떻게 설정할 것인가, 보다 정도 높은 추정량을 얻기 위해 필요하고 사용 가능한 보조정보(행정자료 등)는 어떤 것이 있는가 등, 고려해야 할 많은 부분이 실업률을 실제 정책 집행 현장에서 사용할 지방정부의 담당자들과의 긴밀한 협력을 전제로 하기 때문이다. 따라서 지방자치단체와의 지속적인 협력 강화를 통해 소지역 추정 결과에 대해 지역담당자 및 전문가와 의견을 교환하여 해당 지역에 보다 적합한 모형을 개발할 수 있는 노력이 필요하다. 또한 호주와 일본의 예에서 보듯이, 소지역 추정 결과를 근거로 우리나라의 상황에 맞는 관점을 선정하여 각 추정법에 대한 종합적인 평가를 실시해야 한다. 이는 추정량 연구에 대한 공개적이고 객관적인 평가 결과를 남길 수 있고, 후속 연구에 많은 시사점을 줄 수 있기 때문이다.

참고문헌

- 김달호·김재광(2004), 「가계조사 지역별 추정기법 개발」, 통계청 용역 연구보고서.
- 김재두·신기일·이상은(2005), “공간 시계열 모형을 이용한 소지역 추정”, 「응용통계연구」, 제18권 제3호, pp.627-637.
- 박종태·이상은(2001), “소지역 추정법에 관한 비교연구”, 「한국데이터 정보과학회지」, Vol.12, No.2, pp.47-55.
- 신기일·이상은(2003), “Model-Data Based Small Area Estimation”, *The Korean Communications in Statistics*, Vol.10, No.3, pp.637-645.
- 이계오·김규영(2003), “소지역추정량 실증분석”, 「통계연구」, 제8권 제1호, pp.137-170.
- 이계오(2002), 「소지역 추정법에 의한 시·군·구 실업통계 개발」, 통계청 용역 연구보고서.
- 이상은·진영(1996), “변동계수를 이용한 소지역 통계의 안정성 검토”, 「통계분석연구」, 창간호, pp.23-40.
- 일본 통계국, “소지역 추계와 노동력 조사에의 적용”, <http://www.stat.go.jp/data/roudou/pref/index.htm>.
- 전경배(2001), “공식통계의 소지역 추정 활용사례”, 「계간 국민계정」, 제1호, pp.22-36.
- 통계청(2007), 「2006 경제활동인구연보」.
- _____(2006), “소지역 실업통계 작성을 위한 표본규모 검토”, 「통계연구 발표 자료집」, pp.108-140.
- _____(2003), 「소지역 추정기법에 관한 연구」, 국정과제 추진을 위한 교육훈련 결과보고서.
- 홍태경(2002), “소지역 추정을 위한 표본설계”, 성균관대학교 통계학과 석사학위 논문.

- Falorsi P.D., S. Falorsi, and A. Russo(1994), "Empirical Comparison of Small Area Estimation Methods for the Italian Labour Force Survey", *Survey Methodology*, Vol. 20, No.2, pp.171-176.
- IASS 위성 컨퍼런스 2007 SAE, "Contributed Papers", <http://www.dipstat.ec.nipi.it/SAE2007/speak.php>.
- Rao J.N.K.(2003), *Small Area Estimation*, John Wiley & Sons.
- Statistics Canada(2006), "Guide to the Labour Force Survey", Catalogue no. 71-543-GIE.
- _____(2006), "Improvements in 2006 to the LFS", Catalogue no. 71F0031XIE.
- U.S. Department of Labour Bureau of Labor Statistics(2005), "Local Area Unemployment Statistics: Estimation Methodology", <http://www.bls.gov/lau/laumthd.htm>.
- You, Y. and J.N.K. Rao(2000), "Hierarchical Bayes Estimation of Small Area Means Using Multi-Level Models", *Survey Methodology*, Vol.26, pp.173-181.