

# IASS Satellite Meeting on Small Area Estimation

## I. 개요

### 회의주제 및 참가목적

- 회의주제
  - 소지역추정기법에 대한 최근 연구동향 및 적용사례 발표
- 회의일자 및 장소
  - 2007. 9. 3 ~ 9. 5. 이탈리아 피사
- 참가목적
  - 소지역 추정에 대한 최근 선진 기법 연구동향 및 적용사례 파악
    - 이론과 방법론적인 개선
    - 다양한 조사 실례 등에 대한 해외 사례 파악
  - 국내 소지역 추정에 대한 연구방법 비교 및 향후 연구방향 설정

### 참석자 및 발표논문

- 발표논문
  - 22개 세션발표와 1개 토론으로 구성
  - 73편 발표 : 초청논문 7편, 일반 논문 66편

○ 참석자 : 117명

- 대학교 82명 (70.1%), 정부기관 28명 (23.9%), 협회 및 기구 7명 (6.0%)으로 구성

- 우리 청 참석자 : 사무관 김진(지역통계과),  
주무관 전찬선, 권순필(연구기획실)

○ 발표주제

### 다양한 소지역추정기법 연구

- 기존의 추정기법 이외의 다양한 소지역 추정기법 제시
- 소지역에 대한 분산 및 평균제곱오차(MSE) 추정
- 소지역을 위한 Benchmarking 및 표본설계 기법
- 베이지안, 붓스트랩, 로버스트, 비모수통계 등의 다양한 통계적 개념도입으로 연구영역 확대

### 소지역추정기법 적용범위 확대

- 노동력조사에서의 소지역추정에 대한 연구결과 발표
  - 센서스자료를 이용한 시뮬레이션 결과
  - EBLUP, EB, HB 등 소지역모형 적용
- 다양한 통계조사로 소지역추정 적용범위 확대
  - Household, Business, Agriculture, Health and Medical 등

## II. 발표자료 요약<sup>1)</sup>

### 소지역추정 이론분야

- 다양한 모형의 적용
  - 이변량 이항 자료에 대한 HB 추정량
  - 비대칭(skewed) 자료에 대한 소지역 추정
  - 소지역 추정에서 이상치(outlier) 처리
  - 비모수 소지역 추정, 선형화 추정과 예측방법 등
- 분산 추정 방법
  - 각종 모형기반 추정방법에 대한 분산추정법 연구
  - 붓스트랩, 로버스트, 리샘플링 방법을 이용한 분산추정법 제안
- 베이지안 방법
  - 베이지안 소지역추정기법, 베이지안 Benchmarking 기법 등

### 공식통계 적용사례

- 호주
  - 노동력, 의료, 장애, 가구 소비 등 분야에서 적용 연구
  - 2005년에 「small area estimation practice manual」 발간.
- 영국
  - 실업자수에 대한 모형기반 추정량 제공(분기별)
  - 평균 가구 수입 등에 대한 소지역자료 개발연구 중

1) 자세한 내용은 2007 SAE 홈페이지 : <http://www.dipstat.ec.unipi.it/SAE2007/> 참조

- 스페인
  - Basque 지역의 노동력 조사에 적용.
  - 2001년도 센서스 자료를 이용한 시뮬레이션 연구를 이용하여 여러 추정량들에 대한 효율 비교
- 네덜란드
  - 노동력 조사를 이용한 다양한 소지역 추정 모형과 그 진단에 대한 연구.

## 분야별 적용사례

- 노동력조사 분야
  - 시뮬레이션 연구
    - .. 호주의 Social Insurance Providers의 자료와 호주 노동력조사 자료를 병합하여 실업률에 대한 시뮬레이션 연구 수행
    - .. 스페인, 오스트리아 등에서도 센서스 자료를 이용한 연구수행
  - 모형을 이용한 연구
    - .. 뉴질랜드, 네덜란드, 영국 등 대부분의 국가에서는 우리나라와 마찬가지로 EBLUP, EB, HB와 같은 모형기반 소지역추정기법을 실제 자료에 적용·연구
    - .. 모형기반 소지역추정기법에 의해 노동력 통계조사의 결과를 연간 또는 분기별 생산하여 제공하는 국가(미국, 영국, 일본)도 있음
- 농업통계조사 분야
  - 위성자료를 보조변수로 이용하여 경지의 이용과 사용변화를 예측하는데 소지역모형(신경망 시계열 예측모델) 적용(이탈리아)

- 이탈리아 농가구조조사  
특정 대상의 대다수에서 무응답이 발생하여 표본규모가 작은 경우, 소지역 모형인 EBLUP을 이용하여 무응답층을 추정하여 처리하는 사례연구
- 이탈리아 농촌지역경제 분석  
소지역에 대한 농가소득 자료 추정하기 위해 유사지역간 군집 형성 및 EBLUP을 적용하는 사례연구
- 그 외 적용분야
  - 가계 지출, 빈곤, 건강·의료 등 다양한 분야에서 적용 연구가 진행되고 있음

### Ⅲ. 결론 및 시사점

- 소지역추정기법에 대한 연구는 계속되어야 함.
  - 소지역추정기법에 대한 연구는 1970년대에 시작하여, 90년대 중반 이후 본격적으로 활발하게 진행되었으며, 2003년에 J.N.K.Rao에 의해 “Small Area Estimation”이라는 책이 처음으로 발간되었음.
  - 소지역추정에 대한 연구는 장기간 연구되어야 할 영역으로 학계와 정부기관에서 지속적인 관심을 보이고, 연구를 진행하여야 함.
  - 따라서 소지역 추정법에 대한 개선, 개발을 위해서는 지속적으로 예산과 시간을 투입하고, 연구인력 확보를 위한 노력이 필요하다고 판단됨.

□ 국가통계에서 소지역추정의 현위치는

- 소지역 추정에 대한 필요성에 대해서는 많은 국가들에서 인정하고 있으며, 추정량에 대한 내외부에서의 공식적인 지지 및 승인을 얻는 것이 관건임
  - 해외 국가의 경우, 국가의 공식 통계기관에서 직접 다양한 분야에서 연구를 진행하고 있으며 일부 국가의 경우 그 결과를 공식 발표하고 있음(미국, 영국, 일본)
  - 결과를 공식 발표하고 있지 않은 국가의 경우에는 우리나라의 경우와 마찬가지로 간접추정량에 대한 의구심이 아직 남아있기 때문으로 보임. 그러나 자체 연구는 지속적으로 수행하고 있음.

□ 우리 통계청의 소지역추정 방향설정은,

- 소지역 추정 모형 개발 및 모형에 대한 진단, 분산의 추정에 대한 지속적인 연구의 필요성이 있음.
  - 대규모 센서스 자료를 이용한 시뮬레이션 연구를 통하여 추정량의 효율성, 편의성에 대한 검증 연구가 선행되어야 함.
  - 연구에 대한 전반적인 과정을 공개하거나 매뉴얼화하여 지속적인 연구 수행이 가능토록 해야 함.
  - 연간 또는 분기별 소지역자료를 생산하여 각 지역에 공급한 후, 통계청과 지역간 『소지역 네트워크』를 형성하여야 함.
- 실업자(률) 추정 뿐만 아니라, 가계조사, 농업통계 분야 등 그 적용범위를 넓혀서 연구 수행
  - 특히, 가계조사 및 농어가경제조사와 같이 응답부담이 높고, 고비용조사에 대해서는 표본규모를 대폭 감소하여 현수준의 결과를 공표할 수 있는 소지역추정기법을 개발하도록 노력해야 함.

## [별첨] 호주, 영국, 스페인의 소지역 추정 연구 수행 현황

### □ 호주

#### ○ 소지역 추정의 역사

- 1970년대부터 시작. 합성추정량, 비추정량 등에 초점을 맞춤. 계산 능력의 부족으로 어려움을 겪음
- 1980~90년대에는 의료, 장애, 노동력, 가구 소비 등에 대한 다양한 적용을 시도
- 1990년대 후반 상대제곱근평균제곱오차(RRMSE)에 기반하여 품질 측정
- 2005년 「Small area estimation practice manual ABS(2005)」 발간.
  - 소지역 추정 조사자와 사용자를 위한 소지역 추정량의 생산, 사용, 품질 평가, 타당성 등에 대해 기술

#### ○ 소지역 추정 방법

- 다음의 네 가지 선택 사항을 이용한 회귀 모형 구조 사용
  - 1) 선형 혹은 일반화된 선형 모형을 사용할 지 여부
  - 2) 모형은 일변량 혹은 다변량인지 여부
  - 3) 임의 효과를 포함할 지 여부
  - 4) 모형을 단위/개인 수준으로 할지 지역 수준으로 할 지 여부
- 보조 정보의 주요 소스는 인구 센서스

#### ○ 소지역 추정량에 대한 품질 측정 문제

- 세 개의 단계를 거침
  - 1) 사용된 소지역 모형의 적합과 specification 부분의 통계적 적합성을 평가
  - 2) Brown et. al(2001)이 제안한 품질 진단 과정을 적용
  - 3) 소지역 추정량의 그럴 듯함, 일치성, 표면적 타당성 등에 대하여 사용자, 학교, 정부 기관들로부터 “전문가 의견”을 듣고 그 타당성을 확보

- 호주 통계청(ABS)에서 나온 품질 관련 이슈들
  - 보조 자료가 기대한 만큼 관심(조사) 데이터와 강하게 관련되어지  
지 않음. 모형 예측에 대한 신뢰도와 정확성에 대한 우려 발생
  - 모형 예측의 실행을 평가하기 위한 표준화된 진단 방법이 존재하  
지 않음
  - 품질 측정 문제에서 거쳐야 하는 세 단계의 조화 문제.
  - 최종 결과에 대한 사용자 이해를 높이는 문제. 교육 워크샵이나  
보고서 발간시 품질 측정 결과를 포함함으로 해결해야 함

○ 기타

- 소지역 추정 공표 전 고려해야 할 사항들
  - 1) 좋은 모형을 가지는 소지역 추정량 생산을 위해서는 시간과 비  
용 소비가 수반됨
  - 2) 소지역 추정은 한 묶음으로 대량생산할 수 없음.
  - 3) 궁극적으로 모형에 들인 시간과 노력을 정당화시킬 만큼 모형  
이 잘 적용될지 여부가 명확하지 않음
  - 4) 충분한 시간을 주지 않고 빨리 결과를 공표하도록 하는 압박이 있음
  - 5) 센서스 자료를 보조 정보로 이용하는 한, 센서스 조사 사이의  
기간 동안에 생기는 변화를 모형에 어떻게 반영시킬 수 있는  
가에 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 고려해야할 필요가 있음

○ Reference

- Janice Wooton, Daniel Elazar and Marion McEwin. (2007)  
Small Area Estimation : Issues Facing a National Statistical  
Office, *SAE*.

## □ 영국

### ○ Neighbourhood Statistics DB 이용

- 빈곤층에 대한 공공의 정책 계획 수립에 대한 보조를 목적으로 만들어짐
- 행정자료, 센서스 자료 등을 수록

### ○ 소지역 추정량에 대한 공식 공표 및 연구 활동 수행

- 노동력조사에 대하여 공식 통계 생산
  - .. 모형 기반 추정량을 분기별로 발표 : EBLUP
  - .. 행정 자료인 실업수당요구율을 보조 정보로 이용
- 가구 평균 수입에 대한 실험 통계(experimental stat) 연구
  - .. 1998/99추정 결과를 2001/02년에 발표
  - .. 선거구(4천 가구 정도 포함하는) 수준에서 발표
- 소지역 추정량 개선을 위해 지역간 상관관계를 포함하는 모형 검토

### ○ 사용자의 소지역 추정에 대한 주요 요구 사항

- 소지역 추정량과 보다 큰 지역의 직접 추정량 사이의 일치성 문제
- 반복 추정의 경우, 시간이 흐름에 따라 발생할 수 있는 변화에 대한 평가 문제
- 빈곤층에 대한 소지역 추정적용 시 비율이나 분위수 등 추정 모형 적용 검토

### ○ Reference

- Philip Clarke, Kevin McGrath, Hukum Chandra, Nikos Tzavidis. (2007), Developments in Small Area Estimation in UK with Focus on Current Research Activities, *SAE*.
- Clarke, P. (2005), Small Area Estimation, UK Experiences, *ISI*

## □ 스페인

- 소지역 추정 R&D 프로젝트 수행
  - 2004년에 Eustat과 Navarre 대학 협력 수행
  - 2005년에 경제와 산업 통계에서 소지역 추정량 생산
  - 현재, 노동력조사 특히 실업자수 소지역 추정량을 얻기 위한 시물레이션 연구 수행
  
- 시물레이션 연구
  - 2001년 센서스 자료로부터 표본 추출하여 연구 수행
  - 보조 정보 : 나이, 성별, 교육 수준, 이전 활동 상태 등 최근의 센서스로부터 확보
  - 연구 추정량 : 직접 추정량, 사후 증화 추정량, 합성 추정량과 복합추정량, GREG를 이용한 모형기반 추정량
  - 결과 : 복합추정량의 효율이 가장 좋았음
  
- Reference
  - Lola Ugarte, Ana F. Militino, Cristina Prado, Javier San Vicente, Marina Ayestaran. (2007), Small Area Estimation in Official Statistics : The Labour Force Survey of The Basque Country, *SAE*