

2009 Small Area Estimation Conference

참가 결과 보고

I. 참가 개요

학술대회 일자 및 장소 : 2009. 6. 29~7. 2. 스페인, 엘체

학술대회명 : 2009 Small Area Estimation conference

학술대회 목적

- 유럽공동체 국가들의 소지역 추정 전문가들이 모여 소지역 추정을 위해 학문적, 실질적 방법들에 대한 연구결과를 발표·토의
- 조사환경이 점점 어려워지는 상황에서 조사를 하지 않고 기존의 조사를 이용하여 추정하는 다양한 아이디어를 공유하는 자리
- 특히, 학자들과 통계기관 이용자들이 이론 또는 현실적 측면을 활발하게 논의함으로써 조사통계의 현실적인 대안을 모색한다는 데 이 학회의 의미가 있음

참석자 : 5급 김서영 (통계개발원 연구기획실)

출장 배경 및 목적

- 통계개발원에서 연구하고 있는 소지역 추정에 관한 연구결과를 발표
- 소지역 추정의 선진국인 영국, 호주, 미국, 캐나다 등의 전문가들을 직접 만나 서로의 방법을 공유하고자 하였음
- 이론적 방법보다는 소지역 추정이 공식통계로서 어떻게 자리매김 할 수 있었는지에 대한 현실적인 측면을 배우고자 하였음

II. 소지역 추정 연구 동향

□ 본 학회의 이슈

- 소지역 추정과 관련하여 다양한 주제들이 논의되었으나, 크게 몇 가지 압축해 보면 다음과 같다.
- 즉, 내용별로는 고용실태와 관련한 실업자(율)추정과 수입·지출과 관련한 빈곤도 측정이 핵심 쟁점이었음
- 방법별로는, 로지스틱 모형의 적용 및 응용과 전국 단위 또는 대지역 단위에서는 조사추정치와 소지역추정치의 수준을 맞추는 벤치마킹에 관한 내용이 논의 대상이었음

□ 보고서 전개방법

- 한국의 발표 내용 요약·소개
- 이번 학회에서 논의된 내용 중 소지역 추정의 이론적 측면과 활용성 측면으로 나누어 소개

1. 한국측의 발표내용

- 내용 요약
 - 경제활동인구조사를 이용한 실업자 수 추정
 - 한국의 경제 상황을 잘 반영할 수 있는 추정 방법을 찾는 것이 목적
 - 다양한 추정방법을 비교함으로써 최적의 방법을 찾고자 하였음
 - 추정방법으로 합성, 복합, eblup, hb 추정방법을 고려
 - 보조정보를 실업급여 청구자 수를 사용(이에 따른 한계 직면)
 - 2005년 인구주택 총조사 자료를 이용한 모의실험 연구
 - 결론은
 - 모형에 기반한 HB 추정량이 결과적으로 가장 우수함
 - 그러나 HB는 사전·사후분포 가정에 매우 의존하게 되며, 추정과정이 모의실험을 따르기 때문에 공식통계 추정에 사용하는데 불리
 - 차선으로 모형기반인 EBLUP 을 선택
 - 좋은 모형선택에 대한 연구가 병행되어 진행 중

A comparative study of small area estimators
for the Korean economically active population survey

초록(원문)

Korea, as many other countries, faces the growing need for up to date and reliable data on small areas. In the case of labour market statistics, municipal estimates of unemployment are in high demand. Although a national survey may provide direct survey estimates for a province or municipality, budget constraints limit the sample sizes and other means of improving estimation have to be devised. This paper describes the research at the Korea National Statistical Office(KNSO) on small area estimation of the number of unemployment from the Economically Active Population Survey (EAPS), nationwide monthly survey employing a sample of about 32,000 households.

The goal of the study is to find the best estimator for the Korean situation among many estimators. For the goal, simulation study has been undertaken to compare several estimators, design based and model based, for unemployment. We compared five estimators, i.e., direct estimator, synthetic estimator, composite estimator as indirect estimators, and EBLUP(Empirical Best Linear Unbiased Predictor) and HB(Hierarchical Bayes estimators) as model based estimators. Here, we used the Fay-Herriot model. For all the estimators considered in this study, the relative bias and the relative mean square errors were obtained using a resampling method in which the EAPS design was simulated based on data from the 2005 Korean Census.

2. 활용성 측면에서 소지역 추정(각국의 공식통계로의 적용)

○ 국가별 실업자(율)의 소지역 추정

- 소지역별 실업자 추정은 1990년대 중반을 기점으로 통계 선진국들이 연구의 관심을 가졌던 분야임
- 왜냐하면 아주 작은 지역 단위까지 조사에 의해 실업자수를 추정하기에는 엄청난 예산이 소요되는 어려움이 있기 때문임
- 또한 소지역 수준까지의 조사는 조사부터 공표까지 걸리는 시간이 상당하여 조사시점으로부터 시의성에 문제가 제기되고 있음
- 이러한 점을 고려하여 시의 적절하게, 적용비용으로 수요자와 국가의 요구를 충족할 수 있는 소지역 통계를 작성하기 위해 보다 세련된 소

지역 추정방법들에 연구가 꾸준히 진행되고 있음

- 소지역 실업자수 추정은 특히, 조사환경이 열악한 유럽과 북미를 중심으로 거의 모든 국가가 공식통계의 작성방법으로 도입하고 있고, 각국의 공식통계 작성을 위한 연구에 많은 비용과 인력을 투입하고 있음

(1) 스페인

- 노동력 조사(Labour several survey)의 실업자(율) 추정
 - 다항 로짓 혼합모형(multinomial logit mixed model) 사용
- Spanish Household Budget Survey의 소지역 추정
 - unit level linear model (시공간 정보를 활용한) 적용

(2) 네덜란드

- 노동력 상태와 교육정도 추정
 - 네덜란드 통계청에서 사회조사 프로그램을 재구성하면서 모형기반추정 방법을 적용
 - 연간 빈도에 대해 450개 소지역을 추정
 - 지역기반 또는 단위 기반 모형을 적용 검토

(3) 호주

- 노동력조사에 대해 소지역 추정 방법 개발
 - 현재는 추정결과가 정책결정을 위한 안정적인 품질을 유지하는지에 대해 결정하는 과정
 - 로지스틱 혼합효과 모형을 사용
 - 행정자료(정부연금수혜자수, 실업급여 수급자 수)+센서스자료를 보조정보로 활용하여 모형추정치 효율을 높이고 있음
- ⇒ 현재 개발원 소지역 연구에서도 보조정보 선택에 대해서 신중할 필요성이 있으며 이들의 선택은 우리에게도 좋은 귀감이 됨

(4) 이탈리아

- ISTAT Labour force survey 실업률 추정

- 추정에서 모형과 보조정보 역할의 중요성을 암시
- 공간상관을 갖는 linear mixed model과 로지스틱 혼합모형 사용
- 2001년 센서스 data를 이용한 모의실험 연구

(5) 그 외 국가 : 룩셈부르크, 브라질 등

○ 빈곤도 측정

- 사회복지에 관심이 높아지면서 지역별 빈곤도 측정을 위한 움직임이 활발하게 진행되고 있음
- 가계의 소득(income)과 지출(expenditure)과 관련된 기존의 통계를 활용하여 소지역(영역)별 빈곤도를 측정하는 연구가 한창임
- 유럽공동체는 SAMPLE 프로젝트를 진행 중에 있고, 그 내용을 간단히 소개하면 다음과 같음

SAMPLE

(**S**mall **A**rea **M**ethods for **P**overty and **L**iving condition **E**stimates)

- 연구기금은 유럽공동체에 의해 지원, 2010년을 겨냥
- 목적
 - 사회적 배제와 박탈이라는 특별한 관점에서 불평과 가난 정도를 이해하기 위한 사회적 지표를 개발하는 것
 - 소지역 수준에서 이들 지표와 그와 관련한 통계적 정확성 측정을 위한 통계 모형을 개발하는 것
 - 빈곤지표를 생산하고 해석하기 위한 절차를 수행하는 것
 - 지역수준에서 정의된 가난·빈곤에 관한 믿을만한 지표를 제공하는 것
- 단계: 크게 3가지 단계를 거침
 - 센서스와 지역 행정자료를 조사자료와 연계하는 단계
 - 빈곤 연구와 소지역 추정 전문가 팀과 상호 협력하는 단계
 - 정책입안자와 정부의 조직(NGO)과의 협의하는 단계

(1) 폴란드

- 최소 빈곤율과 빈곤 갭을 추정
- 지역 랜덤효과 모형을 사용하여 EBLUP 추정량에 의한 추정

(2) 브라질(영국의 Pedro silva가 개발)

- poverty mapping
 - 이 방법은 최근 많은 나라에서 사용하고 있는 방법으로
 - 브라질은 5,564 municipalities 의 poverty mapping
 - 2002/03 family budget survey의 수입과 지출을 사용

(3) 스페인

- 빈곤지표의 소지역 추정
 - 스페인 Living conditions survey 자료를 이용
 - 시간 종속 지역 기반 모형을 사용

(4) 이탈리아

- 소득 불일치 추정
 - 시공간 모형(time seires and cross sectional model)
 - Istat의 ECHP와 Bank of Itay의 SHIW에 대해 각 모형 검증

3. 이론적 또는 방법론적 측면에서 소지역 추정

○ 벤치마크(Benchmark) 방법에 관한 연구

- 모형기반 추정치는 추정결과를 보다 더 큰 지역으로 합쳤을 때 그 결과가 직접추정치(direct estimate)와 일치하지 않는다는 문제에 직면
- 대지역에서는 조사가 더 큰 지역에서 더 정확하도록 설계되기 때문에 직접추정치가 더 신뢰할 만 함
- 또한 소지역 추정치의 유일성을 증명하기 위해서라도 벤치마크는 필요하게 됨
- 방법에는 크게 raking 방법과 모형에 기반한 방법이 있음

(1) 베이지안 벤치마킹

- 레이킹 방법은 가장 일반적인 벤치마킹 방법으로서 두 가중합이 일치하도록 특정 요인에 대해서 소지역 추정치를 조정하는 방법
- 이는 확고한 통계적 원칙이 없을 때 사용하면 편리
- 베이지안 벤치마크는 가중평균 또는 가중평균과 변동을 사용함

(2) 이중 벤치마킹

- 미국BLS에서는 Time series 모형에 의해 소지역 실업자수를 추정함
- 이때 소지역은 9개의 census 지역, 50개 state를 대상으로 함
- 이 모형은 조사오차 모형과 모집단 값 모형을 조합하지만 외부 영향에는 둔감함
- 따라서 이 문제를 해결하기 위해 이중 벤치마킹을 사용
- 이 방법은 벤치마크된 추정량의 분산과 공분산을 계산함

(3) 카운트 자료의 모형 벤치마킹

- 벤치마킹 절차는 보다 더 큰 지역에서 직접추정치와 소지역 추정치의 총합이 같도록 조정하는 과정
- 연속형 자료에 대한 자체 pseudo EBLUP 연구를 확장
- 카운트 자료에 대한 unit level logistic mixed model을 이용한 방법

○ 모형(modeling)에 관한 연구

- 이미 모형에 대해서는 많은 연구가 이루어진 상태이지만
- 최근, logistic random effect model,
- 시공간 정보를 이용한 spatial model,
- Bayesian model 기법을 이용한 방법들이 공식통계 작성방법으로 사용되고 있는 추세임
- 그러나 아직 이들 방법들은 많은 경우가 연구 중에 있음

※ 발표된 자료에 대한 자세한 정보는 다음 웹사이트를 참고

<http://cio.umh.es/sae2009>

Ⅲ. 참가소감 및 시사점

- 통계기법의 발전은 그 만큼 조사환경이 열악한 지역(국가) 또는 분야일 수록 발전할 수밖에 없음
 - 조사통계로 해결할 수 없는 다양한 정보를 획득하기 위해서는 보다 더 정확하고 세련된 통계기법이 필요하게 됨
 - 소지역 추정분야는 유럽공동체를 중심으로 조사하기 힘든 작은 지역이나 영역을 대상에 대해 통계를 작성하기 위한 방법으로 이미 많은 국가에서 또는 다양한 통계 분야에서 실용적으로 활용되고 있음

- 통계 선진국일수록 통계분야의 기초 연구가 튼튼하고, 기초연구가 뒷받침될 수 있을 때 통계 선진국 대열에 들어설 수 있음
 - 우리나라도 2000년대를 들어서면서 이미 조사통계에 대한 한계에 직면하고 있음.
 - 물론, 유럽이나 북미의 수준은 아니지만, 사생활 보호문제, 단독가구, 노인 가구 비율의 증가로 조사는 점점 어려워지고 있고,
 - 우리 통계청은 이에 대해 행정등록자료를 통계자료로 활용하기 위한 다양한 방법을 모색하고 있음
 - 또 하나 조사통계의 좋은 대안은 다양한 행정자료를 보조정보로 활용할 수 있는 small area estimation 방법이 될 수 있음

- 공식통계 작성기관은 학계와 꾸준한 연계를 함으로써 공동연구할 수 있는 기반 형성이 중요하다고 보여짐
 - 영국, 호주, 미국, 캐나다 등 많은 통계선진국들은 통계청과 대학 및 연구소간의 협조가 매우 원활하게 이루어지고 있고, 우리의 일회성 용역의 성격이 아닌 장기적인 공동연구의 형태를 유지하고 있음
 - 이는 통계청 발전의 가장 기본적인 기틀이 되어야 한다고 생각됨

- 현재의 유럽이나 북미의 통계작성 상황이 머지않은 미래의 우리의 모습일 수 있다는 생각
 - 오히려 연구현황 및 기관과 학계와의 관계유지를 잘 벤치마킹함으로써 우리의 준비기간을 훨씬 단축할 수 있을 것임
 - 꾸준히 통계선진국들의 연구현황을 주시하면서 우리 통계청을 업그레이드 하는 것이 중요하다고 생각됨
 - 지금 당장의 결과에 연연하기 보다는 북미·유럽 등 통계선진국가들처럼 과제에 따라서 또는 사안에 따라서 보다 원시안적인 시각으로 국가적 프로젝트를 운영하는 것이 절실하다고 판단됨

- 특히, 소지역 추정기법은 조사에 소요되는 엄청난 예산을 절감하고, 동시에 열악한 조사환경을 극복하여 소지역(영역)에서의 필요한 통계정보를 생산할 수 있는 방법으로서 국제적으로 자리매김하고 있음