



## 제2장

# 소지역 소득·빈곤을 추정을 위한 기초연구

권순필 · 박민정

## 제1절 서론

UN의 정의에 따르면, 빈곤은 “기본적으로 선택과 기회에 대한 거부이고, 인간 존엄에 대한 모독이다. 그것은 사회에 효과적으로 참여할 수 있는 기본 능력의 부족을 의미한다”. 이런 의미에서 빈곤은 가장 보편적인 사회정책의 대상이다(오민수, 성은미 등, 2012). 왜냐하면 빈곤과 불평등의 심화는 국가 발전을 저해하는 위해요소이기 때문이다. 그 중에서도 빈곤은 다양한 취약계층을 양산하고 가정의 해체와 같은 사회문제를 야기하며, 결국 공공부조를 비롯한 사회복지 부담을 증가시킴으로써 국가재정운영에도 위협이 될 수 있다(김문길 등, 2012).

우리나라에서는 지난 20여 년 동안 지방자치제도가 안정적으로 정착함에 따라 각종 정책의 시행과 책임이 중앙 정부에서 지방 정부로 이동하였다. 이에 따라 지방 정부에서는 과학적이고 객관적인 정책의 수립과 정책 이행 결과를 확인할 필요가 생겼다. 따라서 지역개발 및 복지정책 수립 구상에 필요한 자료 수요와 지역통계에 대한 필요성은 증가하였다. 빈곤에 대한 통계 역시 그러하다. 그러나 우리나라 전체의 절대적 빈곤율 및 상대적 빈곤율은 통계청과 보건복지부에서 산출하지만, 광역시·도와 기초자치단체인 시군구별 빈곤율은 제시하지 않고 있다. 경기복지재단은 이와 같은 어려움을 해결하고자 경기도 사회조사라는 대규모 표본조사를 시행하여 그 결과를 분석한 「지도로 보는 경기도 빈곤」(오민수, 성은미 등, 2012)이라는 자체 연구 결과를 발간하기도 했다.

빈곤에 대한 지역통계를 생산하기 위해서는 기존 조사통계의 규모를 지역통계 생산이 가능한 수준으로 확대하거나, 여러 기관에서 수집한 행정 자료를 취합하여 정보의 양을 증대시킬 수 있다. 그러나 조사의 규모를 확대하는 방법은 필연적으로 예산과 인력 확보

측면에서 비용의 증가를 수반하고, 응답자들의 응답 부담을 가중시킨다. 그리고 대규모 조사는 시행부터 결과 획득까지 오랜 시간을 요구한다. 한편, 행정 자료는 비통계적 목적으로 수집되기 때문에 요구되는 지역통계의 목적이나 대상과 맞지 않고 목표 모집단에 대한 완전한 커버가 어렵다는 단점이 있다.

“고객은 항상 설계 단계에서 정의된 것보다 더 많은 것을 원한다(Fuller, 1999). 그러므로 우리는 계획되지 않은 소영역을 고려하는 것을 피할 수 없다”(Rao, 2013). 세계적인 통계학자인 Wayne Fuller와 J.N.K. Rao가 한 말이다. 이 말은 전통적인 조사통계가 해결해야 할 많은 숙제들 중에 가장 뜨거운 주제를 표현한 것이다. 즉, 수요자들이 제한된 표본으로 보다 작은 단위의 지역 수준까지 통계를 요구하는 문제를 말한다. 이에 대한 해결 방안으로 학계와 공식 통계 기관들에서는 그 동안 다양한 방법으로 비용 증가없이 조사 설계 당시 계획되지 않은 보다 작은 지역이나 영역에 대한 통계를 생산하는 방안을 연구해왔고, 그 결과 “소지역 추정(Small Area Estimation)”이라는 대안이 제시되었다. “소지역 추정”은 표본설계 단위가 아닌 보다 작은 지역 또는 영역(domain)에 대한 통계를 작성하는 기법 중 하나로, 현장조사(field survey)가 아닌 통계적 추정에 의해 통계를 작성하는 방법을 말한다(김서영, 권순필, 2013). 가장 대표적인 “소지역 추정” 방법은 조사통계와 행정자료 사용의 장점을 모두 살리기 위해 두 정보를 모형으로 결합하여 추정하는 모형기반 추정법이다. 모형기반 추정법은 소지역에서 적은 표본으로 생기는 표집오차를 줄이면서 정도높은 추정통계를 제공할 수 있는 방법이다. 실제로 미국, 영국, 캐나다 등 통계 선진국의 공식통계 생산 기관들에서 이런 “소지역 추정”을 이용한 통계를 연구·생산·공표한 지는 20여 년 이상 되었다. 특히, 이들이 중점적으로 제공하는 소지역 추정 통계는 고용과 소득·빈곤 분야이다.<sup>1)</sup>

본 연구에서는 지역 빈곤 통계 수요 증가라는 외부 환경의 변화에 맞춰 통계청 차원에서 수요에 대응하기 위해 어떤 접근을 시도해야 하는지 검토한다. 특히 국가통계의 생산 주체이면서 정책 수립에 필요한 통계를 적극적으로 생산·제공할 책임이 있는 통계청의 시선으로 필요한 노력을 검토할 것이다. 이를 위해 국내에서 지역 통계를 생산하기 위해 수행했던 노력을 리뷰하고, 일찍부터 소지역 추정을 이용한 소득·빈곤 통계를 생산하고 제공해 온 주요 통계 선진국인 미국, 영국, 유럽의 사례를 중심으로 문헌검토를 실시한다. 이를 통해 얻고자 하는 바는 소득·빈곤을 추정에 i) 어떤 자료를 이용했는가 ii) 어떤 방법을 이용했는가 iii) 예산, 인력, 인구 기간 등의 측면에서 어떤 노력을 기울였는가 iv) 어떻게 결과를 활용하는가에 대한 시사점이다. 실제로 통계청은 소지역의 소득·빈곤

1) 통계개발원에서는 2008년부터 고용 분야에 대해 다양한 소지역 추정 연구를 수행해 왔다. 모형은 Fay & Harriot 모형, Time-Series & Cross-Sectional 모형, 로지스틱선형혼합 모형 등을, 추정량은 EBLUP, HB, EBLUP-type 등을 시험 계산해보았다. 각 모형과 추정 결과에 대해 시계열 안정성, 벤치마킹, MSE 추정 등 여러 측면을 다각적으로 검토하였는데, 구체적인 연구 결과들은 통계개발원 홈페이지(<http://sri.kostat.go.kr>)의 연구결과DB에서 확인할 수 있다.



추정의 중요성을 인식하고, 국가 통계 생산 기관으로서의 역할과 책임을 수행하기 위해 본 연구 주제를 “통계청 중장기 발전 전략 과제”로 선정하였고, 지속 수행할 예정이다.

본 보고서는 제2절은 경기도와 통계청의 소득·빈곤과 관련 지역 통계 생산 사례에 대해, 제3절은 미국, 영국, 유럽의 소지역 소득·빈곤 추정 통계 생산 프로그램에 대해 살펴본다. 제4절에서는 우리나라에서 소지역 추정 통계를 생산하기 위해 사용 가능한 주요정보와 국세청, 연금 자료 등 보조정보를 검토한다. 마지막으로 제5절에는 연구의 결과를 정리하는 것으로 마무리한다.

## 제2절 국내 사례

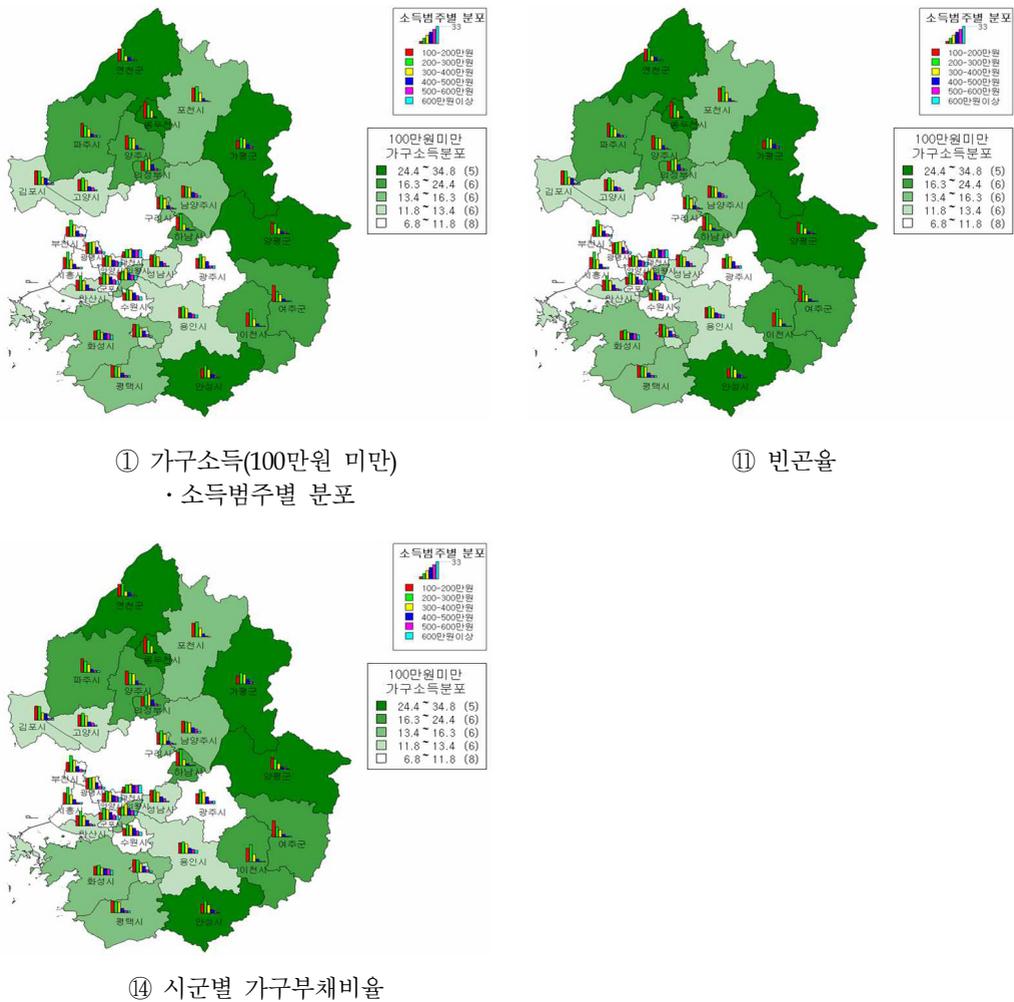
우리나라 전체의 절대적 빈곤율 및 상대적 빈곤율은 통계청과 보건복지부를 통해 산출되지만, 광역시도와 시군구별 자료는 제공되고 있지 않다. 따라서 지역 통계가 필요한 지방 정부는 자체적으로 통계를 생산하려는 시도를 한다. 특히 소득과 빈곤에 대한 우리나라의 지역 통계 생산은 주로 행정 자료의 취합이나 표본 규모의 확대로 이루어져왔다. 이 절에서는 표본 규모의 확대로 지역 통계 생산을 시도한 경기복지재단의 “지도로 보는 경기도 빈곤”과 추정 통계 생산을 시도한 통계청의 “광역시도 소득소비 추정 연구”를 대표적인 국내 지역 통계를 생산하기 위한 노력으로 간주하고 간략히 살펴본다.

### 1. 지도로 보는 경기도 빈곤

“지도로 보는 경기도 빈곤”(오민수, 성은미 등, 2012)은 경기복지재단에서 2012년 발간한 보고서이다. 보고서의 목적은 경기도의 빈곤지도 작성 및 지역별(시군별) 빈곤율과 불균형 정도를 제시하고, 중장기적인 지역개발 및 복지정책 수립 구상에 필요한 환경적 요인들을 탐색하는 것이다. 이를 위해 가독성있는 빈곤지도를 작성하였고, 이를 활용하여 지역별 상대적 빈곤과 빈곤에 영향을 미치는 핵심적 변수들의 분포를 보여준다.

빈곤지도 작성을 위해 사용한 자료는 통계청의 인구동향, 지역별고용조사, 경기도의 경기도민 생활 및 의식조사(2011), 경기도민 복지의식 조사와 정책방향 등이다. 자료는 공간적으로는 경기도의 31개 시군 단위, 시간적으로는 자료 확보 여부에 따라 2010~2011년 자료를 기준으로 산출하였다. 분석 항목은 ① 가구 소득(100만원 미만)·소득범주별 분포, ② 전체·성별실업률, ③ 비임금(자영업), 임시·일용근로자, ④ 4년제 대학교 졸업자비율(30대 이상), ⑤ 고등학교졸업자비율(20세 이상), ⑥ 장애인인구비율, ⑦ 노인인구비율, ⑧ 외국인인구비율, ⑨ 이혼자율, ⑩ 한부모가정, ⑪ 빈곤율, ⑫ 전체 공공부조수급비율, ⑬ 국민기초생활보장, 긴급지원, 무한돌봄수급가구 비율, ⑭ 시군별

가구부채비율, ⑮ 임대주택보급률 및 종류, ⑯ 자살율, ⑰ 경기도 권역별 복지욕구이다. 빈곤지도 제공 결과를 몇 가지만 사례로 살펴보면 [그림 2-1]과 같다. 그림에서 확인하는 바와 같이 지역별 상대적 빈곤과 빈곤에 영향을 미치는 핵심 요인들의 분포를 시각적으로 용이하게 확인할 수 있다. 예컨대, [그림 2-1]을 통해 한 눈에 경기 외곽 지역(농촌 및 도농복합 지역)이 가구소득이 낮은 가구의 비율, 빈곤율, 가구부채비율이 높아 상대적으로 안쪽 지역(도시)들에 비해 빈곤 정도가 심한 것을 알 수 있다.



※ 출처 : 지도로 보는 경기도 빈곤(2012)

※ 자료 : 2011년 경기도민 생활 및 의식조사

[그림 2-1] 경기도 빈곤지도 제공 결과



이와 같은 빈곤지도 작성에 연구자들이 주요정보로 사용한 “경기도민 생활 및 의식조사 (현 경기도 사회조사)”는 12,154 가구의 27천여 15세 이상 가구원을 표본으로 한 조사이다. 이는 통계청 사회조사의 전국 표본 17천여 가구보다 큰 규모로 지방 정부가 조사 통계를 이용하여 지역 통계를 생산하기 위해서는 이와 같은 표본의 확대가 필수적이라는 사실을 보여 준다. 이는 그렇지 않아도 적은 지방 정부의 예산 활용 전략에 적지 않은 부담을 줄 것이다.

## 2. 광역시도 소득소비 추정 연구

전국의 소득소비 수준을 측정하는 가장 대표적인 통계는 통계청의 가계동향조사이다. 가계동향조사는 공표 범위가 도시와 전국이기 때문에 이보다 작은 지역에 대한 통계 수요가 지속적으로 있어왔다. 수요 내용을 보면 다음과 같다(권순필, 김서영, 이정희, 2011).

- 지역별 맞춤형 정책수립(지역위)
- 지역관련 동향지표 부족(산업연)
- 현행 공표 범위로는 지역의 결과 추론에만 이용(지방행정연)
- 16개 시도별로 구분하여 공표 요청(한국은행, 국토연구원, 한국조세연구원, 강원 등)
- GRDP 지출 및 분배 계정 추계 기초 자료 활용(경남, 울산 등)
- 시군구 단위 지역계정의 지출부문 추계에 필요(대구) 등

이처럼 실제로 많은 기관과 지방 정부에서 지역 통계를 필요로 하고 있다. 이러한 수요에 대응하고자 통계청에서는 그 동안 광역시도의 소득소비 추정 연구를 <표 2-1>과 같이 시도했다. 2004년에 처음으로 가계조사의 광역시도에 대한 지역별 추정기법 개발 연구가 시도되었고, 이 연구는 향후 관련 연구의 기초가 된다. 연속적인 연구 수행은 아니었지만, 세 연구 결과가 제시하는 공통적인 시사점은 모형추정은 직접추정에 비해 추정량의 정도(CV) 측면에서 개선이 뚜렷하다는 점이다. 따라서 모형추정을 통한 소지역 통계 생산이 좀 더 성공적이기 위해서는 i) 다양한 모형 검토와 ii) 유용한 보조정보의 탐색이 필요하고 iii) 빈곤율과 같이 추정 방법에 덜 민감한 관심값에 먼저 추정을 시도하여 추정 통계에 대한 이해를 높여야 한다는 것이다. 또한 포괄범위, 추계 방식, 이용자료가 전혀 다름에도 불구하고 유사한 명칭으로 통계가 제공되기 때문에 발생한, 지출 비목에서 광역시도 추정 결과와 GRDP의 항목별 순위 차이 비교와 같은 일이 발생하는 것을 경계해야 한다. 즉, iv) 사용자가 추정 통계를 사용함에 있어 필요한 이런 주의사항을 소지역 추정 통계 제공 시 반드시 함께 알려야 한다.

〈표 2-1〉 광역시도 소득소비 추정 연구 이력(통계청)

---

- 가계조사 지역별 추정기법 개발(김달호와 김재광, 2004)
    - 통계청 발주 용역 연구로 이후 연구의 기초가 됨
    - 모형기반 추정법 제안
      - 모형 : Fay-Herriot 모형(Area-Level)
      - 자료 : 가계동향조사와 국세청, 통계청, 건교부, 노동부 등의 관련 자료
    - 결과
      - 근로소득, 사업소득, 식료품·비주류음료 지출 등 소득소비 주요 항목에 대한 16개 광역시도 모형추정 결과 제시
      - 일부 항목은 직접추정도 가능하나 그 외 항목은 모형추정이 타당
      - 소득소비 분야에서 추정통계 생산의 가능성과 유용성 확인
      - 지출 비목에서 광역시도 추정 결과와 GRDP의 항목별 순위 차이 발생
  
  - 가계조사 지역별 추정기법의 타당성 검증(김혜원과 정동명, 2005)
    - 2004년 연구에 대한 통계청 후속 검증 연구
    - 결과
      - 각 주요 항목의 추정 결과 시계열상 직접추정값과 모형추정값은 큰 차이가 없으나 CV 측면에서는 모형추정값의 개선 정도가 뚜렷
      - 일부 보조정보는 계속 입수에 어려움이 있기 때문에 안정적으로 보조정보를 확보하는 방안 마련 필요
      - 지출 비목에서 광역시도 추정 결과와 GRDP의 항목별 순위 차이는 여전히 발생
  
  - 지역통계 작성을 위한 소지역 추정 방법 연구(권순필, 김서영, 이정희, 2011)
    - 통계개발원 수행 연구
    - 2004년 연구 모형 이용
      - 보조정보 추가 탐색 노력 : 인구, GRDP, 국세, 복지 분야
      - 특히 GRDP 자료는 지출 비목에서 발생하는 광역시도 추정 결과와 GRDP 항목별 순위 차이 발생 정도를 줄여보고자 이용
    - 결과
      - 모형추정은 직접추정의 정도(CV)를 분명하게 개선
      - 모형추정의 성능은 좋은 보조정보 탐색에 의존적
      - 추정 결과와 GRDP 항목별 순위 차이는 포괄범위, 추계방식, 이용자료의 차이로 인해 발생이 불가피하기 때문에 비교 대상이 아님
-



## 제3절 해외 사례

이 절에서는 소득·빈곤율에 대해 통계 선진국인 미국, 영국, 유럽이 수행한 소지역 통계 생산 노력을 문헌을 통해 검토한다. 다양한 통계 생산 방안 중 특히 추정 통계 생산 방법을 살펴본다.

### 1. 미국(SAYPE<sup>2)</sup>)

#### 가. 개요

SAIPE(Small Area Income and Poverty Estimates)은 센서스국에서 모든 학군(school district), 카운티(county)와 주(state)를 위해 소득과 빈곤에 대한 연간 추정값을 제공하는 프로그램이다. 주요 목적은 연방 정부에 적시에 믿을만한 소득과 빈곤 통계를 제공하고 지역정부에 대한 연방 기금을 할당하며 지역 정부 활용 자료를 제공하기 위함이다.

SAIPE가 시작하게 된 계기를 잠시 살펴보면, 미국에서 소득과 빈곤 통계는 역사적으로 센서스에서 나오는 중요한 자료 중의 하나이다. SAIPE 전에는 10년마다 시행되는 센서스 전수조사가 지리적으로 “작은(소)” 지역의 가구와 개인에 대한 소득 분포와 빈곤 통계를 알 수 있는 유일한 원천이었다. 그러나 10년 주기의 자료 공표는 소지역의 경제 상황에 대한 정보를 얻기에는 길다. 그래서 1993년에 미국 센서스국에서는 소지역의 소득과 빈곤 통계 생산을 위해 여러 국가 기관의 자금 지원과 국세청의 자료 제공 동의를 얻어 SAIPE를 시작할 기반을 마련하였다. 1994년에는 미국학교 개선법(Improving America’s Schools Act)이 의회를 통과하면서 법률적 기반까지 확보하였다. 이후 1997년에 처음으로 카운티 수준 소지역 추정 연구 결과가 발표되면서 현재까지 주, 카운티, 학군에 대한 소지역 추정값들이 다양한 방법으로 제공되고 있다.

SAIPE에서 제공한 2011년 추정 결과를 보면 <표 2-2>와 같다. 제공 결과들 중 특히 빈곤 아동에 대한 추정 연구의 가장 큰 후원자는 교육부인데, 교육부는 2006년 당시 약 130억 달러의 기금을 지역에 배분하는데 학군의 빈곤 추정값을 이용했다.

2) SAIPE에 대한 구체적인 내용은 [www.census.gov/did/www/saipe/index.html](http://www.census.gov/did/www/saipe/index.html) 사이트를 이용하여 확인할 수 있다.

〈표 2-2〉 SAIPE 제공 추정 결과

Counties & States	School Districts
all people in poverty	total population
children under age 18 in poverty	children ages 5 to 17
related children ages 5 to 17 in families in poverty	related children ages 5 to 17 in families in poverty
children under age 5 in poverty (for states only)	
median household income	

SAIPE에서 사용하는 빈곤 상태는 가족(family)에 의해 정의된다. 가족 총소득이 빈곤 선보다 작으면 가족 내 모든 구성원이 빈곤 상태에 있다고 하고, 크면 빈곤 상태에 있지 않다고 정의한다. 빈곤을 결정하는 빈곤선은 전국적으로 동일한 기준을 사용하는데 빈곤선에 영향을 미치는 주요 요인은 가족의 크기, 가구원의 연령이다. 이에 따라 빈곤 가구(가구원)를 정의하기 위해 사용 가능한 빈곤선은 모두 48개<sup>3)</sup>이다.

## 나. 추정

SAIPE는 주, 카운티와 학군별로 조금씩 다른 자료를 가지고 조금은 다른 추정방법을 이용한다. 주와 카운티 수준에서는 조사자료와 인구 추정값, 행정자료를 결합한 모형 기반 추정값을, 학군 수준에서는 모형에 기반하여 계산된 카운티의 빈곤·소득 추정값과 연방 세금 정보, 다년(multi-year) 조사 자료를 이용하여 빈곤과 소득에 대한 합성 추정값을 생산한다. 주와 카운티 수준에서 사용하는 모형은 Fay & Herriot이 1979년에 제안한 FH 모형이고, 학군은 이용 가능한 정보가 많지 않기 때문에 카운티 수준의 추정값을 합성 분배(Synthetic-Share)하는 방식을 사용한다.

추정에 이용한 조사 자료는 2005년 이전에는 현재인구조사(Current Population Survey; CPS)의 연간사회경제보조조사(Annual Social and Economic Supplement Survey; ASEC)였는데, 2005년부터는 미국지역사회조사(American Community Survey; ACS)로 바뀌었다. ACS는 인구 6만 5천 명 이상인 지역에 대해서는 단년도 직접추정값을, 인구 2만 명 이상인

3) 48개의 빈곤선은 미국 센서스국의 Poverty 관련 홈페이지 ([www.census.gov/hhes/www/poverty/about/overview/measure.html](http://www.census.gov/hhes/www/poverty/about/overview/measure.html))에서 확인할 수 있다.



지역에 대해서는 3개년 누적 자료를 이용, 모든 카운티와 학군에 대해서는 5개년 누적 자료를 이용해 빈곤과 소득에 대한 직접추정값을 제공한다. 이러한 ACS의 소지역에 대한 통계 제공에도 불구하고 SAIPE는 매년 모든 주, 카운티, 학군에 대해서 모형기반 추정값을 계속 생산하는데, 이는 모형화(modeling)가 표집오차를 줄여주는 추정값을 생산하기 때문이다.

SAIPE에서 주와 카운티 수준 추정을 위해 사용하는 모형은 FH 모형이다. 보다 명확한 이해를 위해 카운티 수준에서 취학연령(5~17세) 아동의 빈곤 추정을 위한 다음의 FH 모형을 기술한다(Bell 등, 2007).

$$\log(y_i) = \log(Y_i) + e_i, \quad e_i \sim \text{ind. } N(0, \nu_i) \quad (2.1)$$

$$\log(Y_i) = x_i' \beta + u_i, \quad u_i \sim \text{i.i.d. } N(0, \sigma_u^2) \quad (2.2)$$

여기서,  $i$  : 카운티,

$y_i$  : ACS의 5~17세 조사 빈곤 추정값 (빈곤인구 혹은 빈곤율),

$Y_i$  : 실제 5~17세 빈곤값 (빈곤인구 혹은 빈곤율),

$e_i = \log(y_i) - \log(Y_i)$  :  $\log(y_i)$ 의 표집오차,

$x_i$  : 회귀 변수들의 벡터,

$\beta$  : 회귀 계수들의 벡터,

$u_i$  : 랜덤 모형 오차 (카운티 랜덤 효과) 이다.

식 (2.1)과 (2.2)의  $\beta$ ,  $\sigma_u^2$ 은 최대우도추정(maximum likelihood)법에 따라 추정하는데,  $\sigma_u^2$ 이 주어졌을 때  $\beta$ 의 최대우도추정량은 가중최소제곱회귀(weighted least squares regression)에 의해 다음과 같은 형태를 갖는다.

$$\hat{\beta} = (X' V^{-1} X)^{-1} X' V^{-1} \mathbf{y} \quad (2.3)$$

$$\text{Var}(\hat{\beta}) = (X' V^{-1} X)^{-1} \quad (2.4)$$

여기서,  $X$  :  $x_i'$ 의 행렬,

$\mathbf{y} = (y_1, \dots, y_m)'$  : 모든  $m$ 개 카운티의 ACS 조사 추정값,

$V$  : 총 분산  $\sigma_u^2 + \nu_i$ 를 원소로 갖는 대각행렬이다.

우도의 최대화는 주어진  $\sigma_u^2$ 에 대해 가중최소제곱으로  $\beta$ 를 계산하는 것과 주어진  $\beta$ 에 대해  $\sigma_u^2$ 의 우도를 최대화하는 작업의 반복을 통해 얻어질 수 있다.

이와 같은 과정을 통해 얻어지는  $\log(Y_i)$ 의 경험적최량예측(Empirical Best Prediction)은 다음의 축소 추정량(shrinkage estimates) 형태로 나타난다.

$$\log(\widehat{Y}_i) = (1 - w_i)\log(y_i) + w_i(x_i'\widehat{\beta}) \quad (2.5)$$

여기서,  $w_i = \frac{\nu_i}{\sigma_u^2 + \nu_i}$ 이다.

$$\text{Var}[\log(Y_i) - \log(\widehat{Y}_i)] = w_i\sigma_u^2 + w_i^2(x_i'\text{Var}(\widehat{\beta})x_i) \quad (2.6)$$

식 (2.5)를 보면  $\log(\widehat{Y}_i)$ 는 직접추정값  $\log(y_i)$ 와 회귀예측값  $x_i'\widehat{\beta}$ 의 가중 결합의 형태인 것을 알 수 있다. 식 (2.6)에서는 모형 오차 분산  $\sigma_u^2$ 을 이미 알고 있는 값으로 취급하는데, 이 값을 구하기 위한 근사적인 접근법은 Datta와 Lahiri의 1995년 연구를 통해 확인할 수 있다.

최종 관심값  $Y_i$ 는 식 (2.5)의 추정 결과를 원래의 스케일로 환원함으로써 다음의 형태로 구할 수 있다.

$$\widehat{Y}_i = \exp(\log(\widehat{Y}_i)) \exp[\text{Var}(\log(Y_i) - \log(\widehat{Y}_i))/2] \quad (2.7)$$

$$\text{Var}(\widehat{Y}_i) = \widehat{Y}_i^2 [\exp(\text{Var}[\log(Y_i) - \log(\widehat{Y}_i)]) - 1] \quad (2.8)$$

$y_i = 0$ 인 일부 카운티는  $\log$  변환값을 얻을 수 없기 때문에 회귀 예측에서 제외한다. 이들 지역에 대해서는 순수한 회귀 예측 결과인  $\log(\widehat{Y}_i) = x_i'\widehat{\beta}$ 를 추정값으로 사용한다. 즉, 식 (2.5)와 (2.6)에서 각각  $w_i$ 를 1로 주는 것이다.

학군 수준에서의 취학 연령 빈곤 아동 추정은 주와 카운티와 달리 합성 분배 방식을 이용한다. 이는 학군 수준에서 이용 가능한 정보가 많지 않기 때문이다. 학군의 빈곤 아동 추정을 위해서는 우선 국세청으로부터 얻은 아동 면세자수(child tax exemptions) 정보로부터 해당 학군의 카운티 대비 세금 기반 아동 빈곤율을 계산하고 이를 카운티 수준에서 얻어진 모형기반 빈곤 추정값에 곱하여 계산한다. 계산식을 간단히 표현하면 다음과 같다(Maples와 Bell, 2007).

$$POV-\widehat{TAX}_{ij} = \frac{T_{ij}^{poor}}{\sum_j T_{ij}^{poor}} \times county-\widehat{POV}_i \quad (2.9)$$

여기서,  $i$  : 카운티,

$j$  : 학군(혹은 학군 조각),

$POV-\widehat{TAX}_{ij}$  : 학군의 빈곤 아동 추정값,

$T_{ij}^{poor}$  : 국세청의 면세자수 정보로부터 계산된 세금 기반 빈곤 아동 추정 인구,

$county-\widehat{POV}_i$  : SAIPE로부터 계산된 모형기반 빈곤 아동 추정 인구이다.

이상과 같이 SAIPE에서 주, 카운티, 학군에서 소지역의 빈곤 인구 추정 방법을 살펴보았는데, <표 2-3>에 이 내용을 간단히 정리해보았다. 주와 카운티는 동일한 모형을 이용하지만, 주 수준에서는 빈곤인구를 추정하기 위해 4개의 연령대별로 추정 빈곤율을 계산하고 추정된 빈곤율에 추계 인구를 곱하여 빈곤인구를 산정한다. 반면에 카운티 수준에서는 빈곤율이 아닌 빈곤인구를 모형을 통해 추정한다. 이는 식 (2.1)과 (2.2)에서  $y_i$ 값으로 빈곤인구나 빈곤율을 쓰는 것으로 추정할 수 있다.  $y_i$ 의 형태에 따라 보조정보  $x_i$  역시 비율 정보를 쓰거나 인구 정보를 쓴다. 또한, 주와 카운티 모형은 모두 연령대별 모형 혹은 관심값에 따라 사용하는 보조정보나 보조정보의 형태가 조금씩 다르다. 예를 들어, 주 수준에서 사용되는 SSI(Supplemental Security Income, 보충적 소득 보장) 수급율 자료는 65세 이상 연령대 모형에서만 이용되고 타 연령대 모형에서는 사용하지 않는 보조정보이고, 카운티 수준의 빈곤 인구 추정에는 총 거주 인구 자료가 사용되지만, 18세 이하 빈곤인구 추정에서는 18세 이하 거주인구 자료가 사용된다.

또한, 주, 카운티, 학군의 빈곤 추정 결과는 보다 상위 지역의 값과 일치되도록 벤치마킹을 적용한다. 즉, 카운티의 빈곤인구 합계는 상위 지역 수준인 주의 빈곤인구와 같고, 주의 빈곤인구 합계는 전국의 빈곤인구 합계와 같도록 하는 것이다. 다만 가구 소득 중앙값에 대해서는 벤치마킹을 적용하지 않는다.

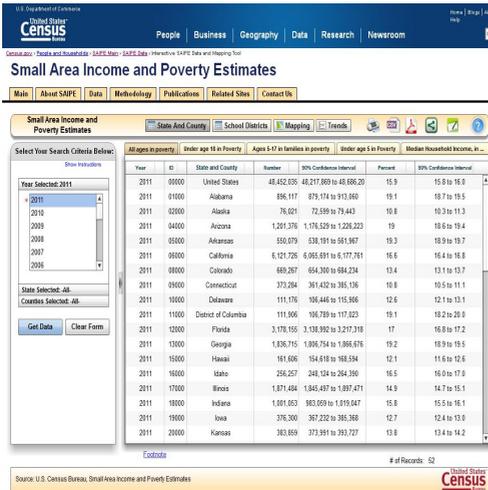


〈표 2-3〉 주, 카운티, 학교의 소득과 빈곤 추정방법

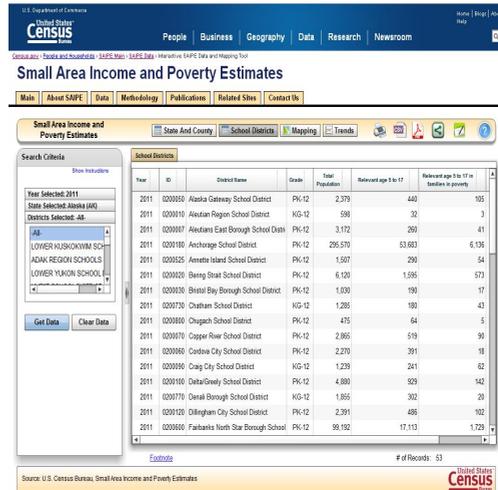
	주(State)	카운티(Counties)	학교(School Districts)
추정 방법	<ul style="list-style-type: none"> <li>모형 : FH 모형</li> <li>추정량 : 모형추정량과 ACS 직접추정량 가중 결합</li> </ul>	<ul style="list-style-type: none"> <li>Empirical Bayes 추정량</li> <li>빈곤율이 아닌 빈곤인구 추정</li> </ul>	<ul style="list-style-type: none"> <li>합성 분배 방식</li> </ul>
이용 정보	<ul style="list-style-type: none"> <li>주요정보 : ACS 직접추정량</li> <li>빈곤 : tax return poverty rate, nonfiler rate, SNAP 참여율, SSI 수급율, 2000년 센서스 65세 이상 빈곤율 등</li> <li>소득 : AGI(보정된 소득 중앙값), 2000년 센서스 자료, 1999 IRS 소득 중앙값 등</li> </ul> <p>※ SNAP : Food Stamp Program, AGI : median Adjusted Gross Income</p>	<ul style="list-style-type: none"> <li>주요정보 : ACS 단년 직접추정량</li> <li>빈곤 : 세금 환급 공제자수, SNAP 혜택자수, 추계인구, 2000년 센서스 인구 등</li> <li>소득 : 2000년 센서스 카운티 가구 소득 중앙값, AGI, BEA의 개인 소득 추정량, nonfiler rate 등</li> </ul> <p>※ BEA : 경제분석국</p>	<ul style="list-style-type: none"> <li>ACS 단년 추정값(2006~2010), 통합된 세금 자료(면세자 정보), 카운티의 모형기반 빈곤 추정값</li> </ul>
벤치 마킹	<ul style="list-style-type: none"> <li>상위 지역 수준 결과에 적용</li> <li>가구소득 중앙값은 벤치마킹 적용하지 않음</li> </ul>		<ul style="list-style-type: none"> <li>상위 지역 수준 결과에 적용(raking)</li> </ul>
비고	<ul style="list-style-type: none"> <li>추정 결과에 대한 90% 신뢰구간 제공</li> <li>각 연령대별 모형 혹은 관심값에 따라 사용하는 보조정보가 조금씩 다름</li> </ul>	<ul style="list-style-type: none"> <li>직접추정값이 0인 경우는 모형에서 제외 하지만 예측 결과는 제공</li> <li>모든 변수는 로그 변환하여 사용</li> </ul>	<ul style="list-style-type: none"> <li>추정 결과에 대한 90% 신뢰구간 미제공</li> <li>낙오 학생방지법의 이행(보조금 지급)을 위해 사용</li> <li>2년마다 갱신되는 학교 경제에 맞춰 추정값도 재계산</li> <li>다양한 인구규모(중간값 7천 5백 명, 평균 2만 5천여 명, 학교의 약 99%가 6만 5천 명 이하)</li> </ul>

## 다. 소지역 추정 결과와 관련 정보 제공

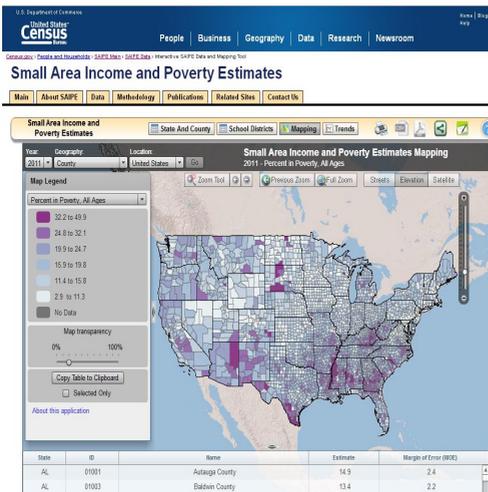
SAIPE는 소득·빈곤에 대한 <표 2-2>와 같은 다양한 소지역 추정 결과를 홈페이지에 제공하고 있다. 제공 형태는 표, 지도, 추세선 등이다([그림 2-2]). 주와 카운티 수준의



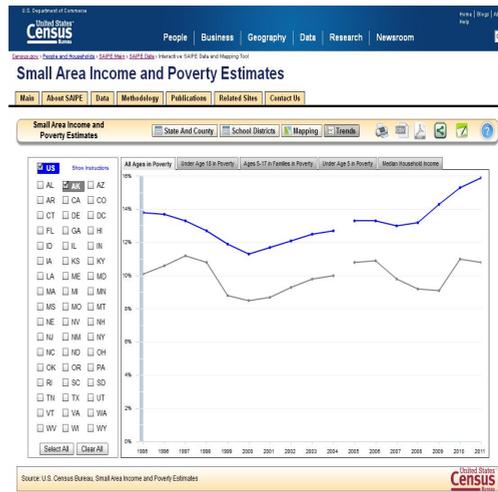
① States & Counties



② School districts



③ mapping

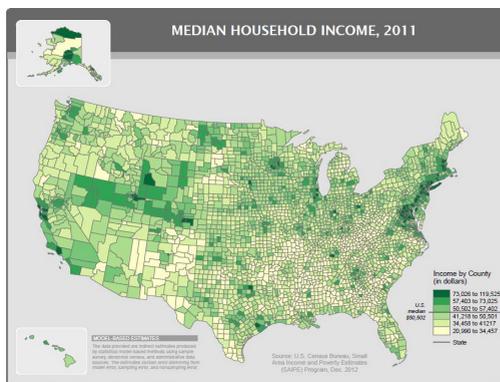


④ Trends

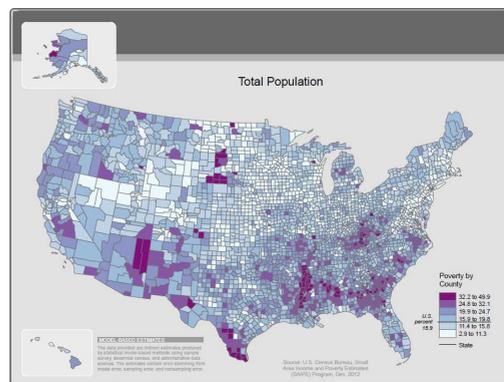
※ 출처 : 미국 센서스국 SAIPE 홈페이지

[그림 2-2] SAIPE 소지역 추정 결과 제공

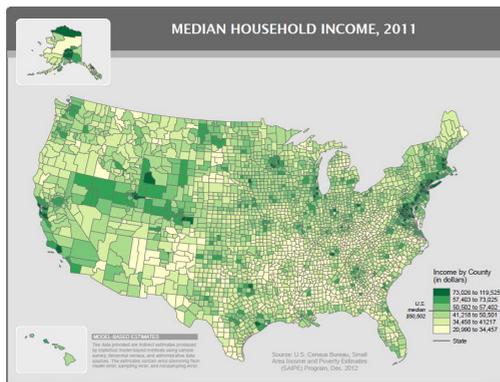
결과에 대해서는 빈곤인구와 빈곤율에 대한 점추정값과 90% 신뢰구간을 [그림 2-2]의 ①과 ②처럼 표(table) 형태로 제공하고, 표 형태와 다르게 한 눈에 미국 전역의 지역별 소득과 빈곤 정보를 확인할 수 있도록 [그림 2-2]의 ③과 [그림 2-3]처럼 지도를 제공하고 있다. 또한 [그림 2-2]의 ④처럼 지난 기간 값들의 변화를 추세 그래프를 통해서도 보여주고 있는데, 이런 시도를 통해 통계 이용자의 편의를 도모하고 통계의 활용성을 증대시키고 있다.



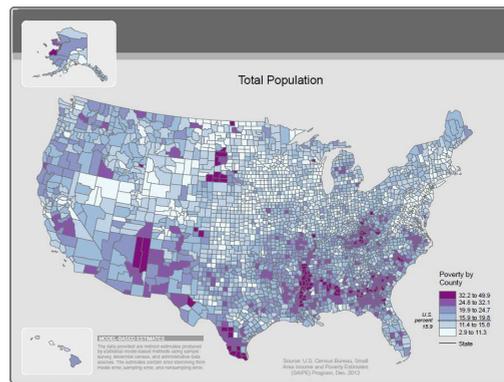
① 카운티별 가구 소득 중앙값



② 카운티별 빈곤 인구 비율



③ 카운티별 가구 소득 중앙값 변화 (2007~2011)



④ 학군별 빈곤 취학 아동 비율

※ 출처 : 미국 센서스국(2012)

[그림 2-3] 카운티와 학군별 소지역 추정 mapping 결과



## 2. 영국(SAEP)

### 가. 개요<sup>4)</sup>

소지역 추정에 관한 각종 보고서들을 살펴보면, 영국 통계청(Office for National Statistics; ONS)은 소지역 추정에 관해 공식 통계기관으로서 선도적 역할을 담당해온 것으로 보인다. 이러한 선도적 역할 수행은 학계와의 꾸준한 교류와 소지역 연구에 대한 공감대 확산을 바탕으로 이루어져 왔다. 특히 1985년(캐나다 오타와)과 1992년(폴란드)의 소지역 관련 국제 심포지엄을 계기로 OPCS(Office for Population Census and Surveys, ONS의 전신)는 1992년 Southampton 대학에 의뢰하여 소지역 연구의 잠재성을 탐색하여 보고서를 발간했다. 또한 1995년에서 1997년 사이에는 여러 가지 모형에 관한 실험적인 공동 연구들을 수행하였다.

이러한 연구 성과에 기반하여, 1998년 ONS는 본격적으로 SAEP(Small Area Estimation Project)를 시작했다. SAEP는 2001년까지 약 4년간 수행되었고, 2003년 최종보고서가 발간되었다. 이 프로젝트의 목표는 사회통계와 관련된 소지역 추정문제에 대해 일반적인 통계적 방법론을 개발하고 시스템을 구축하는 것이었다. 방법론을 일반적으로 개발한다는 것은 소지역을 규정하는 다양한 지역경계에 활용될 수 있도록 한다는 것을 의미하며, 또한 각 분야별로 소지역 추정을 꾸준히 수행해 나가는 기본 방법론을 구축한다는 것도 뜻하고 있다. SAEP 이후, 이를 활용한 영국 통계청의 분야별 소지역 추정 결과 보고서들을 보면, SAEP의 목표는 충분히 달성된 것으로 평가할 수 있다.

SAEP와 관련하여 ONS의 소지역 추정 연구를 분야별로 살펴보면 다음과 같다. 우선 인구 분야에서, 2001 센서스를 대비하여 1998년에서 2000년 사이에 소지역 인구 추정에 관한 실험적인 연구가 있었다. 이를 기반으로 2001년 센서스에 대해 소지역 인구 추정을 구현하는 작업이 2000년에서 2003년에 이루어졌다. 이는 원번호 센서스를 하기 위해 나이-성별 그룹별 지역 수준의 인구 추정값을 보정하는데 이용되었다.

실업과 소득에 관한 소지역 추정 연구의 간략한 개요는 다음과 같다. 먼저 2000년에서 2004년 사이에 2001년 센서스를 활용한 실업과 소득에 관한 소지역 추정연구가 있었다. 이를 바탕으로, SAEP의 최종 보고서가 발간된 이후인 2003년에서 2008년 사이에 실업과 소득에 대한 소지역 추정 구현 작업이 이루어졌다. 2008년 이후 실업의 소지역 추정통계는 공식 통계로 사용되고 있고, 소득에 관한 소지역 추정 통계는 2010년 공식 통계로 승인하는 것을 추진 중인 것으로 확인되었다. 소득을 이용한 빈곤 관련 소지역 추정 연구는 소득과 같은 시기에 연구되어 실험 통계로 공표되고

4) Silva와 Clarke(2008), ONS(2001) 참조

있다. <표 2-4>는 영국의 소지역 추정과 관련된 연구 내용들의 진행상황을 간략히 정리해 보여준다.

지금까지 간략히 설명된 ONS의 소지역 추정 연구에서 큰 변화는 주로 영국 행정구역 개편과 밀접하게 연관되어 일어났다. 또한 연구 대상이 되는 소지역 단위도 인구, 실업, 소득, 빈곤 등 각 분야별 소지역 추정 연구에서 동일하지 않다. 예를 들면, 실업 분야의 소지역 추정은 LAD<sup>5)</sup>단위로 주로 이루어졌으나, 본 연구의 주요 관심 분야인 소득과 빈곤에 관해서는 ward<sup>5)</sup>와 MSOA<sup>5)</sup>가 주요 소지역 단위였다. 다음 절에서 소지역 추정 연구와 관련하여 영국의 행정구역 및 통계구역에 관한 내용을 우선 살펴본 후, 계속해서 소지역 추정과 관련된 일반적 방법론을 개발한 프로젝트인 SAEP를 바탕으로 한 영국 ONS의 소득과 빈곤 분야의 소지역 추정에 대해 자세히 살펴보기로 한다.

<표 2-4> 영국 소지역 추정 연구 진행 이력

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
SAEP	SAEP 수행 (98~01)					보고서					
인구	실험적 연구들			소지역 추정 구현							
실업	2001년 센서스를 활용한 실험적 연구들					소지역 추정 구현 (UA/LAD 단위)					
소득	2001년 센서스를 활용한 실험적 연구들					소지역 추정 구현 (MSOA 단위)					
빈곤						소지역 추정 구현 (MSOA 단위)					

## 나. 지리 체계

영국은 크게 잉글랜드, 웨일즈, 스코틀랜드 및 북아일랜드 네 지역으로 구성되어 있다. 이 중 잉글랜드, 웨일즈, 스코틀랜드는 GB(Great Britain)를 구성하여 지역 구분에서 비교적 유사한 특징을 보이는 경우도 있다. 영국의 주요 행정구역명은 각 지역별로 다른데, 잉글랜드는 county(83개), 웨일즈는 UA(Unitary Authority, 22개), 스코틀랜드는

5) 「2 영국 (SAEP)의 나. 지리 체계」 참조



council(32개), 북아일랜드는 district(26개)가 각각 주요 행정구역이다(<표 2-5>). 주요 행정구역은 또 다시 잉글랜드의 경우 83개 county는 27개 non-metropolitan과 56개 metropolitan county로 구분되며, 상위로는 9개 region으로 묶인다. 또한, metropolitan county는 32개 London borough, 36개 metro-district, 55개 UA 및 2개의 Sui generis로 더 나뉘고 있다. 이처럼 영국의 행정구역 구분은 매우 복잡하다. 게다가 영국은 행정 목적별로 행정구역의 체계가 다른 특징도 가지고 있다. 예를 들면, 교육과 주거를 위한 행정구역명이 서로 다른 지역들도 많이 존재한다. 다만, 각각의 구분이 모두 일치할 경우 이를 UA라고 한다.

<표 2-5> 영국의 주요 행정구역명 및 개수

영국	잉글랜드	웨일즈	스코틀랜드	북아일랜드
주요 행정구역명	county (83)	UA (22)	council (32)	district (26)

소지역 추정에 사용되는 소지역 단위들은 위에 설명된 영국의 일반적인 행정구역들의 하위 수준에 포함되는 Postcode Units and Sectors(PU and PS), wards, Local Authorities Districts(LAD) 등이 있고, 의회의 선거용 구분 지역인 Parliamentary Constituencies(PC)도 있다. 소지역 추정 문제에 있어서 이런 복잡한 행정구역은 지속적으로 대처하기 어려운 수요들을 양산하게 된다. 예를 들어 wards 단위의 소지역 추정 수요가 있어 통계를 작성하면 다시 PC 단위의 소지역 추정 수요가 추가로 발생하는 경우 등의 사례가 있었다.

영국은 이러한 다양한 지역 단위들에 대한 연구와 더불어 일관된 크기와 변하지 않는 경계를 가지면서 투표나 행정적인 이유로 변경되지 않는 새로운 지역 단위 결정에 관한 연구들을 수행해왔다. 다양한 행정구역들이나 의회 선거 구역에 대해 통계 생산을 지속하는 대신, ONS는 각종 지역 단위에 대한 연구들을 바탕으로 새로운 통계 지역 단위를 결정하는 노력을 지속했다. 먼저 Output Areas(OA)는 2001년 센서스에서 사용된 단위이며, 이를 바탕으로 한 Super Output Area(SOA)는 잉글랜드/웨일즈 지역에서 소지역 통계 생산을 향상시키고자 설계된 것이다. 이 SOA들을 결합한 Lower layer SOA(LSOA)가 잠시 고려되었고, 최종적으로는 Middle layer SOA(MSOA)를 소지역 추정의 주요 지역 단위로 사용하게 되었다. MSOA는 2001년 센서스 결과에 사용된 2003년 LAD 경계에 제한된 LSOA들의 그룹이다. 영국의 소지역 통계 생산의 주요 단위인 MSOA의 평균 인구는 7,200명, 최저 인구는 5,000명이다.

역사와 전통이 다양하고 복잡하기는 유럽도 마찬가지여서, 유럽 역시 통일된 지역 구분에 대한 연구를 통해 NUTS<sup>6)</sup>라는 지역 단위를 만들었다. NUTS1이 그 중 가장 큰

상위 지역을 의미하고 숫자가 증가하면서 하위 소지역을 의미하게 되는데, 이런 유럽의 지리 체계를 영국에 적용했을 때 영국 각 지역의 NUTS 개수는 <표 2-6>과 같다.

<표 2-6> 영국의 NUTS 구성

EU기준	잉글랜드	웨일즈	스코틀랜드	북아일랜드
NUTS1	9 (region)	1	1	1
NUTS2	30	2	4	1
NUTS3	93	12	23	5

#### 다. 소지역 추정

영국의 소지역 추정은 모두 SAEP의 방법론을 근거로 하며, 자료의 특성에 따라 SAEP 기본 방법론을 조금씩 변형시킨 모형을 이용해 이루어진다. 본 연구는 소득과 빈곤에 관한 소지역 추정에 관해 살펴보는 것이 주요 목적이거나, 영국의 경우는 하나의 일반적인 방법론을 근거로 각 분야별로 방법론을 변형·적용하는 방향으로 연구가 이루어져 왔으므로, 비교를 위해 실업 분야 소지역 추정의 역사와 내용을 먼저 간략히 정리해 본다.

실업과 관련된 소지역 추정 연구는 주로 UA나 LAD 수준에서 이루어져 왔다. 사용된 주요 자료는 노동력조사(Labor Force Survey) 자료와 보조금 청구횟수 행정자료이다. 이 자료들을 가지고 먼저 여러 방법론들이 검토되었다. 먼저 Structure Preserving Estimation(SPREE)은 성과가 좋지 않았고, 시계열모형은 자료의 기간과 크기가 적합하지 않아 선택되지 않았다. 마지막으로 회귀 모형인 일반화선형혼합 모형을 사용해 고정 효과 설정에 의한 합성추정량 및 랜덤 효과 설정에 의한 EBLUP-type 추정량이 적절한 추정량으로 제시되었다.

이는 보조금 청구횟수라는 보조정보의 실업에 대한 높은 설명력을 활용하여 적합한 모형을 찾은 것으로 추정의 결과가 성공적이었다. 실업 분야의 소지역 추정 성공은 다른 분야의 소지역 추정 연구에 동기를 부여하였다. 또한 실업에 대한 소지역 추정 모형은 우리나라에도 도입되어 통계개발원의 고용·실업 관련 소지역 추정 연구에 큰 기여를 하였다.

영국의 소지역 추정에서는 분야별로 사용한 보조정보의 종류와 기본 소지역 단위의 변화가 주요 쟁점이었다. <표 2-7>은 소지역 추정 연구가 활발하였던 실업, 소득, 빈곤 분야의 소지역 단위, 주요 보조정보, 사용 모형 및 승인여부를 비교하여 정리하였다.

6) 「3. 유럽연합 (SAMPLE)의 나. 지리 체계」 참조



실업은 LAD를 기반으로 통계가 생산되었고, 소득과 빈곤은 ward를 기반으로 추정 연구가 이루어지다가 MSOA를 기반으로 모형을 확정하여 통계 생산을 계속하고 있다.

〈표 2-7〉 영국의 분야별 소지역 추정 현황

	실업	소득	빈곤
소지역 단위	LAD	ward(1998/99, 2001/02) MSOA(2004/05, 2007/08)	ward(1998/99) MSOA(2004/05, 2007/08)
조사자료	노동력조사	가족자원조사	가족자원조사
보조자료	보조금 청구 횟수	센서스 및 각종 행정자료	센서스 및 각종 행정자료
모형	Logistic Mixed 모형	Linear Random Effect 모형	Logistic Mixed 모형
승인여부	2008년 이후 승인통계로 공표	2010년 실험 →승인 계획 중	실험통계로 공표
기타	2009 PC 단위 소지역 추정 (실험통계로 공표)		

이제 본 연구의 주요 관심사인 소득에 관한 소지역 추정의 흐름을 구체적으로 정리 하면 <표 2-8>과 같다. 먼저 스코틀랜드의 소지역을 대상으로 소득의 소지역 추정이

〈표 2-8〉 영국의 소지역 소득 추정 연구 흐름

시기	1998/99	2001/02	2004/05	2007/08
지역	스코틀랜드	잉글랜드/웨일즈	잉글랜드/웨일즈	잉글랜드/웨일즈
소지역 단위	ward (96 wards)	ward (2003 CAS wards)	MSOA (CAS보다 장점 많음)	MSOA (2004/05와 동일)
조사 자료	98/99 가족자원조사	01/02 가족자원조사	04/05 가족자원조사	07/08 가족자원조사
보조 자료	91년 센서스 98/99 보조금 청구횟수 02 의회 과세구분 자료 98/99 내국세무청 자료	01년 센서스 01/02 보조금 청구횟수 02 의회 과세 자료 지역구분지표 01년 부동산 주거비	01년 센서스 04년 보조금 청구횟수 04 의회 과세 자료 지역구분지표 02/05 부동산 주거비 03/04 HMRC 소득자료	01년 센서스 07년 보조금 청구횟수 07 VOA 의회 과세구분 지역구분지표 07 주거비 소유변화 07 HMRC 소득자료

※ 주 : 1. 1998/99 잉글랜드/웨일즈의 소득 소지역 추정은 1998 ward 경계를 사용하여 2001/02 결과와 직접 비교하기 어렵고, 혼란을 방지하기 위해 자료를 삭제하고 2001/02 추정결과만 공표하였다.

2. CAS: Census Area Statistics, HMRC: Her Majesty's Revenue and Customs service

있었고, 이어 잉글랜드/웨일즈 지역의 소지역 추정 연구가 있었다. 이 둘은 모두 ward 단위를 기반으로 연구가 수행되었으나 이후 ward를 대신하여 MSOA를 중심으로 통계 단위가 재설정되면서 MSOA를 기반으로 소득관련 소지역 추정 연구가 다시 수행된다. 연구가 지속되면서 2007/08 자료를 대상으로 한 소지역 추정 모형에서는 보조정보 선택에서 변화를 보이고 있다.

본 연구의 또 하나의 관심사인 빈곤의 소지역 추정을 정리하면 아래 <표 2-9>와 같다. 빈곤은 소득을 이용해 계산이 되므로, 이 표에서는 소득과 연결하여 주요정보를 정리하였다. 보조정보를 살펴보면 소득의 소지역 추정에 이용된 정보를 그대로 사용하는 것을 알 수 있다. 이는 빈곤의 소지역 추정 모형을 보면 왜 그런지 알 수 있는데, 다음 절에서 소득과 빈곤을 소지역 추정의 구체적인 내용을 살펴보도록 한다.

<표 2-9> 영국의 소지역 소득·빈곤 추정 연구 흐름

	소득	빈곤	소득	빈곤
시기	1998/99	1998/99	2007/08	2007/08
지역	스코틀랜드	잉글랜드/웨일즈	잉글랜드/웨일즈	잉글랜드/웨일즈
소지역 단위	ward (96 wards)	ward (98 wards)	MSOA	MSOA
조사 자료	98/99 가족자원조사	98/99 가족자원조사	07/08 가족자원조사	07/08 가족자원조사
보조 자료	91년 센서스 98/99 보조금 청구횟수 02 의회 과세구분 자료 98/99 내국세무청 자료	91년 센서스 98/99 보조금 청구횟수 98/99 내국세무청 자료 지역구분지표	01년 센서스 07 보조금 청구횟수 07 VOA 의회 과세구분 지역구분지표 07 주거비 소유변화 07 HMRC 소득자료	01년 센서스 07 보조금 청구횟수 07 VOA 의회 과세구분 지역구분지표 07 주거비 소유변화 06 HMRC 세금신용자료

## 라. 소득과 빈곤율의 추정 사례<sup>7)</sup>

이제 소득에 관한 ONS의 소지역 추정에 관해 살펴보기로 한다. 소득에 관한 소지역 추정은 2001년 센서스에서 응답자에게 민감한 소득에 관련된 질문을 제외하기로 결정한

7) Model-based ward level estimates of household income(ONS, 1998/99 User guide for Scottish estimates), Model-based estimates of income for wards in England and Wales(ONS, 2001/02 User guide), Model-based estimates of income for middle layer super output areas in England and Wales(ONS, 2004/05 User guide), White 등(2009), Longhurst(2004) 참고



것이 계기가 되어 시작하게 되었다. 추정에 고려된 소득은 주당 가구소득이며 세부적으로 아래와 같은 총 4가지 종류가 있다.

- 주당 가구총소득
- 주당 가구순소득
- 주거비 사용 전 주당 균등화 가구순소득(BHC)
- 주거비 사용 후 주당 균등화 가구순소득(AHC)

총소득과 순소득을 각각 소지역 추정 대상으로 삼고, 순소득에 대해서 다시 주거비 사용전후의 균등화 소득을 고려하였다. 여기서 균등화 가구소득은 가구원수 및 종류에 따라 가중값을 주어(McClement's scale) 계산한 것으로 가구소득을 가구원수의 제곱근 값으로 나눈 우리나라 가계동향조사의 균등화 가구소득과는 정의가 다르다.

위의 네 종류 소득에 대하여 ONS는 모형기반 소득 추정값들을 생산하였다. 추정 과정으로는 먼저 주당 가구소득과 보조정보(센서스 및 행정자료) 간의 관계를 모형화 하였다. 다음으로 상위지역의 공표된 조사 자료값과 일치성이 유지되도록 벤치마킹 과정을 두어 소지역통계량을 추정하였다. 수리적 모형은 SAEP 기본 모형으로부터 다음과 같다.

$$\log(y_{ij}) = \alpha + \beta \bar{X}_{(k)ij} + u_j + e_{ij} \quad (2.10)$$

여기서,  $y_{ij}$  : 표본지역(예를 들면 PC)  $j$ 내의 가구  $i$ 에 대한 주당 소득,

$\bar{X}_{(k)ij}$  : 표본지역  $j$ 내의 가구  $i$ 가 속한 소지역  
(예를 들면 ward)  $k$ 에서 보조정보의 평균,

$u_j$  : 지역 수준 잔차로 평균은 0, 분산은  $\sigma_u^2$ ,

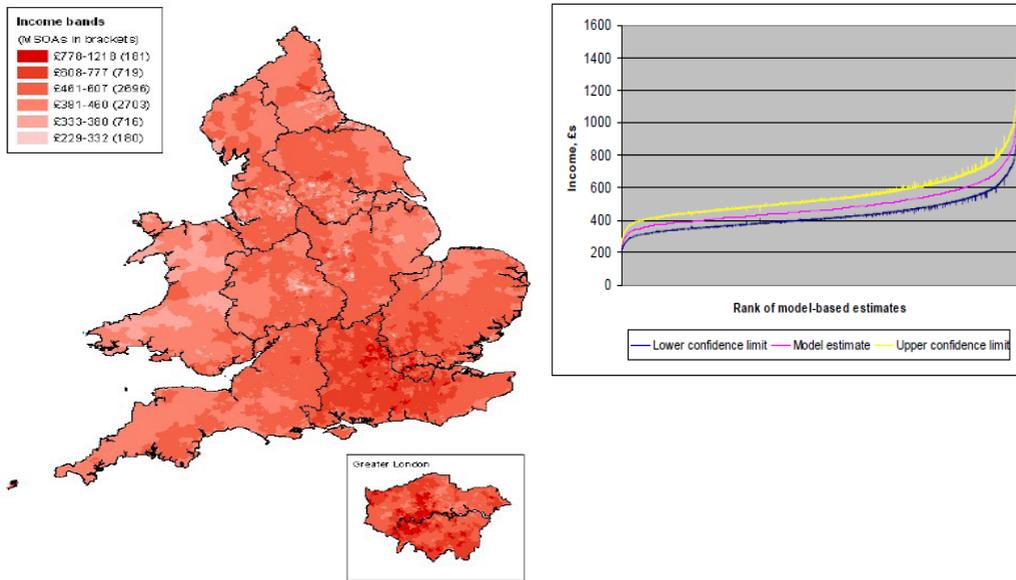
$e_{ij}$  : 개별잔차로 평균은 0, 분산은  $\sigma_e^2$  이다.

모형에 의한 소지역  $k$ 에 대한 주당 가구소득의 평균 추정값은 아래와 같다.

$$\bar{Y}_k = \exp\left(\hat{\alpha} + \hat{\beta}\bar{X}_k + \frac{\sigma_u^2 + \sigma_e^2}{2}\right) \quad (2.11)$$

이렇게 추정된 소지역 소득 추정값은 다양하게 활용되도록 몇 가지 형식으로 제공 되는데, 먼저 신뢰구간을 감안하여 각 소지역 간 소득의 차이 여부를 알 수 있도록 도표로

제공되었다. 또한 소지역 소득을 상하위 각 2.5%, 차상하위 각 10%, 중간 각 37.5%의 총 6개 구간으로 나누어 지도로도 제공하였다. 이러한 예는 다음 [그림 2-4]에 나타나 있다. [그림 2-4]는 2007/08 주거비 사용 후 주당 균등화 가구순소득(AHC)의 소지역(MSOA) 추정 결과이다. 오른쪽의 그림을 보면 신뢰구간을 추정값과 함께 선으로 나타내주어 신뢰구간이 서로 겹치지 않는 소득의 값들이 유의미하게 차이가 있음을 쉽게 파악하도록 하였다. 또한 왼쪽 그림은 지도의 명암을 통해 각 구간들을 나타내어 관련 정보를 한눈에 파악하기 쉽도록 하였다.



※ 출처 : White 등(2009)

[그림 2-4] 2007/08 주거비 사용 후 주당 균등화 가구순소득(AHC)의 소지역(MSOA) 추정 결과

다음은 ONS의 빈곤 관련 소지역 추정 내용이다.<sup>8)</sup> 빈곤 관련 소지역 추정 연구는 소지역 단위의 저소득(빈곤) 가구의 비율을 추정하고자 하는 수요에 부응하기 위한 것이 계기가 되어 시작되었다. 사용한 소득의 종류는 주거비 사용 전 주당 균등화 가구순소득(BHC)과 주거비 사용 후 주당 균등화 가구순소득(AHC)이다.

소득의 소지역 추정과 마찬가지로 여기서 균등화 가구소득은 가구원수 및 종류에 따라 가중값을 주어(McClement's scale) 계산한 것이다. 1998/99 자료를 대상으로 한

8) Longhurst와 Goldring(2004), Rahman 등 참고



추정에서는 BHC를, 2004/05 및 2007/08 자료에서는 AHC를 중심으로 소지역 빈곤율을 추정하였다.

빈곤율의 소지역 추정도 다른 분야와 마찬가지로 SAEP 기본 방법론을 바탕으로 하였는데, SAEP 방법론을 적용하기 위해서는 소득을 빈곤선과 비교한 이진변수로 생성하는 것이 필요하다. 즉 어떤 가구의 소득이 빈곤선 이하이면 빈곤, 빈곤선 이상이면 빈곤하지 않음으로 판단하여 빈곤율을 계산하는 것이다. 이때 빈곤선은 전체 소득 중간값의 60%를 이용하였다. 빈곤선을 전체 소득 중간값의 50% 혹은 70%가 되는 값으로 정할 수도 있으며, 이와 비교한 분석들도 이루어져 있으나 최종 빈곤선은 전체 소득 중간값의 60%를 사용하였다. 이와 관련된 수리적 모형은 소득과 마찬가지로 SAEP 기본 모형으로부터 다음과 같이 표현할 수 있다.

$$y_{ij} \sim \text{Binomial}(1, \pi_{ij}) \quad (2.12)$$

$$y_{ij} = \pi_{ij} + e_{ij} \quad (2.13)$$

$$\text{logit}(\pi_{ij}) = \alpha + \beta \bar{X}_{k(ij)} + u_j \quad (2.14)$$

여기서,  $y_{ij}$  : 표본지역  $j$  내의 가구  $i$ 에 대한 빈곤 여부(이진변수),

$\pi_{ij}$  :  $y_{ij}$ 의 기대확률,

$\bar{X}_{k(ij)}$ : 표본지역  $j$  내의 가구  $i$ 가 속한 소지역  $k$ 에서 보조정보들의 벡터,

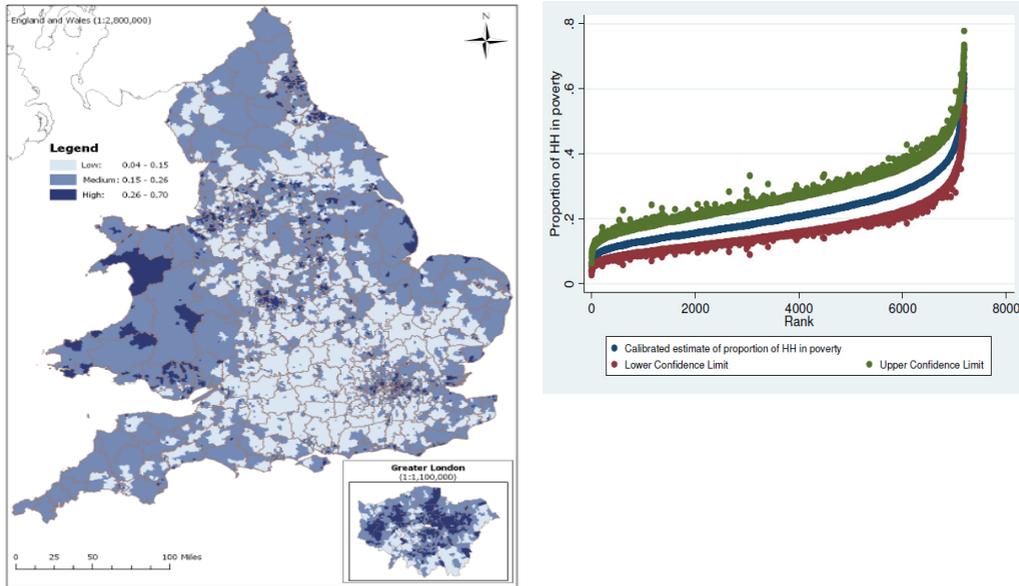
$u_j$  : 지역 수준 잔차로 평균은 0, 분산은  $\sigma_u^2$ ,

$e_{ij}$  : 개별잔차로 평균은 0, 분산은  $\sigma_e^2$ 이다.

모형에 의한 소지역  $k$ 에 대한 빈곤율의 추정값은 아래와 같다.

$$\hat{Y}_k = \text{anti logit}(\hat{\alpha} + \hat{\beta} \bar{X}_k) = [1 + \exp(\hat{\alpha} + \hat{\beta} \bar{X}_k)]^{-1} \quad (2.15)$$

빈곤율의 소지역 추정값 활용은 소득의 경우와 유사하다. 신뢰구간을 감안하여 각 소지역 간 빈곤율의 차이 여부를 알 수 있다. 한편 소지역 빈곤율 지도는 소득과 달리 3단계로 이루어졌다. [그림 2-5]는 2007/08 AHC 기준 빈곤율의 소지역(MSOA) 추정 결과이다.



※ 출처 : Rahman 등

[그림 2-5] 2007/08 AHC 기준 빈곤율의 소지역(MSOA) 추정 결과

### 3. 유럽연합(SAMPLE<sup>9)</sup>)

#### 가. 개요

유럽연합(EU)에서는 7th Framework Programme의 지원으로 “빈곤과 생활 조건 추정을 위한 소지역 방법(Small Area Methods for Poverty and Living Condition Estimates; SAMPLE)”이란 프로젝트를 2008년부터 2011년까지 3년간 수행했다. 이 프로젝트의 목적은 다음과 같다.

■ 프로젝트 목적

- ① 사회적 배제와 박탈에 초점을 맞춘 불평등과 빈곤을 위한 새로운 지표의 개발
- ② 소지역 수준에서 새로운 지표와 그에 부합하는 정확성을 제공할 수 있는 통계적 모형 개발
- ③ 빈곤 지표를 생산하고 해석하기 위한 절차 시행

9) SAMPLE에 대한 구체적인 내용은 [www.sample-project.eu](http://www.sample-project.eu) 사이트를 이용하여 확인할 수 있다.



- ④ 지방 정부에 유용한, 신뢰할 수 있고 유익한 지역 수준의 빈곤과 박탈 지표 제공

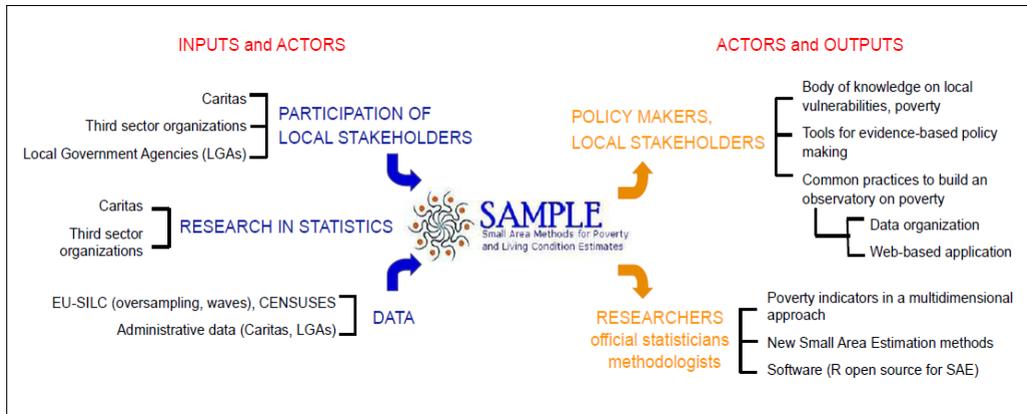
SAMPLE은 추정을 포함한 종합적인 소지역통계 생산에 초점을 맞추고 있으며, 이탈리아, 영국, 스페인, 폴란드 4개국 9개 기관에서 수행하였다. 프로젝트 구조는 6개의 Work Package(WP)로 이루어져 있는데 각 WP의 내용은 다음과 같다.

■ Work Package 내용

- WP1 : 불평등과 빈곤에 대한 새로운 지표와 모형
- WP2 : 빈곤과 불평등 지표의 소지역 추정
- WP3 : EU-SILC와 행정 자료의 통합
- WP4 : 거주 조건 추정을 위한 소프트웨어의 표준화와 응용 프로그램 개발
- WP5 : 관리
- WP6 : 결과의 보급

WP1에서는 델파이 방법을 통하여 다양한 이해 관계자를 참여시키고 의견을 수렴함으로써 새로운 지표와 모형을 개발하기 위해 노력했고, WP2에서는 M-Quantile 모형을 이용하여 NUTS3, LAU1 수준의 소지역을 대상으로 소득과 빈곤지수를 계산하였다. WP3에서는 임금과 생활 조건 통계(EU-Statistics on Income and Living Conditions; EU-SILC)라는 조사자료와 국세청 등 행정자료를 매칭시켜 각 개체의 정보량을 증대시킨 후 oversampling을 통해 소지역통계 생산을 시도하였다. WP4에서는 R-engine을 이용하여 추정 모형을 수행할 프로그램을 작성하고, MySQL DB에 결과를 저장하여 Google Visualization APIs를 이용하여 결과를 제공하였다. 이와 같은 SAMPLE의 절차를 간단한 그림으로 표현한 것이 [그림 2-6]이다. 즉, 지역 통계에 대한 이해 관계자와 전문가들의 의견, 통계적 연구, 수집된 자료들이 SAMPLE을 통해서 정책 수립을 위한 정보로 재탄생하는 것이다.

6개의 WP 중 우리가 관심을 가지는 부분은 WP2와 WP4이다. 소지역 추정을 위해 프로젝트 팀은 10여개 이상의 다양한 소지역 추정 방법에 대한 비교·검토 연구를 수행하였고, 이 중 M-quantile 모형을 주요 모형으로 제안하였다. 그리고 이 다양한 모형들에 대한 R 프로그램을 공개하였다(SAMPLE, 2010a, SAMPLE, 2011).



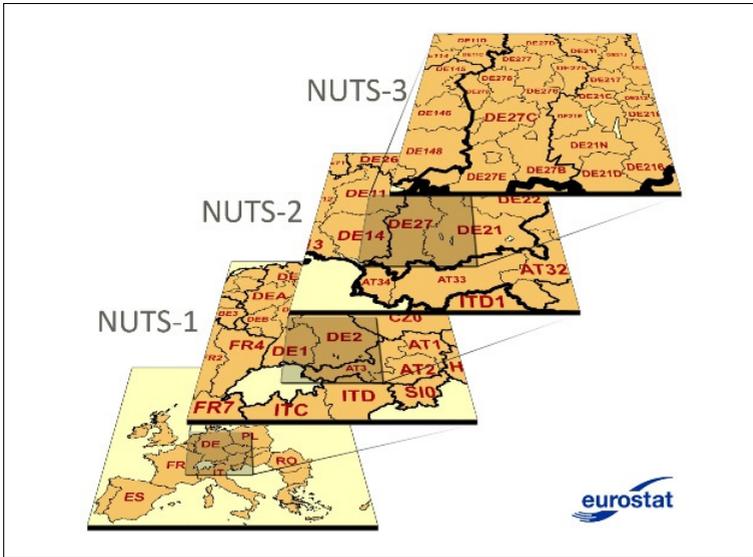
※ 출처 : European PolicyBrief - SAMPLE (2011)

[그림 2-6] SAMPLE 개요도

## 나. 지리 체계

소지역 추정을 위해 사용한 통계적 지리 체계는 NUTS(Nomenclature of Territorial Units for Statistics)와 LAU(Local Administrative Units)라는 두 개의 시스템이다. NUTS는 EU의 경제 영역을 지리적으로 분할하기 위한 계층적 시스템으로 숫자가 높을 수록 하위 계층이고 LAU는 지역 수준 통계 요구에 부응하기 위해 NUTS와 호환되도록 설정한 구역이다([그림 2-7]). NUTS와 LAU의 2006년도 구성 현황을 보면 <표 2-10>과 같다. 27개국으로 이루어진 EU에서 NUTS는 3단계로 구성되어 있고, 그 밑으로 LAU1과 LAU2라는 보다 작은 지리 단위가 있다.

소지역 추정을 위해 사용하는 EU-SILC 조사는 NUTS2 수준까지 공표할 수 있도록 설계되어 있고, 소지역 추정에서는 이를 좀 더 작은 지역인 NUTS3과 LAU1 수준의 통계를 생산하기 위한 노력을 하였다. NUTS3은 유럽 전역을 대상으로 하면 평균 37만 8천 명의 인구를 가지는 단위이며 최소 만 명에서 최대 5백만 명의 인구 분포를 가진다. 2010년 인구주택총조사 결과, 우리나라의 시도는 평균 3백만, 최저 53만(제주도), 최대 979만(서울)의 인구 분포를 가지며 시군구는 평균 21만, 최소 8천(울릉군), 최대 1백만 명(수원시)의 규모를 갖는다. 인구 규모로만 단순 비교 시, 우리나라의 시군구가 유럽연합의 NUTS3보다는 조금 더 작은 크기로 보인다. 다만 최소 인구에서 우리나라 시군구의 경우 울릉군, 옹진군 등 도서 지역을 제외하면, 최소 인구는 영양군의 1만 7천여 명으로 NUTS3과 비슷한 수준이다.



※ 출처 : Eurostat 홈페이지

[그림 2-7] 유럽연합의 통계적 지리 체계인 NUTS 구성

<표 2-10> 유럽연합의 통계적 지리 체계

	Level	지역수 (개)	면적 (km <sup>2</sup> )			인구 (천 명)		
			최소	평균	최대	최소	평균	최대
EU-27	NUTS1	97	161	44,335	302,946	27	5,081	18,067
	NUTS2	271	13	15,869	153,439	27	1,819	11,360
	NUTS3	1,303	13	3,300	98,249	10	378	5,218
	LAU1 (NUTS4)	8,398						
	LAU2 (NUTS5)	121,601						
ITALIA	NUTS1	5	49,301	60,267	72,063	6,668	11,750	15,195
	NUTS2	21	3,239	14,349	25,409	123	2,798	9,434
	NUTS3	107	211	2,816	7,400	90	549	3,854
	LAU1 (NUTS4)	-						
	LAU2 (NUTS5)	8,101						

※ 출처 : Eurostat (2007)

※ 주 : 인구는 2005년도 자료임

### 다. 소지역 추정

소지역 추정을 위해 사용한 주요정보는 EU-SILC로 이 조사는 유럽의 소득 및 생활 수준을 파악하는 것을 목적으로 가구와 개인에 대한 금전적·비금전적 측면에서 사회적 배제 등에 대한 횡단적·종단적 자료를 수집한다. EU-SILC 조사내용은 <표 2-11>과 같다. 모형추정을 위한 보조정보는 주로 센서스의 마이크로데이터를 이용하였으며, 구체적인 추정 사례는 다음의 이탈리아 추정 사례에서 확인할 수 있다.

<표 2-11> EU-SILC 조사내용

가 구	<ul style="list-style-type: none"> <li>◦ 기본 : 도시화 정도 포함</li> <li>◦ 소득 : 가구 총소득 및 가처분소득, 가구 총소득의 구성 요소</li> <li>◦ 사회적 배제 : (비)주택 관련 체납금, 비금전적 가구 박탈 지표, 간신히 연명하는 문제, 부채 범위와 기본 필수품의 부족 정도, 물리적/사회적 환경</li> <li>◦ 주택 : 주거유형, 주거소유형태, 주택상태, 주거시설, 주거비용</li> </ul>
개 인	<ul style="list-style-type: none"> <li>◦ 기본 : 개인 자료와 인구 자료</li> <li>◦ 소득 : 개인 총소득, 개인 총 소득의 구성 요소</li> <li>◦ 노동 : 현재 고용 상태와 직업에 대한 기본 정보(실업일 경우, 가장 최근의 직업), 소득 기준 기간 동안의 고용 상태, 부업의 총 근로 시간, 상세한 노동 정보, 고용 이력, 활동 일정</li> <li>◦ 교육</li> <li>◦ 건강 : 건강상태 및 만성 질병이나 상태, 의료 서비스 접근성</li> </ul>

※ 출처 : SAMPLE Deliverable 9(2010)

모형은 M-quantile 회귀 모형(Chambers와 Tzavidis, 2006)을 제안하였다. M-quantile 모형은 개체 수준(unit level) 자료를 이용하고  $x$ 가 주어진  $y$ 의 조건부 분포 분위수가  $x$ 에 대해 선형이라고 가정한다. 소지역 랜덤 효과는 직접적으로 모형에 포함되지 않는다. 일반적인 형태는 다음과 같다.

$$Q_q(y|X) = X\beta(q) \tag{2.16}$$

여기서,  $q \in (0,1)$  : 분위수이다.

M-quantile 회귀 모형의 특징은 i) 평균 뿐만 아니라 분위 추정도 가능하고 ii) 강한 분포 가정이 불필요하고, 이상값에 강건(robust)하며 iii) 비모수 모형 접근이 용이하고 조사 가중치 결합이 쉽다는 것이다. 반면 M-quantile 추정량은 편향이 발생할 수 있다. SAMPLE은 이런 M-quantile 모형의 장점에 공간 정보를 활용할 수 있는 M-quantile GWR(Geographically Weighted Regression) 모형(Salvati et al., 2008)도 제안하였다. M-quantile GWR 모형의 특징은 지리적으로 가까울수록 더 연관이 있다는 가정 하에



공간 좌표를 사용하고, 전역 변수보다 지역 변수를 사용함으로써 공간 변이를 포착하는 데 용이하며 소지역에서의 MSE 개선 효과를 기대할 수 있다는 것이다. 그러나 편향이 보정된 M-quantile 추정량은 MSE 증가를 이끌 수 있고, Fuzzy Monetary Index 같은 좀 더 복잡한 빈곤 지수를 어떻게 다루는지, 비표집된 모수를 어떻게 추정하는지, 어떻게 지역 수준 자료를 사용하는지와 같은 부분은 선명하지 않다(Rao, 2013)는 단점이 있다.

SAMPLE에서 소지역 추정을 통해 얻고자 하는 관심 추정량은 다음과 같다.

- 가구 평균 균등화 소득(Mean of Household Equivalised Income)과 RMSE
- 분위별 소득(Quantiles of Household Equivalised Income)과 RMSE
- 빈곤 지수(Poverty Indicators)
  - HCR(Head Count Ratio) : 빈곤 인구 비율
  - PG(Poverty Gap) : 빈곤 갭, 빈곤의 정도 측정 가능

여기서, 빈곤 지수는 1984년 Foster 등이 제안한 다음의 식을 이용한다(Foster 등, 1984).

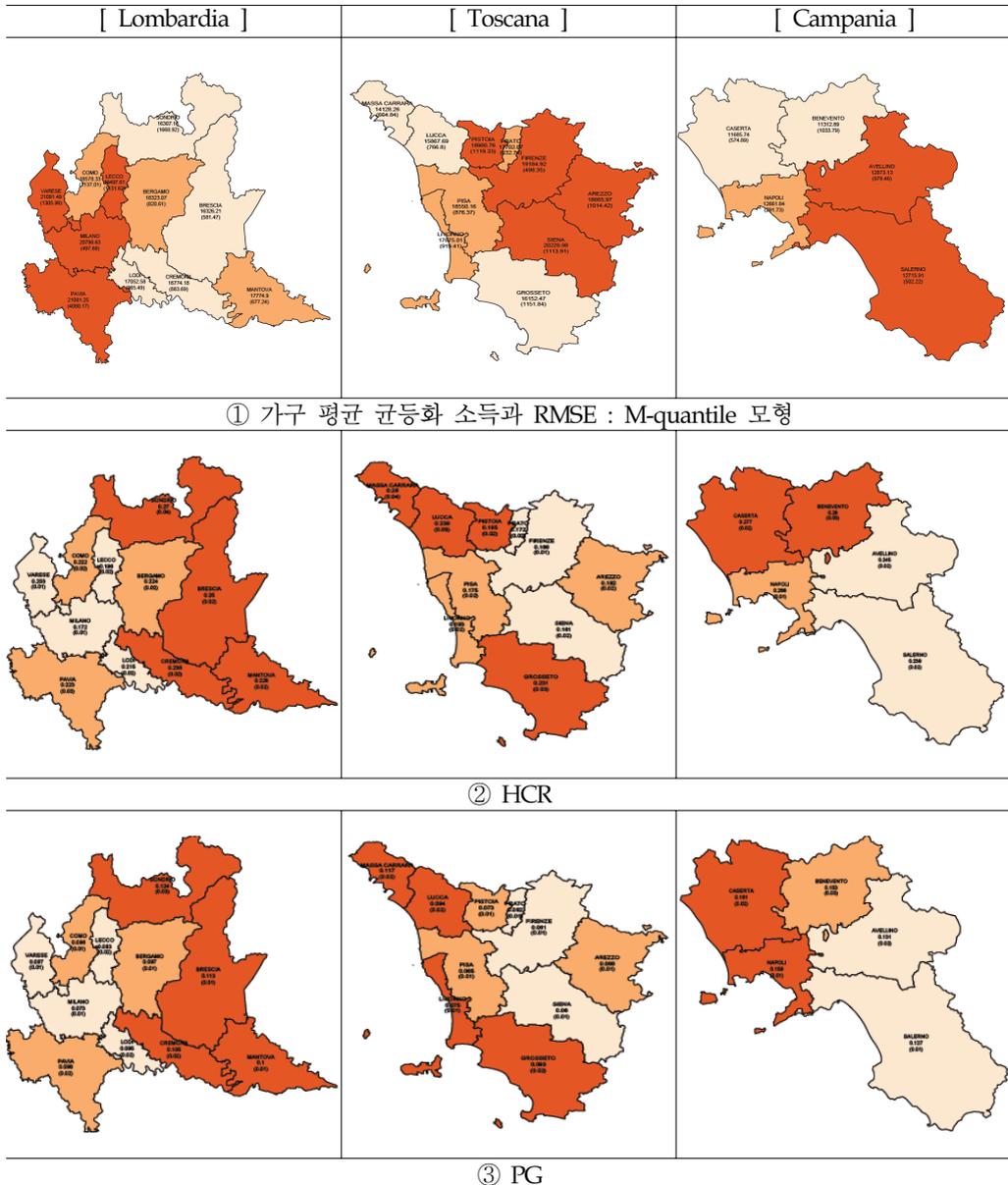
$$F_{\alpha d} = \frac{1}{N_d} \sum_{j=1}^{N_d} \left( \frac{t - y_{jd}}{t} \right)^{\alpha} I(y_{jd} \leq t), \quad \alpha = 0, 1, 2, \quad d = 1, \dots, D \quad (2.17)$$

여기서,  $y_{jd}$ : 양적인 복지량(e.g. 소득),  
 $t$ : 빈곤선(e.g. 소득 중앙값의 60% 수준),  
 $d$ : 지역,  
 $j$ : 개체,  
 $\alpha = 0$ 일 땐 HCR,  $\alpha = 1$ 일 땐 PG이다.

## 라. 이탈리아 추정 사례

SAMPLE 연구팀은 이탈리아 지역을 대상으로 소지역 추정법을 실제로 적용해보았다(SAMPLE, 2010b). 추정을 위해 사용한 자료는 이탈리아 통계청에서 조사한 2007년 EU-SILC 결과와 2001년 센서스 자료이다. EU-SILC는 NUTS2 수준까지 결과를 제공하는 조사이고, 센서스에서는 가구의 연령, 성, 고용상태, 교육정도와 가구 크기 변수를 보조정보로 이용하였다.

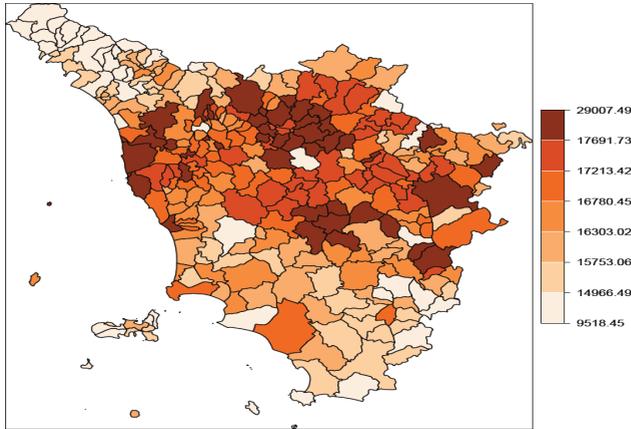
소지역 추정에 이용된 지역은 북부의 Lombardia, 중부의 Toscana, 남부의 Campania 지역으로 NUTS3 지역을 각각 11개, 10개, 5개 가지고 있다. 이들 NUTS3 지역의 가구 평균 소득과 빈곤 지수 소지역 추정은 M-quantile 모형을 이용하였고, 추정 결과는 [그림 2-8]과 같이 나타났다.



※ 출처 : SAMPLE(2010b)

[그림 2-8] 이탈리아의 소득·빈곤을 소지역 추정 결과

[그림 2-9]는 NUTS3보다 작은 LAU2 수준인 Municipality 단위에 대한 소지역 추정을 M-quantile GWR 모형을 이용하여 추정한 결과이다.



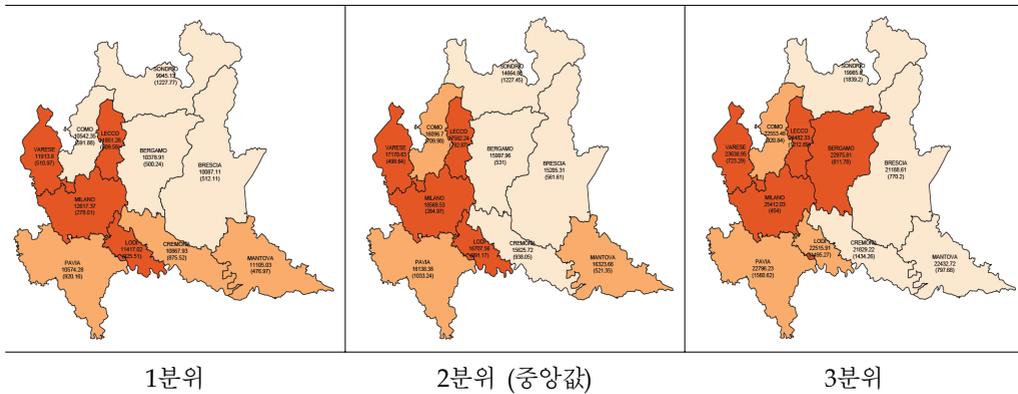
※ 출처 : SAMPLE(2010b)

※ 주 : 1. Municipality 수준 = LAU2 수준

2. 총 287개 소지역 중 65개 지역이 표본이고, 5개 표본 지역에서는 관측 개체수가 1 이하

[그림 2-9] M-quantile GWR 모형을 이용한 Toscana 지역의 가구 평균 균등화 소득 추정 결과

[그림 2-10]은 Lombardia 지역의 분위별 소득 추정 결과를 지도에 표현한 것이다.



※ 출처 : SAMPLE(2010b)

[그림 2-10] M-quantile 모형을 이용한 Lombardia 지역의 분위별 소득

#### 4. 요약

이 절에서는 통계 선진국인 미국, 영국, 유럽의 소지역 소득·빈곤을 추정 프로그램을 각각 살펴보았다. <표 2-13>은 이들 국가의 방법을 요약하여 비교한 것이고, 이를 통해 얻을 수 있는 시사점을 <표 2-12>와 같이 항목별로 간단히 정리해보았다. 앞서 언급한

바와 같이 이들 통계 선진국의 사례를 통해 얻고자 하는 바를 정리하면 다음과 같다.

i) 어떤 자료를 이용했는가

우리나라는 빈곤 관련 지역 정보가 부재하고 소득 정보에는 민감하기 때문에 미국, 유럽과 같이 우선 빈곤인구와 빈곤율에 초점을 맞추는 연구를 수행하면서, 가구 소득 중앙값 혹은 분위값 추정 연구를 수행할 필요가 있다. 그러나 우리나라는 부처 간 정보 공유에 어려움이 있어 자료 확보에는 어려움이 있다. 본 연구에서도 빈곤과 소득을 설명할 수 있는 행정자료를 최대한 확보하고자 노력하였으나, 실제 소지역을 의미하는 시군구 자료는 거의 획득할 수가 없었다. 따라서 보다 작은 지역의 행정자료를 제공하고 제공받기 위한 관련 노력이 추가되어야 한다. 이는 미국의 경우처럼 통계청 뿐만 아니라 관련 부처들의 지역 통계 필요에 대한 공감대가 선행되어야 한다. 또한 보조정보에 보다 독립적이기 위해서 유럽과 영국처럼 공간 정보와 센서스 정보를 활용하는 방안을 구체적으로 모색할 필요가 있다.

ii) 어떤 방법을 이용했는가

미국, 영국, 유럽 모두 모형기반 추정방법을 이용하였다. 미국은 FH 모형, 영국은 영국형 회귀 모형, 유럽은 M-quantile 모형을 사용하였다. 이들 국가는 모두 자국의 실정에 맞는 방법을 선택하고자 다년간 관련 연구를 수행하였다. 우리 역시 최신 기법 도입과 응용(모형 탐색, 공간 정보 활용 방안 등)을 통해, 여러 가지 소지역 추정 모형에 대한 실제 자료 적용과 평가를 시행하여 우리나라의 현실에 맞는 방법을 찾는 것이 필요하다. 특히, 영국과 같이 우리나라의 자료·지역·관심 특성에 맞고, 확장이 용이한 일반적인 추정 방법을 개발하는 노력이 필요할 것으로 보인다. 이를 통해 다양한 지역 경계와 다양한 분야에 대해 개발된 방법을 보다 빠르고 쉽게 적용하게 된다면 지역 통계 생산에 큰 기여를 하게 될 것이다. 이와 같이 조사 자료의 단점(예를 들어, 소지역에서의 소표본)과 행정자료의 장점을 모형을 이용해 결합하여 통계를 생산하는 방식은 세계적으로 활발한 연구와 적용이 이루어지고 있는 분야이기도 하다.

iii) 예산, 인력, 연구 기간 등의 측면에서 어떤 노력을 기울였는가

미국은 센서스국이 1993년에 빈곤과 소득에 대한 지역 통계를 생산할 필요가 있다고 알리는 활동을 시작으로 처음으로 구체적인 지역 통계를 생산하는데 최소 5년 이상의 연구 기간이 소요되었다. 그 과정에서 관련 부처의 자금 지원, 국세청 자료 확보와 법률적 지원이 있었다. 영국 역시 1998년 시작된 SAEP 연구를 소득, 빈곤에 적용하여 소득과 빈곤 모두 승인 통계로 인정받고자 하는데 10년 이상의 기간을 소요하고 있다. 유럽은 미국과 영국에 비해서는 상대적으로 연구 기간이 적었으나, 4개국 9개 기관에서 3개년의 연구를 수행하였다. 반면 우리나라는 지역의 빈곤 통계 필요에 대해서는 공감하지만,



〈표 2-12〉 미국, 영국, 유럽의 사례가 주는 시사점

- 소지역 추정 방법 : 모형기반 추정
- 관심값
  - 미국, 유럽 : 가구 소득보다 빈곤 추정에 초점이 맞춰져 있고, 소득은 주로 가구 소득 중앙값이나 분위값을 추정
  - 영국 : 센서스에 소득 문항이 불포함되면서 소지역의 소득에 관심을 가지게 되어 소득과 빈곤 모두에 대해 활발한 연구 수행이 이루어졌음
- 보조정보의 획득
  - 미국, 영국 : 국세청 자료가 주요 보조정보, 관련 기관들의 자료 협조가 원활
  - 유럽 : 공간정보, 센서스 자료 이용
- 보조정보 사용 방법
  - 빈곤인구, 빈곤율 : 센서스, 국세청 자료 등의 인구와 비율 정보
  - 가구 소득 : 센서스, 국세청 등의 소득 추정값 등 금액 정보가 다수
  - 관심값에 따라 조금씩 다른 형태의 보조정보 사용 검토
- 단계적 연구 수행
  - 미국 : 지역 통계 필요성에 대한 센서스국의 홍보 노력 → 관련 부처(5개)의 자금 지원, 국세청 자료 확보, 법률 지원 → 5년 이상의 연구 → 결과 공표
  - 영국 : 센서스 소득 문항 불포함 → ONS의 노력, 국세청 자료 확보 → 결과 공표 (승인 통계)
  - 유럽 : EU의 관심 → 프로젝트 발주 자금 지원 → 관련자 의견 청취 → 관련 연구 수행

그 과정에서 필요한 노력에 대한 이해는 부족한 듯 보인다. 따라서 이와 관련한 부처 및 전문가의 의견을 청취하고 청취된 의견을 정리하여 다시 피드백하는 환기 작업을 통해 끊임없이 지역 빈곤 통계의 필요성에 대한 이해를 높여야 한다. 이 과정에서 예산, 인력 등 인프라의 확보와 관련자들의 협조 유도는 필수적이다.

#### iv) 어떻게 결과를 활용하는가

미국과 영국은 결과 자료를 국가통계 기관에서 문서와 홈페이지 등을 통해 공표한다. 미국의 경우 법률적으로 지역 통계를 이용하여 예산을 배분하도록 규정하고 있어 소지역 추정 통계에 더 큰 힘을 실어주고 있다. 또한 지도를 이용한 결과 제공을 통하여 사용자의 이해와 편의를 높이고, 시계열 자료 제공을 통한 소지역 소득·빈곤율 변화 추이를 한 눈에 확인할 수 있도록 제공하고 있다. 그리고 개발된 방법을 모두 공개하고 배포함으로써 이용자의 관심을 높이고 있다.

〈표 2-13〉 미국, 영국, 유럽의 소지역 소득·빈곤율 추정 프로그램 요약 비교

	미국 SAIPE	영국 SAEF	유럽 SAMPLE
정식 명칭	Small Area Income and Poverty Estimates (소지역 소득과 빈곤 추정)	Small Area Estimation Project (소지역 추정 프로젝트)	Small Area Methods for Poverty and Living Condition Estimates (빈곤과 생활 조건 추정을 위한 소지역 방법)
목적	연방 정부에 적시에 민을만한 소득과 빈곤 통계를 제공 지역 정부에 대한 연방 기금의 할당	2001년 센서스에 소득 문항을 포함하지 않으면서 대안 필요	불평등과 빈곤을 위한 새로운 지표의 개발 새로운 지표에 대한 소지역 통계 모형 개발
생산 통계	빈곤율, 취약 연령 아동 빈곤율 가구 소득 중앙값	총/순 주간 가구 소득(불균등) 주거비 사용 전/후 순 주간 가구 소득(균등) 빈곤율	빈곤율 가구 평균 균등화 소득과 분위별 소득
주요 정보	American Community Survey(ACS)	Family Resources Survey(FRS)	EU-Statistics on Income and Living Conditions (EU-SILC)
보조 정보	ACS의 소득과 빈곤 직접추정값 2000 센서스 표본의 소득과 빈곤 직접추정값 우세청 세금 자료 SNAP 수급자료 경제분석국의 개인 소득 추정값 SSI 수급자료 연령·그룹별 인구추정값 등	2001 센서스 보조금 청구 횟수 VOA 의회 파세구분 지역구분지표 주거비 소유 변화 HMRC 소득자료, 세금신청자료	2001 센서스 -가구주 인구 특성과 가구 크기
추정 방법	주, 카운티: FH 모형기반 추정(Area level) 학군: 합성 분배 방식	자체 모형기반 추정 -소득은 Linear Random Effect 모형 -빈곤은 Logistic Mixed 모형	M-quantile (GWR) 모형기반 추정(Unit level)
소 지역 범위	주, 카운티, 학군 지역 목적 마감	UA/LAD, MSA 등 지역	NUTS3 수준
비고	관심도에 따라 모형과 사용 변수가 조금씩 다름 홈페이지에 결과 제공 비슷한 단년도 조사에 비해 정도가 높음 유사 소지역 통계로 건강보험(SAHIE)이 있음 빈곤선: 가족의 크기, 가구원의 연령 등을 고려한 전국 통일 기준	공식통계 승인 계획 중 홈페이지에 결과 제공 빈곤선: 전국 소득 중간값의 60%	EU의 자금 지원으로 4개국 9개 기관이 3년간 수행한 실험적 프로젝트임 모형에 공간 정보를 활용하나 지역랜덤효과과는 미포함 빈곤선: 지역 소득 중간값의 60%

※ 출처 : 1. 미국 센서스국 SAIPE 사이트 : [www.census.gov/did/www/saipe/](http://www.census.gov/did/www/saipe/)  
 2. 영국 Model-Based Estimates of Income for Middle Layer Super Output Areas, 2007/08 Technical Report  
 3. 유럽 SAMPLE 사이트 : [www.sample-project.eu](http://www.sample-project.eu)



## 제4절 자료 검토

모형을 이용한 소지역 추정에서 가장 중요한 자원은 자료(data)이다. 이 절에서는 모형화에 필요한 주요자료로 통계청의 가계금융복지조사를 보조자료는 각 지역별로 소득과 빈곤을 설명할 수 있는 국세청, 보건복지부 등의 행정자료를 검토한다.

가계금융복지조사는 가계생활수준의 정도, 변화, 지속기간, 변화요인 등을 종합적으로 파악하여 재정 및 복지관련 정책과 연구에 활용하고 자산, 부채, 소득 등의 규모, 구성 및 분포와 미시적 재무건전성을 파악하여 사회 및 금융관련 정책과 연구에 활용하기 위하여 2012년부터 시작한 조사통계이다.<sup>10)</sup> 조사주기는 1년이고, 표본은 전국 2만여 가구이다. 조사는 자산·부채의 규모, 자산·부채의 분포, 자산·부채의 구성, 가계의 재무건전성, 경제활동 등 복지부문 등에 대한 결과를 공표하고 있다. 가계금융복지조사를 주요정보로 선정한 가장 큰 이유는 가구와 개인의 소득 정보를 포함하는 큰 규모의 표본조사 자료이면서 시도별 소득 평균 및 중앙값이 공표될 수 있도록 설계되어 있기 때문이다.<sup>11)</sup>

가구 경상소득은 크게 근로소득, 사업소득, 재산소득, 사적이전소득, 공적이전소득으로 구성되어 있다. 이들 각 부문별 소득에 대해 이용 가능하고 설명 가능한 행정 자료를 검토한 결과 <표 2-14>와 같은 행정자료 사용이 가능할 것으로 보인다. 그러나 광역 시도보다 작은 시군구 지역에 대해 자료를 제공가능하다고 응답하거나 공식적으로 발표하는 경우는 국민연금 자료 밖에 없어 나머지 자료의 입수에 대해서는 부처 간 혹은 기관 간 추가적인 획득 노력이 필요하다. 이 외에도 이전소득과 관련하여 산재보험, 별정우체국연금 등의 보조정보도 사용을 검토했으나 자료 활용에 따른 장점이 지역별 자료 입수 노력에 비해 크지 않은 것으로 보여 활용하지 않기로 한다. 필요한 행정자료는 한국보건사회연구원의 2012년 빈곤통계연보에서 주로 탐색하였다.

10) 실제 가계금융복지조사 연혁은 다음과 같다.

- 2006년 : 제1회 가계자산조사 실시(5년 주기)
- 2010년 : 제1회 가계금융조사 실시(1년 주기)
  - 가계자산조사(통계청), 가계신용조사(금융감독원), 가구패널조사(한국은행)를 통합
- 2011년 : 제2회 가계금융조사 실시
- 2012년 : 제1회 가계금융·복지조사 실시
  - 전국을 대표하는 2만 가구를 표본으로 추출하여 횡단 및 패널 분석이 가능하도록 표본 재설계(금융 부문 : 1만 가구, 복지부문 : 1만 가구)
- 2013년 : 제2회 가계금융·복지조사 실시
  - 추정의 연속성을 위해서는 2012년 조사 결과부터 활용하는 것이 가장 좋지만, 이럴 경우 너무 짧은 시계열 때문에 추정 통계의 안정성 검토에 제약이 발생한다. 따라서 연구 수행시 적어도 2010년까지 확장하여 연구를 수행할 필요는 있다.

11) 앞서 언급한 통계청의 광역 시도 소득소비 추정 연구 시에는 가계동향조사 자료를 주요정보로 이용했는데, 당시 연구는 가계동향조사의 적은 표본 규모 때문에 소지역을 광역시도 수준으로 하는 한계가 있었다.

〈표 2-14〉 소득·빈곤을 추정을 위해 고려된 보조정보

소관기관	행정자료명	세부항목	자료제공 범위
국세청	소득세	징수실적	시도
	법인세	"	
	상속세	"	
	증여세	"	
	종합부동산세	"	
국민연금공단	국민연금	(가입형태별) 가입자수, 수급자수, 급여실적	시군구
국민건강보험공단	건강보험	(직장/지역) 적용인구, 급여실적	"
고용정보원	고용보험	피보험자수(혹은 비중), 수급자수, 급여실적	시도
공무원연금공단	공무원연금	공무원수, 연금수급합계	전국
사립학교교직원 연금공단	사립학교교직원 연금	인원, 지급액	"
국방부	군인연금	수급자수, 연금액	권역별
보건복지부	기초생활보장	수급자수, 가구수, 급여실적	시도
	기초노령연금	수급자수, 지급실적	"
국민건강보험공단	의료급여	대상자수	"
	노인장기요양보험	인정자수	"

## ※ 출처

1. 국세청 자료 : 국세통계연보(2012)
2. 국민연금 : 국민연금통계연보(2011)
3. 건강보험 : 건강보험통계연보(2011)
4. 고용보험 : 고용보험통계연보(2011)
5. 공무원연금 : 공무원연금주요통계자료(2011)
6. 사립학교 교직원연금 : 통계연보(2011)
7. 군인연금 : 군인연금통계연보(2011)
8. 기초생활보장, 기초노령연금, 의료급여, 노인장기요양보험 : 보건복지통계연보(2012)

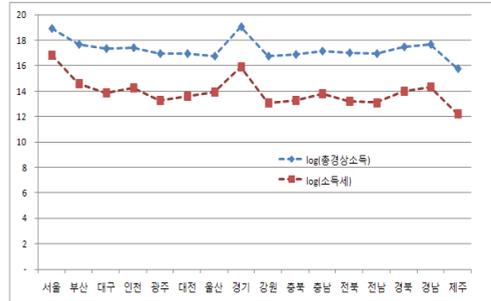
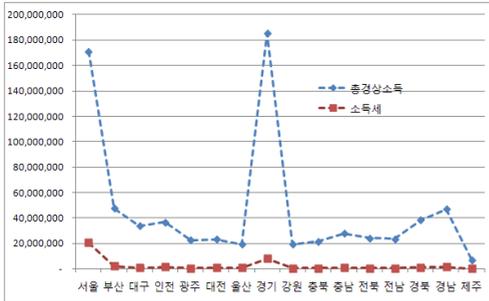
## ※ 주

1. 공무원연금, 사립학교교직원연금 자료는 요청 시 시도 자료 제공 가능
2. 군인연금 자료는 광역권별로 제공, 요청 시 시도 자료 제공 가능
3. 관련 자료는 한국보건사회연구원의 2012년 빈곤통계연보로부터 탐색

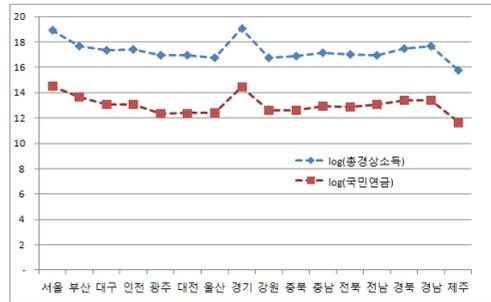
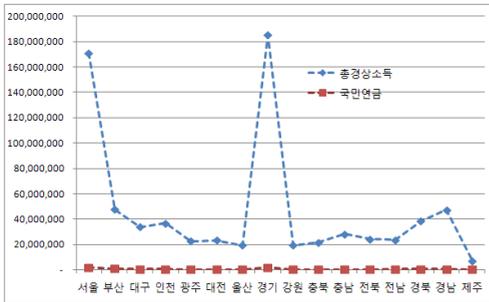
다음으로 주요정보와 보조정보가 모형에서 어느 정도의 설명력을 가질 수 있는지 확인하기 위하여 가구 경상소득과 각 보조정보의 16개 시도 단위에서의 상관 정도를 분석해 보았다(〈표 2-15〉). 여기에서 가구 경상소득은 가계금융·복지조사의 2011년 시도별 평균 가구소득(2012년 공표)에 시도별 추계가구수를 곱하여 산출한 값이고, 행정자료는 2011년 시점의 자료를 집계하여 2012년에 공표된 자료이다. 각 정보는 모두 금액 자료로, 지역별, 항목별 규모 차이가 크기 때문에 상관분석을 위해서 모두 로그 변환하였다. 소득세와 더불어 국민연금과 같은 공적이전소득이 가구 소득과의 상관 정도가 높음을 알 수 있다.

〈표 2-15〉 시도 수준에서 가구 경상소득과 각 보조정보 간의 상관 정도

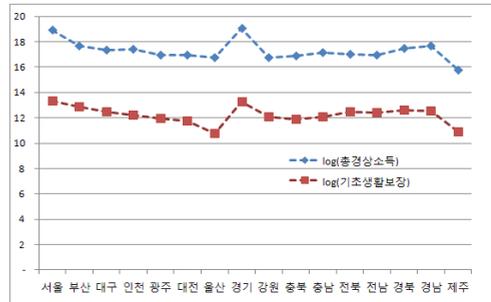
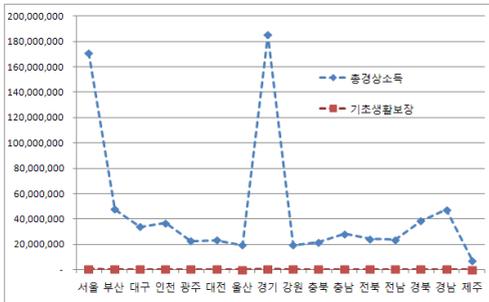
소득세	법인세	상속세	증여세	종합부동산세	국민연금	건강보험	실업급여	기초생활보장	기초노령연금
0.96	0.90	0.90	0.91	0.90	0.97	0.98	0.98	0.85	0.90



① 총경상소득 vs. 소득세 징수실적 원자료 vs. 로그변환 후 분포



② 총경상소득 vs. 국민연금 급여실적 원자료 vs. 로그변환 후 분포



③ 총경상소득 vs. 기초생활보장 급여실적 원자료 vs. 로그변환 후 분포

[그림 2-11] 시도 수준에서 가구 경상소득과 각 보조정보 원자료 vs. 로그변환 후 분포

## 제5절 결론

국내의 소득과 빈곤 관련 지역 통계 생산 노력은 주로 행정자료를 취합하거나 표본 규모를 확대하는 방향으로 이루어져 왔다. 그러나 행정자료는 비통계적 목적으로 수집되어지기 때문에 요구되는 지역통계의 목적이나 대상과 맞지 않고 목표 모집단을 대표하지 못한다. 또한 표본 규모를 확대하는 방법은 많은 인력, 예산과 시간을 필요로 한다. 이런 어려움을 극복하고 지역 통계 수요에 부응하기 위해 통계청은 “소지역 추정”을 이용한 지역 통계 생산을 시도한 바 있다. 그러나 추정 통계에 대한 이해 부족과 보조 정보 획득의 어려움 때문에 계속적인 연구에 난항을 거듭하였다. 본 연구는 이러한 어려움을 극복하고 안정적으로 소지역 소득·빈곤 추정 통계를 생산하기 위한 초석을 다지기 위한 기초연구이다.

본 연구에서는 지역 빈곤 통계 수요 증가라는 외부 환경의 변화에 맞춰 통계청 차원에서 수요에 대응하기 위해 어떤 접근을 시도해야 하는지 검토해 보았다. 이를 위해 국내에서 지역 통계를 생산하기 위한 노력을 리뷰하고, 일찍부터 소지역 추정을 이용한 소득·빈곤을 통계를 생산하고 제공해 온 주요 통계 선진국인 미국, 영국, 유럽의 사례를 중심으로 문헌 검토를 실시했다. 이를 통해 얻고자 한 정보는 소득·빈곤을 추정을 위해 i)어떤 자료를 이용했는가 ii)어떤 방법을 이용했는가 iii)예산, 인력, 인구 기간 등의 측면에서 어떤 노력을 기울였는가 iv)어떻게 결과를 활용하는가였다.

결론부터 말하면 우리는 비용, 시의성, 안정성 측면에서 모형을 이용한 지역 통계 생산 노력을 지속할 필요가 있다. 앞서 살펴본 미국, 영국, 유럽은 이미 조사자료와 센서스, 행정자료와 같은 보조정보를 이용한 소지역의 소득·빈곤 관련 모형기반 추정값을 생산하고 있다. 모형기반 추정은 큰 비용의 추가없이, 소지역에서는 직접추정값에 비해 보다 안정된 추정값을 생산할 수 있다. 또한 확정된 모형에 이미 확보된 정보를 입력자료로 활용하면 빠른 시간 안에 원하는 결과를 손에 넣을 수 있다. 이는 통계 생산자와 이용자의 욕구를 모두 충족시킬 수 있다. 왜냐하면 양측 모두 가용한 예산과 자원이 효율적으로 분배되고 사용되길 바라기 때문이다.

특히, 미국과 영국의 사례는 우리에게 시사하는 바가 크다. 미국은 센서스국이 1993년에 빈곤과 소득에 대한 지역 통계를 생산할 필요가 있다고 알리는 활동을 시작으로 처음으로 구체적인 통계를 생산하는데 최소 5년 이상의 연구 기간이 소요되었다. 그 과정에서 관련 부처의 자금 지원, 국세청 자료 확보와 법률적 지원이 있었다. 이는 우리나라의 이해관계자들이 지역의 소득·빈곤 통계 필요에 대해서는 공감하지만, 정보 제공 협조 등과 같이 지역 통계 생산 과정에서 필요한 지원에 난색을 표하는 경우를 설득시키는 데 중요한 사례가 될 것이고, 지역 통계를 생산하고자 하는 통계청의 향후 연구 진행 방향을



설정하는데 중요한 나침반이 될 것이다. 또한 법률적 지원을 통해, 생산된 지역 통계의 활용성을 높이고 보다 품질좋은 통계를 생산하는 노력(예산, 인력의 확보)에 힘을 실어 줄 수 있다. 또한 영국은 사회통계와 관련된 소지역 추정 문제에 대해 일반적인 통계적 방법론을 개발하고 시스템을 구축하였다. 이를 통해 다양한 지역 경계와 분야에 대해 소지역 추정을 꾸준히 적용하고 실제 공식통계로도 활용하고 있다. 이와 같은 영국 사례를 벤치마킹하는 것은 급속히 늘어나고 있는 우리나라 지역 통계의 다양한 수요에 대해 보다 유연하게 대처할 수 있는 최선의 방안이 될 수 있다.

지금까지 소지역의 소득·빈곤을 추정을 위하여 국내외 소득과 빈곤 관련 지역 통계 생산 현황을 살펴봤다. 본 연구에서는 특히, 모형을 이용한 추정에 초점을 두고 해외 사례를 검토했다. 그에 맞춰 우리나라의 자료를 이용하여 기본적인 모형추정 적용까지 시도하려고 했으나, 원하는 소지역(시군구) 수준까지 주요정보와 보조정보를 획득하지 못해 이를 수행하지 못한 아쉬움이 있다. 또한 해외의 사례를 검토하면서 느낀 바는 우리나라도 여러 가지 이해에 따라서 변할 수 있는 시도, 시군구와 같은 행정구역에 따른 통계 생산보다는 NUTS(유럽), MSOA(영국) 등과 같은 변하지 않는 통계적 지리 단위를 설정하여 이에 맞는 지역 통계를 생산하려는 노력이 보다 필요할 것으로 보인다. 이런 아쉬움과 한계는 이후 수행될 관련 연구를 통해 해결되길 바란다.

## 참고문헌

- 공무원연금공단 (2011), 공무원연금주요통계자료.
- 국민건강보험공단 (2011), 건강보험통계연보.
- 국세청 (2012), 국세통계연보.
- 국방부 (2011), 군인연금통계연보.
- 국민연금공단 (2011), 국민연금통계연보.
- 권순필, 김서영, 이정희 (2011), 지역통계 작성을 위한 소지역 추정 방법 연구, 통계개발원.
- 근로복지공단 (2011), 고용보험통계연보.
- 김달호, 김재광 (2004), 가계조사 지역별 추정기법 개발, 통계청 용역 연구 보고서.
- 김문길, 김태완, 우선희, 이서현 (2012), 2012년 빈곤통계연보, 한국보건사회연구원.
- 김서영, 권순필 (2010), 소지역 고용통계 추정, 통계개발원.
- 김서영, 권순필 (2013), 고용통계 소지역 추정 실무 적용 매뉴얼, 통계개발원
- 김혜원, 정동명 (2005), 가계조사 지역별 추정기법의 타당성 검증, 통계청.
- 보건복지부 (2012), 보건복지통계연보.
- 사학연금공단 (2011), 사립학교교직원연금 통계연보.
- 오민수, 성은미 등 (2012), 지도로 보는 경기도 빈곤, 경기복지재단.
- 한국보건사회연구원 (2012), 빈곤통계연보.
- Denis B. N. Silva and Philip Clarke (2008), Some initiatives on combining data to support small area statistics and analytical requirements at ONS-UK.
- European PolicyBrief - SAMPLE (2011).
- Eurostat, <[epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home](http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home)>
- Eurostat (2007), Regions in the European Union (Nomenclature of territorial units for statistics, NUTS 2006 / EU-27), Methodologies and working papers, Eurostat.
- Jane Longhurst (2004), Model-based estimation of income: measuring change over time, 7th meeting of the national statistics methodology advisory committee.
- Jane Longhurst and Shayla Goldring (2004), Modelling households on low income by ward, 7th Meeting of the national statistics methodology advisory committee.
- Jerry Maples, William Bell (2007), Small Area Estimation of School District Child Population and Poverty: Studying Use of IRS Income Tax Data, Research Report Series, U.S. Census Bureau. <<http://www.census.gov/srd/papers/pdf/rrs2007-11.pdf>>
- J.N.K. Rao (2013), Small Area Estimation : Methods, Applications and New Developments, NTTS 2013 conference.
- Nargis Rahman, Philip Clarke, Kevin McGrath and Denise Britz do N. Silva, Small area estimates of household in poverty for England and Wales, ONS.
- Nicolar White, Alistair Dent, Philip Clarke, Denise Silvia, and Jane Naylor (2009), Model-based estimates of income for middle super output areas, 2007/08 technical report.

- ONS, Model-based ward level estimates of household income, 1998/99 User guide for Scottish estimates.
- ONS, Model-based estimates of income for wards in England and Wales, 2001/02 User guide.
- ONS, Model-based estimates of income for middle layer super output areas in England and Wales, 2004/05 User guide.
- ONS (2001), Small area estimation in the ONS, National statistics methodology advisory committee (01) 3.
- Ray Chambers and Nikos Tzavidis (2006), M-quantile models for small area estimation, *Biometrika*, 93(2) 255-268.
- SAMPLE, <[www.sample-project.eu](http://www.sample-project.eu)>
- SAMPLE (2010a), Final Small Area Estimation Developments and Simulation Results, SAMPLE Deliverable 12 and 16.
- SAMPLE (2010b), Pilot Applications, SAMPLE Deliverable 17.
- SAMPLE (2011), Software on Small Area Estimation, SAMPLE Deliverable 22.
- U.S. Census Bureau, <[www.census.gov/did/www/saipe/index.html](http://www.census.gov/did/www/saipe/index.html)>
- U.S. Census Bureau (2012), Small Area Income and Poverty Estimates (SAIPE): 2011 Highlights.
- William Bell, Wesley Basel, Craig Cruse, Lucinda Dalzell, Jerry Maples, Brett O'Hara, David Powers (2007), Use of ACS Data to Produce SAIPE Model-Based Estimates of Poverty for Counties, U.S. Census Bureau. <<http://www.census.gov/did/www/saipe/publications/files/report.pdf>>

