

제4장

국가별 소지역 고용통계 추정 현황

소지역의 통계를 생산하기 위하여 소지역 추정방법을 이용하는 것은 더 이상 특별한 일이 아니며, 다른 나라에서는 이미 추정을 통한 소지역 별 통계 생산이 증가하고 있다. 뿐만 아니라 통계적 추정 방법의 개발·개선·적용으로 소지역 통계의 질적 향상을 위한 꾸준한 노력이 진행되고 있다. 본 연구에서는 소지역 고용통계 생산에 있어서 모형 기반 추정 기법을 잘 적용하고 있는 국가들 중 미국, 일본, 캐나다, 영국의 사례를 중심으로 그 현황을 파악하고, 이로부터 시사점을 얻고자 한다. 현황 파악을 위해서 ① 각국의 고용통계조사 개요(목적, 조사방법, 표본 체계와 추출 방법 등), ② 소지역 고용통계 발전 과정(공표 현황, 활용, 행정구역에 따라 구분되는 지역 인구·표본 규모 포함), ③ 소지역 추정방법(모형 종류, 보조정보, 벤치마킹¹⁾ 등) 등에 대해 살펴보고자 한다.

특히 국가통계 작성에 있어서 소지역 추정은 그 지역의 인구와 표본 규모에 따라 추정방법이 다르게 적용될 수 있다. 뿐만 아니라 소지역 통계의 작성 목적과 소지역 정의에 따라서도 추정방법은 달라질 수 있다. 그러한 측면에서 본 연구는 각 국가의 소지역 정의와 그 소지역의 인구와 표본규모 등을 파악하고, 각 국가들이 추정방법을 선택함에 있어 이런 정보들이 어떤 관련이 있는지에 대해서 살펴보고자 한다. 이는 우리가 추정방법을 선택함에 있어 단지 통계적 측면에서의 효과만을 고려하지 않고, 사용상의 활용성과 편리성을 충분히 반영하고자 하기 위함이다.

제1절 미국

1) 소지역들의 간접추정량의 합과 대지역의 직접추정량을 일치시켜주는 것

1. 상시인구조사 개요

미국의 고용 동향 파악은 상시인구조사(Current Population Survey: CPS)를 통해 이루어지며 이 조사 결과가 소지역에 대한 고용통계 추정 시 주요 정보가 된다. CPS는 미국 전체와 각 주(State)에 대한 노동력 실태 파악을 목적으로 매월 약 60,000여 가구의 표본에 거주하는 16세 이상 가구원을 대상으로 실시하는 노동력 통계조사이다. 이들 중 시설가구 내에 살고 있는 사람 등은 제외하고 국내외 직업군인, 외국인 취업자는 조사 대상에 포함된다. 조사의 수행은 가구 조사에 대한 경험이 풍부한 센서스국(Census of Bureau)에서, 조사 결과의 처리와 분석은 노동통계국(Bureau of Labor Statistics: BLS)에서 하고 있다.

CPS는 1953년부터 4-8-4 연동표본제(rotation sampling system)를 도입하였다. 이 방법은 표본가구가 연속되는 4개월 동안 조사에 참여, 이후 8개월 동안은 조사를 중단, 그 다음 4개월 동안 다시 조사에 참여한 후 교체되는 방식이다. 따라서 8개의 연동그룹 중 6개 그룹이 전월에 이어서 연속으로 조사를 받고 2개 그룹은 조사에서 교체되어 중복표본비율이 전월대비 75%, 전년 동월대비 50%인 표본 구조를 가지고 있다(통계청, 2003).

미국 전체 주 지역은 2,007개의 PSU들로 나뉘어져 있고, 다시 이들 PSU는 각 주 내에서 층으로 구분되어 있다. CPS는 우선 모집단에서 754개의 PSU(Primary Sampling Unit, 일차표집단위 또는 표본지역)를 추출한 다음, 이들 PSU로부터 다시 60,000여 가구를 추출하는 방식으로 표본을 뽑고 있다. 각 층에서 PSU를 뽑을 때는 PSU의 인구에 비례하여 추출한다. 이처럼 CPS의 표본설계는 기본적으로 주라는 지역에 기반하는 방식을 채택하고 있다. 그러나 하나의 주에 최소한 한 개 이상의 PSU가 포함되어 있다는 측면에서는 주 기반 설계방식이지만, 모든 주를 하나의 층으로 보고 주 내에서 가구를 선택하지는 않는다는 측면에서는 엄격한 의미의 주 기반 설계와는 구분될 수 있다. 또한, CPS 표본설계가 주 단위 기반임에도 불구하고 <표 3-1>에 따르면 주의 표본이 매우 작고(작

은 주에서 평균 950여 가구, 큰 주에서 약 2,200여 가구) 결과값들은 유의한 정도의 표집오차를 포함한다(BLS, 2004). 이런 이유에서 미국은 카운티 단위는 물론이고 주별 통계생산에서도 소지역 추정기법을 적용하고 있다.

2. 소지역 고용통계 발전 과정

미국의 소지역을 위한 실업자 추정은 60여년 이상의 기간 동안 개발, 개선되어져 왔다. 1950년대에 이미 실업보험(Unemployment Insurance: UI) 자료를 보조정보로 이용하는 “소책자 방법”을 사용하면서 표본에 대한 규모의 확장 없이 소지역의 취업과 실업 추정량을 생산하기 시작하였다. 1973년에는 통계국이 주관이 되어 CPS와 이전의 소책자 방법을 결합(개념·정의·추정 분야 등)함으로써 주 단위와 주의 세부단위의 노동력도 추정할 수 있는 기법을 개발하였다.

1976년에는 CPS의 사용이 모든 주로 확대되었으며 그 후에 BLS는 실업률이 6%일 때 실업에 대한 최대 기대 CV를 10%로 사용해 왔다. 이는 월 CPS 자료를 노동력에 대한 직접 추정 자료로 공식 발표하기 위해 사용하는 기준치이다. 이런 기준에 따라 1978년 초에 10개의 큰 주(캘리포니아, 플로리다, 일리노이주, 매사추세츠, 미시간, 뉴저지, 뉴욕, 오키오, 펜실베이니아, 텍사스)와 2개의 지역(PMSA, 뉴욕)에 대한 공식적인 노동력 추정량이 생산되었다. 더불어 1976~1978년에 주와 지역의 노동력 추정을 위해 사용되던 모든 고용보험 실업수당자 자료를 표준화하였다. 이러한 작업을 통해 매달 12일(CPS 참조 주간)을 포함하는 주간(week)의 UI 실업수당자에 대한 자료를 자동갱신하게 되었다. 이로써 UI 정보가 CPS에서 사용되는 실업률에 대한 보조 정보로써 좀 더 일관성을 유지하게 되었고, 이는 주 사이의 비교를 보다 가능하도록 하였다.

1985년에는 북캐롤라이나가 직접 추정량 제공 지역으로 추가되었고, 월 실업률 추정량에 대한 최대 기대 CV를 11개의 큰 주들에 대해 8%로

4 고용통계 소지역 추정 연구 I

낮았다. 나머지 39개의 간접추정량 제공 지역 주와 DC에 대해서는 연평균 실업에 대하여 실업률이 6%일 때 최대 기대 CV가 8% 정도가 되도록 기준을 정했다. 이처럼 실업률의 신뢰도에 대한 CV가 줄어들게 된 것은 1976년부터 CPS 표본이 지속적으로 늘어났기 때문에 가능한 일이었다. 이와 같은 39개의 간접추정량 제공 지역 주들과 DC, 주 하부 지역들에 대한 월별 공식 추정량의 생산은 1989년까지는 부분적으로는 소책자 방법에 기초했다. 그러나 1989년 초, 이 39개 주와 DC에 대한 추정량은 BLS에 의해 개발된 시계열모형에 기초하도록 변경되었다.

1992년에는 시계열모형으로부터 주 추정량에 대한 계절 조정이 도입되었고, 1994년에는 좀 더 발전된 회귀 모형을 도입하게 되었다. 이후 1996년 초, 예산 절감에 따라 CPS 표본 크기가 56,000에서 50,000가구로 줄어들면서, 과거와 같은 직접 추정 방법은 지속될 수 없었다. 그 이후로 11개의 큰 주들에 대한 고용통계 역시 보다 작은 주들에서 사용되었던 시계열모형에 의해 추정되는 것으로 바뀌었다.

그리고 2005년 1월, 지역 실업통계(The Local Area Unemployment Statistics: LAUS) 프로그램의 품질을 높이기 위하여 큰 변화가 반영되었다. 우선, 주와 보다 큰 지역에 대하여 실시간 벤치마킹 과정을 포함하게 되었고, 보다 많은 지역에 대한 최선의 추정 기술을 제공하도록 하였다. 또한 다른 모든 지역에 대해서도 보다 좋은 기술과 입력 자료를 사용하고 2000년 센서스 조사에 기반하여 지역 정보를 갱신하였다. BLS는 이와 같이 생산된 소지역의 노동통계를 경제분석이나 노동시장 연구, 연방기금 배정 등에 이용하고 있다(BLS, 2005).

이런 과정을 거쳐 현재 미국은 약 7,200여개의 소지역에 대하여 경제활동인구, 취업자 수, 실업자 수, 실업률 등의 고용통계량을 제공하고 있다. 이들 소지역에는 모든 주, 워싱턴 DC, Puerto Rico, 대도시(Metropolitan areas), 작은 노동시장 지역(Labor Market Areas: LMAs), 카운티, 뉴잉글랜드에 있는 도시들, 인구 25,000명 이상의 도시들이 포함되어 있다(BLS, 2005). 고용통계 추정에 있어서, 미국은 주 및 주와 대등

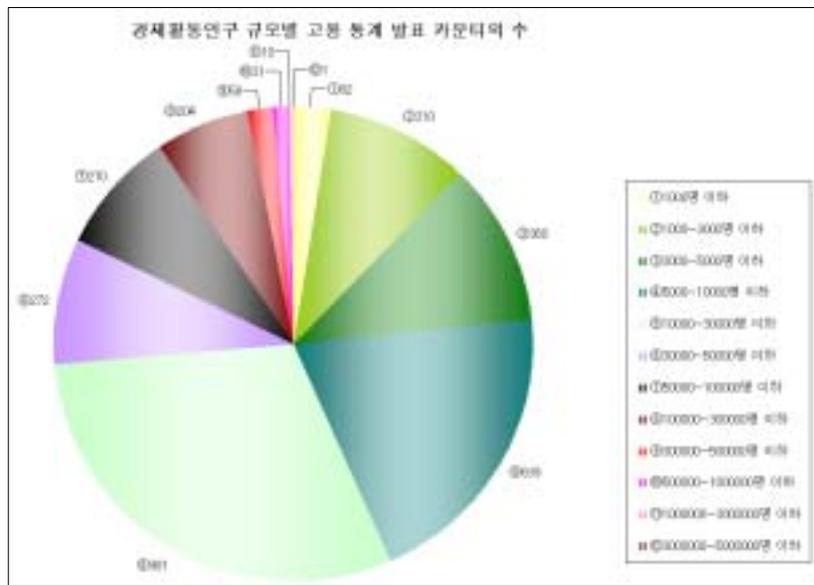
한 규모의 주 하위 지역(substate)에 대해서는 모형 추정법을 사용하고, 그 이하 하위 지역에 대해서는 분해기법(disaggregation technique)을 사용한다. 특히, 카운티, 타운, LMA의 하위 지역의 추정은 각 지역의 인구와 실업급여 청구자 정보를 활용하여 배분하는 방식을 사용한다. 카운티와 같은 아주 작은 지역단위들은 표본규모가 작기 때문에 모형 추정은 오히려 추정치들의 분산을 크게 할 위험이 있다. 더구나 미국에서 사용하고 있는 시계열 모형은 한 지역 내에서 자신의 과거 정보를 사용하여 현 시점을 추정하기 때문에 표본수가 작을 경우, 추정치의 신뢰성이 떨어질 수밖에 없다. 이에 BLS은 카운티와 같은 아주 작은 지역에 대해 모형 추정법을 적용하는 대신에, 대안으로 적절한 부차 정보를 활용한 할당 또는 분해 기법을 사용하는 것으로 보인다. 분해기법 절차와 구체적 방법은 BLS의 'LAUS Handbook'을 참고할 수 있다.

그렇다면 미국의 주 또는 카운티 단위의 인구 또는 표본규모는 어느 정도일까. 대표적으로 카운티별 인구 규모를 살펴보기 위해 <표 4-1>과 같이 2000년 센서스 조사에서 나타난 미국의 지방 조직 현황, 즉 주별 인구와 CPS의 표본 분포를 확인해 보았다. 미국은 1개 수도와 50개 주로 이루어져 있으며, 3,200여개의 카운티를 가지고 있다. 각 주별 CPS 표본규모를 보면 주에 따라 많게는 2,931(텍사스 주)가구에서 적게는 557(미시시피 주)가구로 큰 차이를 보인다 (<표 4-1>). 이것을 주보다 작은 카운티 단위로 본다면 주 내에서 카운티 당 표본가구의 수는 평균 약 20여 가구(전국 표본가구 수/전국 카운티 수)로, 카운티 수준 내에서의 표본가구 수에 있어서 큰 차이를 보인다. 이는 우리나라의 시군구당 표본가구 수에 비하면 훨씬 적은 규모이다. 참고로 2008년 현재 우리나라 시군구당 가장 작은 조사구 수는 3개로, 가구로 보면 60가구(월별 조사구당 표본가구 수는 20가구)가구에 해당된다.

경제활동인구 규모로 카운티별 고용통계 발표 현황을 보면 [그림 4-1]과 같다. 경제활동인구가 1,000명 이하인 카운티가 82개(2.6%)이며 1만 명 이하인 카운티는 1,391개(43.2%)였다. 이 중 가장 작은 인구 규모

6 고용통계 소지역 추정 연구 I

를 가지고 있는 카운티는 텍사스 주의 러빙카운티로 인구가 67명으로 이 지역의 2007년도 고용 현황을 보면 경제활동인구 41명에 취업자 수 37명, 실업자 수 4명(실업률 9.8%)으로 발표되었다.



[그림 4-1] 미국의 경제활동인구 규모별 고용통계 발표 카운티 수

3. 소지역 추정 방법

BIS는 모든 주와 콜롬비아, 주 하위 지역(Los Angeles-Long Beach-Glendale, metropolitan division, New York City, the respective balances of state)의 고용통계 작성에 있어서 시계열모형을 사용하고 있다(Scott and Smith, 1974; Bell and Hillmer, 1990). 위에서 언급한 주와 주 하위 지역들보다 작은 지역들에 대해서는 모형에 의한 추정보다는 분해

기법(disaggregation technique)을 사용한다.

분해 기법은 LMA 하위 지역, 카운티, 시, 타운 등에 대한 고용통계 추정에 사용된다. 두 가지 방법이 LMA 추정치를 분해하는데 사용된다. 첫 번째 방법은 “population-claims” 방법으로 선호되는 기법이다. 이 기법은 LMA 내 총 요구자 수에 대한 하위 지역의 요구자 수의 비를 사용하여 하위 지역 수준에서의 실업 추정치를 분해해 내는 방법으로, LMA 지역 내의 하위 지역들에 대해서 거주자 기반(residence-based)의 실업보험수급 요구자 자료를 이용할 수 있을 경우에 사용된다. 두 번째 방법은 “census-share” 방법으로, 실업보험수급 요구자 자료를 이용할 수 없는 경우에 사용된다. 이 방법은 실업자와 취업자를 분해하기 위해 총 LMA의 실업자와 취업자가 각 하위 지역의 센서스 자료를 공유하는 방식을 사용한다. 본 절에서는 시계열모형추정법을 위주로 설명하는 관계로 분해기법은 자세하게 설명하지 않기로 한다.

주 단위 고용통계 추정을 위한 ‘signal-plus-noise’ 모형 추정 방법에 대해서 설명해 보자. 이 모형에서 자료는 소지역에서 수행된 CPS 조사의 과거와 현재 자료를 모두 이용한다. 이는 소지역들의 표본 규모가 작기 때문에 현 시점의 CPS 자료만 사용하여 직접추정을 하게 되면 소지역들 간 추정치의 변동성이 매우 커지기 때문이다.

8 고용통계 소지역 추정 연구 I

〈표 4-1〉 미국의 주별 인구 및 CPS 표본 분포

No	STATE	COUNTY				CPS 표본	
		개수	인구 분포(천명)			가구 수 (호)	구성비 (%)
			최소	평균	최대		
1	Alabama	67	10	66	662	631	1.06
2	Alaska	27	1	23	260	800	1.35
3	Arizona	15	9	342	3072	773	1.30
4	Arkansas	75	6	36	361	652	1.10
5	California	58	1	584	9519	4,576	7.71
6	Colorado	63	1	68	555	1,253	2.11
7	Connecticut	8	109	426	883	1,360	2.29
8	Delaware	3	127	261	500	912	1.54
9	District of Columbia	1	572	572	572	997	1.68
10	Florida	67	7	239	2253	2,478	4.17
11	Georgia	159	2	51	816	1,217	2.05
12	Hawaii	5	0	242	876	818	1.38
13	Idaho	44	1	29	301	662	1.11
14	Illinois	102	4	122	5377	1,758	2.96
15	Indiana	92	6	66	860	901	1.52
16	Iowa	99	4	30	375	1,132	1.91
17	Kansas	105	2	26	453	879	1.48
18	Kentucky	120	2	34	694	891	1.50
19	Louisiana	64	7	70	485	549	0.92
20	Maine	16	17	80	266	1,177	1.98
21	Maryland	24	19	221	873	1,453	2.45
22	Massachusetts	14	10	454	1465	924	1.56
23	Michigan	83	2	120	2061	1,424	2.40
24	Minnesota	87	4	57	1116	1,410	2.37
25	Mississippi	82	2	35	251	557	0.94
26	Missouri	115	2	49	1016	1,000	1.68

〈표 4-1〉의 계속

No	STATE	COUNTY				CPS 표본	
		개수	인구 분포(천명)			가구 수 (호)	구성비 (%)
			최소	평균	최대		
27	Montana	56	0	16	129	641	1.08
28	Nebraska	93	0	18	464	921	1.55
29	Nevada	17	1	118	1376	933	1.57
30	New Hampshire	10	33	124	381	1,291	2.17
31	New Jersey	21	64	401	884	1,174	1.98
32	New Mexico	33	1	55	557	589	0.99
33	New York	62	5	306	2465	2,570	4.33
34	North Carolina	100	4	80	695	1,216	2.05
35	North Dakota	53	1	12	123	763	1.29
36	Ohio	88	13	129	1394	1,686	2.84
37	Oklahoma	77	3	45	660	718	1.21
38	Oregon	36	2	95	660	838	1.41
39	Pennsylvania	67	5	183	1518	1,824	3.07
40	Rhode Island	5	51	210	622	1,031	1.74
41	South Carolina	46	10	87	380	803	1.35
42	South Dakota	66	1	11	148	926	1.56
43	Tennessee	95	5	60	897	852	1.43
44	Texas	254	0	82	3401	2,931	4.93
45	Utah	29	1	77	898	644	1.08
46	Vermont	14	6	43	147	917	1.54
47	Virginia	135	3	52	970	1,237	2.08
48	Washington	39	2	151	1737	1,078	1.82
49	West Virginia	55	6	33	200	720	1.21
50	Wisconsin	72	5	74	940	1,114	1.88
51	Wyoming	23	2	21	82	788	1.33
	계	3,219	0	89	9,519	59,387	100

주: Puerto Rico 지역 제외

자료: 1. 인구 분포 자료 : 미국 센서스국(2000년)

2. CPS 표본 수 자료 : 미국 노동통계국(2007년, 담당자 e-mail)

시계열모형에 의한 월간 CPS 표본 추정량의 형태를 살펴보면 다음

과 같이 두 가지 성분으로 구성되어 있다.

$$y_t = \theta_t + e_t$$

여기서, y_t 는 CPS 추정량, θ_t 는 실제 노동력값, e_t 는 표집오차이다. 노동력값인 θ_t 는 시그널(signal), 표집오차 e_t 는 노이즈(noise)로 이 모형은 시그널 플러스 노이즈(signal plus noise) 형태로 표현된다. 여기서 시그널은 실업자 추정의 경우는 CPS 추정량과 실업보험수급요구자 수를, 취업자 추정의 경우는 상시고용통계(Current Employment Statistics: CES) 프로그램으로부터 얻어진 보조정보와 함께 과거 시계열 자료에 의존하는 시계열모형으로 표현된다. 이 시계열모형과 실제 노동력값의 추정량에 대한 표집오차의 주요 특성을 반영하는 노이즈 모형이 결합하여 소지역 추정에 사용되는 모형(이하 시계열모형)으로 나타난다.

시그널의 구체적 모형은 다음과 같다.

$$\theta_t = \beta_t X_t + T_t + S_t + I_t$$

여기서, X_t 는 설명변수, T_t 는 추세, S_t 는 계절성분, I_t 는 불규칙성분이다. 회귀계수 β_t 는 시간이 변하면서 CPS와 설명변수 사이에 발생하는 변화량을 측정한다. 추세와 계절성분은 설명변수에 의해 설명되지 않는 CPS내의 체계적 변동(systematic variation)을 제어하게 된다. 불규칙성분은 모형의 다른 성분에 의해 설명되지 않은 일시적인 변동들을 설명한다. 시간의 경과에 따른 회귀 계수와 시계열 성분이 변하는 정도는 경험적으로 각 주(state)에 따라 결정된다. 때때로, 추세는 상수, 즉, 고정된 절편처럼 활동하고, 계절 성분은 전년 같은 기간에 대해 고정된 경향을 갖는 것으로 추정되어진다.

미국의 소지역 고용통계는 실업과 취업에 대해 서로 다른 설명변수 X_t 를 이용한다. 취업 모형에서는 CES 프로그램으로부터 얻어진 비농업 분야 임금근로자의 월간 주 전체의 추정량을 센서스로부터 얻은 노동

연령의 주별 인구조로 나눈 비율값을 사용하며, 실업모형에서는 주 노동자의 실업보험수급요구자 수의 비율을 사용한다.

시계열모형의 특징 중 하나는 CPS의 표집오차를 반영할 수 있다는 것이다. 시계열모형에서 표집오차 $e(t)$ 는 미국의 소지역 추정 중에 가장 중요한 부분이다. 미국처럼 표본에 연동구조가 포함되어 있는 경우에는 표집오차에 시계열 상관성이 있을 것으로 예상되므로, 이에 따라 $e(t)$ 는 시계열 상관을 갖고, $e(t)$ 의 분산은 시점 t 에 따라 다르다고 가정한다. CPS는 연동표본제를 사용하는데, 이에 따라 현 시점에서의 자료는 전년 또는 전월의 자료와 중첩되고, 표본의 재설계에 따른 표본크기의 변동은 표집오차의 분산 변화를 초래하게 된다. CPS의 이러한 현상은 표집오차에서 강한 양의 자기상관과 이분산성을 야기한다. 그러므로 표집오차 추정 과정에서 이런 특성들이 무시된다면 노동력의 추정량을 심각하게 왜곡시킬 수 있다. 이러한 이유로 표집오차를 CPS의 관측되지 않은 성분을 추정하기 위한 시그널 모형에 결합시키는 것은 중요한 부분이다. 구체적인 모형이나 이론적 설명은 참고문헌(통계개발원, 2008)을 참고할 수 있다. 모형에 대한 추정량은 BLUP(Best Linear Unbiased Predictor)을 이용하며 표집오차의 분산은 일반화된 분산 함수 방법에 의해 추정된다. 시그널에 대한 미지의 초모수는 Kalman Filter 알고리즘을 이용한 최대우도추정방법으로 추정한다(Tiller, 1992).

소지역 추정에 있어서 중요한 부분 중 하나는 벤치마킹에 관한 것이다. 대영역 내에서 추정된 소지역 추정값들의 합계는 대영역의 추정값과 일치해야 하지만, 대개의 경우 일치하지 않는다. 따라서 최종적으로 총계를 일치시키는 벤치마킹 작업을 수행하게 되는데, BLS는 덴톤법을 이용해서 벤치마킹을 수행해 왔다. 그리고 2005년 1월에는 LAUS 프로그램 재설계하면서 주와 보다 큰 지역에 대해서는 실시간 월간 벤치마킹 작업을 하게 되었다(BLS, 2005).

제2절 일본

1. 노동력조사 개요

일본은 노동력 실태 파악을 목적으로 표본가구를 대상으로 1947년 7월부터 본격적으로 매월 노동력조사(Labor Force Survey: LFS)를 실시하고 있다. 조사는 약 3,000여개 조사구에서 약 40,000여 표본가구에 거주하는 15세 이상 인구²⁾(100,000명 정도로 상한 연령은 없음)를 대상으로 하며 매월 말일을 포함한 1주에 조사한다. 노동력조사는 한 개의 조사구를 4개월 연속 조사하고, 8개월 동안 표본에서 제외시켰다가 다시 4개월 동안 조사한 후 표본에서 제외하는 연동표본제를 도입하여 수행하고 있다(통계청, 2003). 따라서 전체 표본 조사구 중에서는 매달 1/4씩이 교체되며 전체 조사 가구 수에서는 1/8씩 변동이 생긴다. 그러나 연속 조사가 이루어지는 4개월 동안 조사구 내, 표본가구들의 절반은 앞의 2개월 동안, 나머지 반은 뒤의 2개월 동안에 조사가 이루어진다. 따라서 가구를 기준으로 보면 실제로는 2개월 연속 조사 후 10개월 동안 표본에서 제외되었다가 다시 2개월 동안 조사가 수행되는 2-10-2 연동체계이다(일본 통계국). 일본의 노동력조사는 공식적으로 전국(월별, 분기별, 반기별, 연별) 및 10개 총 지역³⁾(표본설계 단위, 분기별, 연별)의 결과를 공표하고 있다.

노동력조사의 표본 추출 과정을 간단히 살펴보면 다음과 같다. 이 조사는 층화2단 추출법으로 표본을 뽑는데, 조사구를 1차 추출단위로 하고, 조사구 내 가구를 2차 추출단위로 한다. 우선 각 10개 표본권역별로, 권역 내의 모든 조사구를 국세조사에 근거한 특성으로 층을 나눈 후 각 지역의 층별로 정해진 추출율과 난수번호를 이용해서 조사구를 계통추출한다. 조사구 내에서 가구 표본 역시 1개 조사구 당 약 15가구가 되도록 계통추출한다. 이처럼 일본은 행정구역단위인 도도부현 단위로 표본 설계를 하지 않고, 이들 도도부현을 포함하는 보다 넓은 10개의 광역권의 통계 공표를 목적으로 표본 설계를 하였다. 그러므로 일본은 소지역을 도도부현 단위로 정하고, 이에 대해서 소지역 추정법을 적용하고 있다.

2) 시설가구 내 거주자, 외국외교관, 외국군대의 군인 및 그 가구원 등은 제외

3) 홋카이도, 토호쿠, 북관동·카츠노부, 남관동, 호쿠리쿠, 토카이, 킨키, 추코쿠, 시코쿠, 규슈

2. 소지역 고용통계 발전 과정

2000년 전후, 일본 역시 다른 나라와 마찬가지로 국가의 실업정책에 대한 관심과 함께, 보다 작은 지역단위의 고용통계 자료에 대한 수요가 증가하였다. 그래서 일본 통계국은 당시 10개의 큰 권역과 함께 47개 도도부현 단위에 대해서도 고용통계를 작성하기로 하였다. 도도부현별 결과에 대해서는 2002년부터 참고적으로 비추정치에 의한 연평균 결과(시산치) 공표를 시작하였다. 이를 계기로 시계열 회귀모형에 의해 추정치 의 정도를 향상시킨 다음, 2006년 5월부터 새로운 사분기 평균 결과(모형 추계치)를 공표하고 있다. 모형에 의한 추估치는 1997년 이후부터 노동력인구, 취업자 수, 완전실업자 수, 비노동인구 수, 완전실업률 항목에 대해 도도부현별 사분기 및 연평균 결과를 인터넷을 통해 참고자료의 형태로 공표하고 있다.

그러나 도도부현이라는 행정구역 단위가 상위 10개 권역에 비해 작기 때문에 소지역이라고는 할 뿐, 이를 우리나라나 다른 나라들과 비슷한 개념의 소지역이라고 보기는 어렵다. 왜냐하면 이들 도도부현의 경우 최소 인구를 가지는 돗토리현의 인구가 60만 명이 넘는 등 일반적으로 인구 규모가 100만 명 이상이고, 표본 조사구도 최소 16개(조사구당 약 15가구 포함) 이상을 확보하고 있기 때문이다.

예를 들어 일본이 발표한 소지역 고용통계 결과에 따르면 돗토리현의 2008년 1분기(1~3월) 평균 15세 이상 인구는 518,000명, 노동력 인구는 308,000명, 그 중 취업자와 실업자는 각각 295,000명, 13,000명으로 완전실업률이 4.2%로 발표되었다. <표 4-2>는 일본의 도도부현별 인구 및 노동력조사의 표본 분포 현황을 나타낸 것으로, 이와 같은 현상을 확인할 수 있다. 이처럼 일본의 소지역은 어느 정도 이상의 표본 수를 확보하고 있고 인구 규모 또한 작지 않다. 그래서 일본은 일정 정도 표본 수를 확보할 수 있는 지역(표본조사구 수 100개 이상)⁴⁾에서는 비추정량(직접추정량)을 그대로 공표하고 있다. 이는 어느 정도의 표본 수가 확보된 다면 굳이 간접적인 방법으로 추정할 필요 없이 직접추정량을 사용할

4) 홋카이도, 사이타마, 가나가와, 아이치, 오사카, 효고, 후쿠오카, 오키나와

수 있음을 시사하며 이에 대한 연구도 소지역 추정의 주요 과제가 될 수 있을 것으로 보인다(통계개발원, 2008).

김서영과 권순필의 출장보고에 따르면 일본에서 소지역 추정 결과는 도도부현별 지역 경제 동향을 파악하는 데 소비자 물가지수와 함께 필립스 곡선을 찾기 위한 주요 지표로 사용되고 있다(통계개발원, 2007). 그러나 도도부현보다 더 작은 시정촌 단위(우리의 시군구 수준)에서 고용통계를 추정할 경우에는 표본규모가 너무 작아진다는 이유에서 아직까지는 소지역 추정 통계를 생산할 계획이 없음을 확인하였다(통계개발원, 2008).

3. 소지역 추정 방법

일본의 소지역 추정 방법은 연동표본의 영향을 반영할 수 있는 시계열모형이다. 이는 미국의 소지역 추정 모형을 벤치마킹한 것으로 일본의 상황에 맞게 보다 간단하게 재설계하였다. 모형에서 보조정보로는 이웃 지역, 즉 해당 도도부현이 속한 10개 층의 대지역의 실업률을 사용하고 있다. 다른 보조정보로써 고용보험자료와 유효구인배율을 검토한 바 있으나, 이들 정보들은 노동력조사의 실업정보와의 상관성 측면에서 볼 때 유익한 정보는 아닌 것으로 판단한 것 같다(통계개발원, 2008).

시계열 모형은 다음과 같이 5개 요인을 포함하고 있다.

$$Y(t) = X(t)\beta(t) + T(t) + S(t) + I(t) + e(t)$$

모형에서 각각의 요소는 다음과 같은 변동을 나타내고 있다. 회귀항, $X(t)\beta(t)$ 는 각 도도부현의 변화와 도도부현이 속한 상위 지역 트렌드와의 관계를 나타낸다. 트렌드항, $T(t)$ 는 경제 성장 등에 따른 장기적 변동과 경기 순환에 동반되는 변동 등 일정한 주기를 갖는 변동으로 주기가 12개월을 넘는 순환변동 등을 포함한다. 예를 들면 이 항은 경기 후퇴와 회복에 있어서 완전실업자가 증가 또는 감소하는 경향을 말한다. 계절변동항, $S(t)$ 는 12개월을 주기로 하는 계절변동으로, 예를 들어 취업자 수는 신년도가 시작되는 3~4월까지 증가하고, 6~7월에 최고가 되며 연 후반기에 감소하는 움직임이 있다. 불규칙변동항, $I(t)$ 는 과거

석유 파동 및 지진 등 자연재해 등에 의해 발생하는 일시적 변화와 같은 순환변동과 계절변동 이외의 변동으로, 돌발적 상황에 의한 변동 및 경기의 단기적 변동을 말한다. 표집오차, $e(t)$ 는 추계 모형에서 가장 중요한 부분으로, 표집오차의 변동 패턴과 변동의 진폭을 나타낸다. 특히 표본에 연동 구조가 있는 경우에는, 표집오차에 시계열 상관성이 있을 것이라고 가정하고 모형에서 계산한다는 특징이 있다(통계개발원, 2008).

벤치마킹에 있어서 미국의 BLS에서는 모형에 의한 추정치의 연평균과 원래 관측치의 연평균이 일치하도록 벤치마킹을 수행하고 있다. 그래서 일본도 당초는 토탈법이나 미국에서 이용되고 있는 방법으로 벤치마킹을 고려한 바 있지만, 연평균으로 계산함으로써 12월과 익년 1월 추정치에 단차가 발생할 수 있는 등의 문제점이 있기 때문에 최종적으로 벤치마킹을 보류한 상태이다. 이런 모형 추정량의 여러 가지 제약 사항 때문에 일본 통계국은 소지역 추정량 공표 시 사용상의 유의점 등을 포함하여 추정치에 대한 신뢰성 확보를 위한 노력을 꾸준히 하고 있다.

〈표 4-2〉 일본의 도도부현별 인구 및 노동력조사 표본 분포

No	도도부현	인구		노동력조사		10개 층
		인구수 (천명)	구성비 (%)	표본조사 구수*	구성비 (%)	
1	홋카이도	5,628	4.4	178	6.1	홋카이도
2	아오모리	1,437	1.1	34	1.2	토호쿠
3	이와테	1,385	1.1	35	1.2	
4	미야기	2,360	1.8	54	1.9	
5	아키타	1,146	0.9	26	0.9	
6	야마가타	1,216	1.0	30	1.0	
7	후쿠시마	2,091	1.6	55	1.9	
8	이바라키	2,975	2.3	66	2.3	
9	토치기	2,017	1.6	46	1.6	
10	군마	2,024	1.6	50	1.7	
11	사이타마	7,054	5.5	114	3.9	남관동
12	치바	6,056	4.7	96	3.3	
13	도쿄도	12,577	9.8	220	7.6	
14	가나가와	8,792	6.9	148	5.1	
15	니가타	2,431	1.9	75	2.6	후쿠리쿠
16	토야마	1,112	0.9	30	1.0	
17	이시가와	1,174	0.9	41	1.4	
18	후쿠이	822	0.6	23	0.8	북관동· 카츠노부
19	야마나시	885	0.7	21	0.7	
20	나가노	2,196	1.7	52	1.8	
21	기후	2,107	1.6	45	1.6	토카이
22	시즈오카	3,792	3.0	74	2.6	
23	아이치	7,255	5.7	148	5.1	
24	미에	1,867	1.5	40	1.4	

〈표 4-2〉의 계속

No	도도부현	인구		노동력조사		10개 층
		인구수 (천명)	구성비 (%)	표본조사 구수	구성비 (%)	
25	시가	1,380	1.1	24	0.8	킨키
26	교토	2,648	2.1	49	1.7	
27	오사카	8,817	6.9	175	6.0	
28	효고	5,591	4.4	106	3.7	
29	나라	1,421	1.1	28	1.0	
30	와카야마	1,036	0.8	20	0.7	
31	돗토리	607	0.5	16	0.6	츠키쿠
32	시마네	742	0.6	21	0.7	
33	오카야마	1,957	1.5	50	1.7	
34	히로시마	2,877	2.3	77	2.7	
35	야마구치	1,493	1.2	45	1.6	시코쿠
36	도쿠시마	810	0.6	30	1.0	
37	카가와	1,012	0.8	33	1.1	
38	에히메	1,468	1.1	55	1.9	
39	코치	796	0.6	33	1.1	규슈
40	후쿠오카	5,050	4.0	107	3.7	
41	사가	866	0.7	18	0.6	
42	나가사키	1,479	1.2	33	1.1	
43	마모토	1,842	1.4	38	1.3	
44	오이타	1,210	0.9	26	0.9	
45	미야자키	1,153	0.9	27	0.9	
46	카코시마	1,753	1.4	39	1.3	
47	오кина와	1,362	1.1	144	5.0	
계	47개	127,768	100	2,895	100	

주: 1. 표본 수는 2000~2002년 평균 조사구 수입

2. 1조사구당 약 15가구 포함

자료: 인구 분포 자료 : 일본 통계국 국세조사 자료(2005년)

제3절 캐나다

1. 노동력조사 개요

캐나다는 매월 실시하는 노동력조사를 통해 고용동향을 파악한다. LFS는 다단계 층화 표본설계에 의해 선택된 54,000가구에 대해 조사를 수행하고 있으며 연동표본제를 실시하고 있다. 캐나다의 연동표본제는 1/6 교체 시스템으로 표본으로 선정된 가구는 연속되는 6개월 동안 조사에 참여한 후 교체되는 방식이다. 중복 표본 비율은 전월대비 5/6이다. 이 조사는 표본가구 내 거주하는 15세 이상 인구에 대해 15일이 포함된 1주(일요일부터 토요일까지)를 대상으로 실시한다(통계청, 2003).

캐나다는 53개의 소지역으로 구분된 각 고용보험(Employment Insurance : EI) 지구의 실업률을 근거로 하여 실업보험수당 수혜자와 수혜기간을 결정하고 있다. LFS는 일부 EI 지역에서 너무 작은 표본의 크기를 가지고 있어 의미 있는 추정량을 산출할 수가 없다. 그래서 캐나다 통계청에서는 LFS의 표본설계를 이원화함으로써 각 지역의 실업률을 직접 추정하고 있다. 즉 표본설계 첫 단계에서는 전국·주별 통계의 산출에 필요한 적정수의 표본을 배분하고, 두 번째 단계에서는 나머지 표본을 소수의 인구를 가진 EI 지역에 집중 배분함으로써 대지역과 소지역에서 각각 의미 있는 추정량을 끌어내는 설계 기반 추정법을 제공하고 있다.

2. 소지역 고용통계 발전 과정

캐나다는 1970년대 초부터 소지역 추정에 관한 연구를 수행해왔다. 그 적용분야는 매우 다양하여, 고용뿐만 아니라 사업체와 가구조사, 총조사의 과소포함 분야 등 작은 지역(영역) 단위에서의 통계를 생산하기 위해 꾸준한 노력을 해왔다. 특히 고용통계 분야에서는 앞서 기술한 바와 같이 캐나다는 LFS의 표본설계를 이원화함으로써 각 EI 지역의 직접

추정량을 얻기 위한 최대한의 노력을 기울이고 있다. 그러나 이러한 LFS의 주요 목표가 아닌 통계의 산출, 즉 초기 단계에 계획되지 않은 단위에 대한 통계 등이 요구될 때는 시간·공간상의 보조 정보를 이용하는 간접추정법을 이용하고 있다. 이 방법은 국가통계로서의 공식 추정량은 아니지만 사용자로부터 특별한 요구가 있을 때 주로 이용하고 있다. 본 연구에서는 You 등(2003)을 참고하여 캐나다의 고용통계의 소지역 추정방법에 대해 설명하였다.

캐나다에서 간접추정을 이용하는 소지역은 CMA(Census Metropolitan Areas), CA(Census Agglomeration)지역, 경제구역(Economic Region: ER), 고용보험지역으로 구분되고 있다. CMA는 전체 인구가 10만 명 이상이며 그 중 도심 인구가 5만 명 이상인 지역, CA는 도심 인구가 1만 명 이상인 지역, ER은 지역 경제 활동 분석을 위한 표준 지역 단위, EI 지역은 53개의 고용보험 지역이다(Statistics Canada, 2007). 캐나다의 소지역인 EIER과 ER 별 인구 규모와 표본 규모를 <표 4-3>에 정리하였다. EIER은 가장 작은 규모의 지역이 Manitoba주의 41번 EIER로 인구 8만 명에 700여 표본가구를 가지고 있다. 이는 우리의 시(city) 수준의 인구 규모이지만, 표본 수는 훨씬 많다는 것을 알 수 있다. ER은 EIER에 비해 작은 규모를 가지는 경우가 많은데, 가장 작은 ER은 Saskatchewan주의 4760번 ER로 인구 12,000명에 31개 표본가구를 가지고 있다.

〈표 4-3〉 캐나다의 EIER, ER별 인구 및 노동력조사 표본 분포
(단위 : 천 명)

구분	EIER				ER				
	번호	인구	LFS 표본		번호	인구	LFS 표본		
			가구수 (호)	구성비 (%)			가구수 (호)	구성비 (%)	
P r o v i n c e	Newfoundland and Labrador	01	171	664	1.24	1010	243	932	1.75
		02	329	1,280	2.40	1020	42	195	0.37
						1030	100	380	0.71
						1040	115	437	0.82
	소계		500	1,944	3.64				
	Prince Edward Island	03	135	1,378	2.58	1110	135	1,378	2.58
	Nova Scotia	04	181	615	1.15	1210	142	486	0.91
		05	369	1,115	2.09	1220	156	457	0.86
		06	349	1,143	2.14	1230	120	369	0.69
						1240	122	385	0.72
						1250	359	1,176	2.20
	소계		899	2,873	5.38				
	New Brunswick	07	351	1,236	2.32	1310	167	578	1.08
		08	119	647	1.21	1320	181	639	1.20
		09	251	871	1.63	1330	167	643	1.20
						1340	124	452	0.85
						1350	83	442	0.83
	소계		721	2,754	5.16				
	Quebec	10	147	577	1.08	2410	95	409	0.77
		11	682	672	1.26	2415	201	403	0.76
		12	137	651	1.22	2420	638	637	1.19
13		146	638	1.20	2425	383	817	1.53	
14		152	793	1.49	2430	286	1,006	1.88	
15		467	747	1.40	2433	218	249	0.47	
16		3429	2,199	4.12	2435	1278	1,219	2.28	
17		949	924	1.73	2440	1814	1,153	2.16	
18		234	498	0.93	2445	343	271	0.51	
19		449	727	1.36	2450	387	286	0.54	
20		258	758	1.42	2455	461	369	0.69	
21		155	589	1.10	2460	315	862	1.62	
					2465	142	329	0.62	
				2470	252	753	1.41		
				2475	276	754	1.41		
				2480	89	206	0.39		
				2490	26	50	0.09		
소계		7,204	9,773	18.31					

〈표 4-3〉의 계속

(단위 : 천 명)

구분	EIER				ER				
	번호	인구	LFS 표본		번호	인구	LFS 표본		
			가구수 (호)	구성비 (%)			가구수 (호)	구성비 (%)	
P r o v i n c e	Ontario	22	823	872	1.63	3510	1123	1,502	2.81
		23	401	841	1.58	3515	422	1,040	1.95
		24	146	677	1.27	3520	337	433	0.81
		25	1,028	1,259	2.36	3530	4996	3,205	6.01
		26	297	653	1.22	3540	1062	1,660	3.11
		27	4,748	2,640	4.95	3550	1274	2,282	4.28
		28	665	758	1.42	3560	587	1,204	2.26
		29	377	847	1.59	3570	606	1,347	2.52
		30	416	872	1.63	3580	279	595	1.11
		31	255	724	1.36	3590	537	1,371	2.57
		32	301	648	1.21	3595	210	777	1.46
		33	416	702	1.32				
		34	314	719	1.35				
		35	501	1,056	1.98				
		36	150	701	1.31				
		37	120	608	1.14				
		38	476	839	1.57				
			소계	11,433	15,416	28.88			
Manitoba	30	671	1,917	3.59	4610	83	359	0.67	
	40	311	1,017	1.91	4620	52	189	0.35	
	41	81	727	1.36	4630	101	314	0.59	
					4640	44	156	0.29	
					4650	622	1,800	3.37	
					4660	79	356	0.67	
					4670	42	163	0.31	
					4680	40	324	0.61	
	소계	1,063	3,661	6.86					
Saskatchewan	42	193	902	1.69	4710	267	1,183	2.22	
	43	226	799	1.50	4720	104	472	0.88	
	44	314	1,289	2.42	4730	282	1,011	1.89	
	45	195	790	1.48	4740	84	339	0.64	
					4750	179	744	1.39	
					4760	12	31	0.06	
	소계	928	3,780	7.08					

3. 소지역 추정 방법

〈표 4-3〉의 계속

(단위 : 천 명)

구분	EIER				ER				
	번호	인구	LFS 표본		번호	인구	LFS 표본		
			가구수 (호)	구성비 (%)			가구수 (호)	구성비 (%)	
P r o v i n c e	Alberta	46	971	1,206	2.26	4810	234	303	0.57
		47	938	1,392	2.61	4820	179	439	0.82
		48	199	1,010	1.89	4830	1041	1,291	2.42
		49	840	1,808	3.39	4840	72	268	0.50
						4850	151	333	0.62
						4860	974	1,476	2.77
						4870	209	766	1.44
						4880	88	540	1.01
	소계		2,948	5,416	10.15				
	British Columbia	50	591	954	1.79	5910	672	1,542	2.89
		51	148	573	1.07	5920	2270	3,154	5.91
		52	1983	2,346	4.40	5930	447	697	1.31
		53	307	914	1.71	5940	144	260	0.49
		54	491	834	1.56	5950	158	327	0.61
55		319	756	1.42	5960	52	149	0.28	
					5970	39	89	0.17	
					5980	58	166	0.31	
소계		3,839	6,377	11.95					
계		29,670	53,372	100.00					

자료: Methodology of the Canadian Labour Force Survey(2008)

캐나다 통계청은 약 20여년 동안 소지역추정을 위해, 표본설계당시의 표본크기에 의존하는 표본크기의존(Sample Size Dependent: SSD) 추정량을 사용해왔다. 이 추정량은 LFS자료에 근거하며, 소지역에 대한 일반화 회귀추정량과 합성추정량의 가중평균의 형태를 갖는 복합추정량의 일종이다(Drew 등, 1982). 이러한 추정량은 모형에 의존하는 방법들에 비해 간단한 반면 사용되는 보조정보에 대해 상당한 신뢰성을 필요로 하기도 한다. 이와 같은 방법 적용은 캐나다의 LFS가 많은 수의 표본을 확보하고 있기 때문에 가능한 것으로 보인다⁵⁾.

5) 캐나다의 LFS 표본가구 수는 약 54,000가구이며 이는 인구수 대비 표본 수를 비교해 보면 우리나라의 약 2배가 넘는 수준이다.

그러나 최근에는 시간과 공간 정보를 모두 활용할 수 있는 시계열·횡단면 모형(Time series & Cross-sectional Model)을 이용한 추정 연구가 활발하다. 이 방법은 횡단면과 종단면으로부터의 모든 정보를 이용할 수 있다는 장점을 가지고 있다. 캐나다의 표본제는 매월 1/6의 표본이 새로운 표본으로 교체되며, 두 연속되는 월에서는 표본의 5/6가 동일한 연동표본제로, 매월 중복된 표본으로 인해 자료들 간에 상관성이 생기게 된다. 따라서 표본 중복으로 인한 상관성을 추정에 반영할 필요가 제기되었고, 이는 지역뿐만 아니라 시간으로부터도 정보를 빌려옴으로써 어느 정도는 해결할 수 있을 것으로 보았다

이와 같은 시계열·횡단면 모형은 Fay-Herriot 모형을 다변량 모형으로의 확장한 형태이다. 보조정보로는 고용보험 수혜율과 이웃 지역의 정보 등이 이용되고 있다. 모형을 좀 더 자세히 살펴보면, y_{it} 는 시점 t 에서 i 번째 CA(소지역)의 실업률의 참값 θ_{it} 의 직접 LFS 추정값이다.

$$y_{it} = \theta_{it} + e_{it}$$

여기서 e_{it} 는 지역 i 내에서 t 시점에 대한 표집오차이며, LFS의 표본연동에 의해 소지역 내에서 짧은 시간에 걸쳐 상당한 표본이 중복되므로 e_{it} 와 e_{is} ($t \neq s$) 간의 상관은 반드시 고려되어야 한다. θ 는 실업률에 대한 참값으로 보조변수와 랜덤 지역효과를 포함한 선형회귀모형을 사용하여 추정할 수 있다. 구체적 모형은 다음과 같다.

$$\theta_{it} = \mathbf{x}_{it}' \boldsymbol{\beta} + \nu_i + u_{it}$$

여기서, \mathbf{x}_{it} 는 보조자료, $\boldsymbol{\beta}$ 는 회귀계수, ν_i 는 랜덤 지역효과, u_{it} 는 랜덤 시간성분이다. u_{it} 는 표본의 교체 형식과 조사기간에 따라 확률보행과정(random walk process)과 정상자기회귀모형(stationary autoregressive model)을 통해 추정할 수 있다(통계청, 2004).

벤치마킹에 있어서 캐나다는 주로 대영역의 직접추정값을 소지역들의 상대적 간접추정값들의 비율로 재배분하는 방법을 이용한다. 즉, 소지역 i 에 대한 총계 y_a 의 추정량을 \bar{y}_a , 해당 소지역 i 를 포함하는 대지역 I 에 대한 \bar{y}_a 의 합계를 $\bar{y}(A)$ 라 하자. 이때 소지역 추정값들은 다음과 같은 \bar{y}_a^{ADJ} 를 통해 보정된다.

$$\bar{y}_a^{ADJ} = \frac{\bar{y}_a}{\bar{y}(A)} \hat{y}(A)$$

여기에서, $\bar{y}(A) = \sum_{a \in A} \bar{y}_a$ 이다(통계청, 2002).

제4절 영국

1. 노동력조사 개요

영국의 노동력조사는 분기별로 조사되며 컴퓨터보조개별면접(Computer Aided Personal Interviews: CAPI)조사로 실시된다. 표본가구 내 16세 이상 인구를 대상으로 상한 연령은 없으며 기숙사 및 국립보건수용시설 등에 거주하는 사람도 포함하여 조사한다. 표본 수는 2007년에 약 53,000가구에 거주하는 138,000명 정도이고, 표본을 연속 5분기 조사 후 완전 교체하는 연동표본제를 사용하고 있다(통계청, 2003).

표본 설계는 면접지역(interview area)을 <표 4-4>의 3개의 권역에 할당하는 방법을 사용하고 있다. 즉, England에는 162개, Wales에는 26개, Scotland에는 24개로 총 212개의 면접지역을 우선 할당한다. 할당지역에서 가구를 선정하는 과정에는 표본틀 구성을 위해 우편번호 코드와 국가 보건 서비스 시설(National Health Service)의 등록된 명부의 2개 정보를 활용하고 있다. 이처럼 영국의 LFS는 소지역 단위인 UA/LADs(Unitary Authority/Local Authority Districts) 수준에서 표본추출을 한 것이 아니기 때문에 이들 소지역에 대해서 직접 추정치를 사용하는 것은 신뢰성 측면에서 바람직하지 않다. 따라서 영국은 UA/LADs 수준에서의 통계는 소지역 추정방법을 사용하여 생산하고 있다(ONS, 2006).

<표 4-4> 영국의 UA/LAD별 인구 및 노동력 조사 표본 분포

구분	UA		LAD		
	개	인구 분포(천명)	개수	인구 분포(천명)	제외

		수	최소	평균	최대		최소	평균	최대	지역
England	North East	5	89	128	178	18	24	104	281	
	North West	4	118	147	191	39	50	157	439	
	Yorkshire And The HumberR	5	153	210	314	16	47	244	715	
	East Midlands	4	34	200	280	36	48	94	194	
	West Midlands	3	158	191	240	31	37	151	977	
	East	4	143	161	184	44	56	108	166	
	London					33	147	223	331	1
	South East	12	110	170	249	55	60	108	169	
	South West	9	130	204	381	36	35	88	151	1
Wales					22	56	132	305		
Scotland					32					
계	46	34		381	362	24		977	2	

자료: 2001년 센서스 자료

2. 소지역 고용통계 발전 과정

영국은 406개의 UA/LADs를 가지고 있다. UA/LADs는 우리나라의 시군구(city 또는 town 개념)와 유사한 지방 행정 조직을 말한다. 영국의 2001년 센서스 조사 결과에 의한 UA/LADs의 인구규모는 <표 4-4>와 같다. LAD 가운데 제외 지역은 런던시와 실리제도이다. 이 두 개의 지방 정부는 노출의 문제로 소지역 추정에 포함하지 않았다. 소지역 추정 결과에 대한 발표 사례를 보면, 가장 작은 인구 규모를 가지고 있는 지방은 North East의 Teesdale로 인구가 약 24,000명 정도의 소도시이다. 이 지역의 2007

년도 실업자 수는 500 ± 100 명, 실업률은 $4.0 \pm 1.0\%$ 로 발표되었다.

2000년 이후 LFS 표본은 각 지역마다 적어도 최소한의 경제활동인구가 표본에 포함되도록 증가되어 왔다. 이 자료들은 국제노동기구(International Labour Organisation: ILO) 정의에 맞는 실업 통계를 연간에 기초하여 측정하는 데 사용된다. 그러나 소지역 내의 표본 크기는 종종 당시의 공표 규칙 하에서 신뢰할 만한 추정량을 제공하기에는 너무 작았다. 이후 신뢰도에 대한 허용 규칙이 좀 더 완화되긴 했지만, 여전히 많은 지방 정부들은 실업에 대해서는 공표할 만한 어떠한 추정량도 없으며, 추정량이 있는 경우라도 상당수는 정도(precision)가 부족했다. 이런 배경 하에서, 영국의 통계청(Office for National Statistics: ONS)은 사우스햄턴대의 Ray Chambers 교수에 의해 개발된 모형을 적용하여 UA/LADs 수준에서 실업자 수와 실업률에 대한 추정량들을 생산하고자 하였다(ONS, 2006).

ONS가 Ray Chambers 교수와 함께 개발한 새로운 방법은 1996/97년부터 2003/04년뿐만 아니라 2005년까지 포함하는 과거의 자료를 이용하여 소지역의 실업률과 실업자 수를 관련 신뢰구간과 함께 모형에 기반하여 추정한다. 이들 추정량은 학문적 리뷰와 타당성 검토를 수차례 거친 후 현재 국가통계로서 인정받고 있다. 또한 ONS는 소지역 추정량의 사용에 있어서 제약사항에 대한 사용자 가이드와 함께, 연간자료를 분기별로 갱신하여 제공하고 있다(ONS, 2006). 공표 시 실업자 수와 실업률과 각각의 95% 수준에서의 신뢰구간도 제공하고 있다. 실업자 수에 대한 추정량은 100명 단위로, 실업률은 소수점 첫째 자리까지 공표하고 있다.

3. 소지역 추정 방법

ONS의 모형기반 방법은 LFS에 의해 측정되어진 실업과 보조정보(일반적으로 센서스나 행정자료) 사이의 강한 관계에 의존한다. 이런 관계는 UA/LADs를 위한 실업에 대해 좀 더 신뢰성 있는 추정량을 제공하게 한다. 주요 보조정보는 구직자수당의 수령자 수(claimant count: 요구자

6) 지역 교육 정부 및 런던 자치구를 위한 지역으로 UA/LADs를 의미하지는 않음

수)이다.

ONS가 사용하는 모형은 지역임의효과 모형(Area Random Effects Model)이다. 이 모형은 일반적으로 고정효과모형에 비해 더 정확한 추정을 할 수 있다고 알려져 있다. 그리고 이 모형기반추정량은 LFS에서 생산되어진 직접추정량들보다는 항상 더 정확한 것으로 평가되었다.

영국에서 사용하고 있는 지역임의효과 모형의 구체적인 형태는 다음과 같다. 우선 보조정보로는 추정하고자 하는 기간에 얻어진 매월 실업 급여 요구자 수 자료에 대한 연평균값을 이용한다. 모형은 로짓모형으로 연령별×성별 그룹(남성/여성, 16~24세/25~49세/50세 이상) i 와 LAD j 에 대한 실업률 p_{ij} 를 추정하기 위한 형태이다. 모형에서 예측변수로는 UA/LAD 내에서 다음과 같은 변수들을 사용한다. ① 각 UA/LAD내의 각 연령별/성별 그룹에서의 요구자 수 비율에 대한 로짓⁷⁾, ② UA/LAD내의 요구자 수 비율에 대한 로짓, ③ 성별/연령별 그룹, ④ 12개 정부 청 지역(Government Office Regions: GORs), ⑤ 영국 지방 권역에 대한 7개 상위그룹(2001년 지역 분류하⁸⁾)에서의 7개 supergroup들, cluster라고도 불림).

모형의 결과 변수는 각 연령별×성별 그룹 내에서의 16세 이상 인구의 비율에 대한 LFS 직접추정량의 로짓으로, 구체적인 모형은 다음과 같다(ONS, 2006).

$$\begin{aligned} \text{logit}(p_{ij}) = & \beta_0 \\ & + \beta_{1-5}[\text{dummy age/sex variables}] \\ & + \beta_{6-16}[\text{dummy GOR variables}] \\ & + \beta_{17-22}[\text{dummy Cluster variables}] \\ & + \beta_{23}[\text{logit}(\text{claimant count proportion}_{ij})] \\ & + \beta_{24-28}[\text{logit}(\text{claimant count proportion}_{ij}) \\ & \quad \cdot [\text{dummy age/sex variables}]] \\ & + \beta_{29}[\text{logit}(\text{claimant count proportion}_j)] \\ & + U_j(\text{area random effect term}_j) \end{aligned}$$

7) $\text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$

8) National Statistics 2001 Area Classification for Local Authorities.

http://www.statistics.gov.uk/about/methodology_by_theme/area_classification

이것은 예측변수들과 함께 지역임의효과 항, U_j 를 포함하고 있다. 즉 보조 정보에 의해 설명되지 않는 지역 간 변동을 모형에 포함시킨 것이다. 모형 내에 지역임의효과를 포함하면 추정량은 직접추정량과 고정효과추정량의 가중평균의 형태가 되는데 이는 사실상 표본 크기가 증가함에 따라 직접추정량에 좀 더 많은 가중을 주게 된다. 이것은 표본의 크기가 커짐에 따라 직접추정량으로의 수렴을 보장하고, 직접추정량이 모형추정량보다 정도가 높을 때, 추정량의 선택을 쉽게 한다(ONS, 2006).

모형에 사용되는 주요자료는 LFS와 실업급여 요구자 수 자료이다. LFS 자료는 연간 자료가 사용된다. 이것은 앞선 4분기(예를 들면, 3월부터 2월까지 12개월을 포함하는 추정을 하기 위해서 앞선 분기들 3-5월, 6-8월, 9-11월, 12-2월)로부터 얻은 자료)로부터 얻은 독립적인 자료들로 구성되어 있다. 실업급여 요구자 수 자료는 같은 연간 기간 동안에 월 실업급여 수급자수의 평균이다. 이 자료는 모형에서 두 개의 항에 의해 삽입된다. 하나는 성별/연령별과 지역에 따른 실업과 실업급여 요구자 수 사이의 관계에 대한 관측자료에 반영하기 위한 것이고, 다른 하나는 하나의 그룹으로써 UA/LAD의 수준에서 이들 지역에 속한 전체적 고용 환경은 UA/LAD에 속한 다른 그룹의 고용 전망에 영향을 줄 것이라는 측면을 반영하기 위한 것이다(ONS, 2006).

영국 역시 대지역의 직접 LFS 추정량에 소지역 추정량들의 합을 맞추기 위해 벤치마킹을 실시하고 있다. 모형기반 방법은 UA/LAD 추정량이 LFS로부터 나온 직접추정량에 맞춰지도록 개발되어왔다. 예를 들면, 웨일즈의 UA/LAD 수준의 모형 기반 추정량은 웨일즈의 연간 실업 수준 LFS 추정량에 부합한다(ONS, 2006).

제5절 요약

9) 이것들은 이전에 사용된 계절 분기이다. 현재 사용되는 달력 분기는 1-3월, 4-6월, 7-9월, 10-12월이다.

지금까지 소지역의 고용통계를 생산하기 위해 추정 기법을 사용하고 있는 국가들 중 미국, 일본, 캐나다와 영국의 사례를 살펴보았다. 각국가별로 소지역에 대한 정의는 달랐지만, 대체적으로 기존의 공표 단위보다는 하위 지역이면서 인구나 표본 규모는 크지 않은 지역을 대상으로 하고 있다는 것을 알 수 있다. 이를 위해 직접추정량보다는 각국의 상황에 적합한 추정 방법을 채택하고 있는데, 즉, 미국과 일본은 시계열 모형, 캐나다는 시계열·횡단면 모형 그리고 영국은 로짓 모형을 사용한다. 모형기반 소지역 추정은 어떤 모형이나에 관계없이 모두 좋은 보조정보를 필요로 하고, 이를 위해서 대부분의 국가는 실업보험 자료나 이웃 지역(상위 지역)의 정보를 이용하고 있다. 미국은 다른 나라와 달리 취업과 실업에 대해 서로 다른 보조정보를 이용하여 모형을 적용한 반면, 일본, 캐나다 등은 취업과 실업에 대해 특별히 다른 모형을 적용하지는 않았다. 일본을 제외한 미국, 캐나다, 영국 세 나라에서는 소지역 추정치를 상위 지역 직접추정치에 대해 벤치마킹을 적용하는 것으로 파악되었다. 캐나다를 제외하고는 추정 결과를 인터넷이나 책자를 통해 외부에 공표하고 있었으며 사용자를 위한 가이드도 같이 제공하고 있었다.

추가적으로 호주는 2008년 현재 고용통계에 대한 소지역 추정 결과를 공표할 예정에 있는 것으로 파악되었다. 간단히 설명하자면 호주는 414개의 지역정부(Local Government Areas: LGAs)를 가지고 있으며 LFS 전체 표본 규모는 57,000여 개이다. 평균 LGA 표본크기는 256개로, 성별과 5개의 연령별 그룹을 가지고 있다. 이때 5개의 연령별 그룹은 15-24, 25-34, 35-44, 45-54, 55-64세 그룹이다. 성별/연령별 그룹에 대한 표본 규모는 평균적으로 약 13개 정도가 포함된다.

사례를 통해 본 결과 다른 많은 국가들은 소지역의 고용통계 생산을 위해서 짧게는 5년에서 길게는 50여년 동안 관련 방법과 정보를 개선하고 보완해왔다. 우리나라의 소지역 추정 연구 시작을 90년대 후반이라고 볼 때, 지속적인 연구 활동과 관심이 적었다는 점에서 상대적으로 추정량의 개발·개선이 늦어진 경향이 없지 않다. 이런 면에서 각국의 연구와 적용사례는 우리에게 시행착오를 줄이면서 나아갈 방향을 모색하는데 좋은 귀감이 될 수 있다. 우리나라는 현재 학계를 중심으로는 소

지역 추정 방법에 관한 연구가 활발히 진행되고 있다. 국가 통계 측면에서는 통계개발원을 중심으로 한창 연구 중에 있으며 조만간 우리나라에 적합한 추정방법을 찾을 수 있을 것으로 본다.

<표 4-5>는 독자의 이해를 돕기 위해 각국의 고용통계에 대한 소지역 추정 현황을 간략하게 비교 정리한 것이다. <표 4-5>에서 보는 것처럼 어느 나라에서 채택하고 있는 방법이 가장 좋다는 식의 결론을 내리기는 어렵다. 그러나 자국의 상황에 비추어 또는 이용 가능한 자료의 정보에 의해서 추정 방법을 선택하고 있다는 것만은 확실한 것 같다. 고용보험 등 복지정책이 잘 되어 있고 이에 대한 통계자료 구축이 잘 되어 있는 나라일수록 보조정보로부터 보다 정확한 추정치를 구할 수 있을 것으로 보인다. 그리고 현 시점의 자료가 현 시점의 이웃 지역들 간의 상관관계보다는 과거 자료에 더 많은 영향을 받을 것이라고 판단할 경우에는 시계열모형에 의존하는 경향이 있을 것이고, 반대로 이웃 지역들 간의 노동형태라든지 지역특성이 상당히 유사하다고 판단되는 경우에는 로짓 모형이라든가 Fay-Herriot 모형과 같은 일반적인 회귀모형 형태를 선호하게 될 것이다.

현재 우리의 소지역 추정에 있어서 어떤 방법으로 추정을 할 것인가에 대한 문제는 신중한 판단을 요한다. 이러한 판단은 다양한 모형에 의한 방법들을 충분히 검토한 후에 결정되어야 한다. 최소한 국가통계로서 소지역 추정 통계가 신뢰를 얻기 위해서는 일반적으로 적용가능한 방법인지, 재현가능한지, 사용하기 쉬운지, 정부통계로서 적용가능한지, 추정 결과가 보편타당한지, 현실적으로 얼마나 실용성이 있는지 등의 측면에서 검토가 이루어져야 할 것이다. 이를 위해서는 학계 및 통계청 내·외 다양한 분야의 전문가들의 의견 수렴이 있어야 하며, 이는 추정결과에 대한 신뢰성과 설명력을 확보하기 위한 확실한 장치가 될 수 있다.

〈표 4-5〉 각국의 고용통계에 대한 소지역 추정 현황

	한국	미국	일본
소지역 정의	모든 시군구 (2008년 현재 230개)	주, DC, Puerto Rico, 대도시, 작은 노동시장 지역, 카운티, 뉴잉글랜드에 있는 도시들, 인구 25,000명 이상의 도시 7,000여개 지역	도도부현 47개
인구규모	1만 명 이상	-	60만 명 이상
고용통계 조사	경제활동인구조사	상시인구조사	노동력조사
연동체계	매월 1/36 교체	4-8-4, 4개월 조사, 8개월 중지, 4개월 재조사	2-10-2, 2개월 조사, 10개월 중지, 2개월 재조사
모 형	-	시계열 회귀 모형	
보조정보	-	실업보험자료(UI), 경상고용통계조사자료(CES)	대지역 정보*
공표여부	-	○	○
추정범위	경찰인구 수, 취업자 수, 실업자 수, 실업률 등		
기간	-	연평균	분기, 연평균
벤치마킹	-	○	×
중 점 연구기간	1996년 연구 시작 2002년 경찰조사, 2004년 가계조사 용역 발주 연구	50여년 (1950 ~ 1996)	5년 (2002 ~ 2006)
활 용	-	경제분석, 노동시장 연구, 연방 기금 배정 등	지역경제현황과악 소비자 물가지수와 함께 필립스 곡선을 찾는 데 활용
기 타	연구 중	-	

* 대지역 정보란, 해당 소지역을 포함하는 상위 지역의 실업(률) 정보를 이용하는 것

자료 : 1. 미국 공표 사이트 : <ftp://ftp.bls.gov/pub/special.requests/la/laucnty07.txt>

2. 일본 공표 사이트 : <http://www.stat.go.jp/data/roudou/pref/index.htm>

〈표 4-5〉의 계속

	캐나다		영국
	조사	추정	
소지역 정의	53개의 EI 지역	CA, CMAs	UA/LADs 406개
인구규모	-	1만 명 이상	2만 명 이상
고용통계 조사	노동력조사		
연동체계	1/6 교체. 6개월 조사 후 완전 교체		1/5 교체. 5분기 조사 후 완전 교체
모 형	× (표본설계 이원화)	횡단면 및 시계열모형	지역 랜덤 효과 모형(로짓모형)
보조정보	×	고용보험(EI), 이웃 지역정보	실업수당요구자 수(율)
공표여부	○	×	○
추정범위	경활인구 수, 취업자 수, 실업자 수, 실업률 등		실업자 수, 실업률
기간	월평균	-	연평균(분기별 갱신)
벤치마킹	×	○	○
중 점 연구기간		1970년대 초 ~	2000* ~ 2006
활 용	예산 배분, 정책 결정		-
기 타	조사통계가 공식통계이나, 내부이용 또는 수요가 있을 시는 소지역 추정 통계를 이용함		공식통계

* 영국의 중점 연구 기간은 모형 기반 추정량 연구 기간 표시

자료 : 3 영국 공표 사이트 :

<http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=13574&Pos=&ColRank=1&Rank=272>

참고문헌

- 통계청(2002), “소지역 추정법에 의한 시·군·구 실업통계 개발”, 통계청 용역 연구 보고서.
- 일본 통계국, “소지역 추계와 노동력 조사에 적용”, <http://www.stat.go.jp/data/roudou/pref/index.htm>.
- 통계개발원(2008), “소지역 추정 연구를 위한 일본 통계국 출장 결과”, 출장보고서.
- 통계청(2003), “외국 고용통계 자료집”(내부자료).
- _____ (2004), “모형기반 소지역 추정 방법”(내부자료).
- Bell, W.R. and Hillmer, S.C.(1990), "The time series approach to estimation for repeated surveys", *Survey Methodology*, 16, 195-215.
- Drew, J.D. Singh, M.P. and Choudhry, G.H.(1982). "Evaluation of small area techniques for the Canadian Labour Force Survey", *Survey Methodology*, 8, 17-47.
- ONS, “National Statistics 2001 Area Classification for Local Authorities”, http://www.statistics.gov.uk/about/methodology_by_theme/area_classification.
- ONS(2003), "Development of improved estimation methods for local area unemployment levels and rates", Technical Report. Labour Market trends. January, 37-43.
- ONS(2006), "Model-Based Estimates of ILO Unemployment for UA/LADs in great Britain Guide for Users".
- Scott, A.J. and Smith, T.M.F.(1974), "Analysis of repeated surveys using time series methods", *Journal of the American Statistical Association*, 69, 674-678.
- Statistics Canada(2007), “Guide to the labour force survey”, Catalogue no. 71-543-G IE.
- Statistics Canada(2008), “Methodology of the Canadian Labour Force Survey”, Catalogue no. 71-526-X.
- Tiller, R.(1992), "Time series modeling of sample survey data from the U.S. current population survey", *Journal of Official Statistics*, 8, 149-166.

- U. S. Bureau of Labor Statistics(2005), "Local Area Unemployment Statistics", <http://www.bls.gov/lau/publications.htm>.
- U. S. Census Bureau, Data Sets, <http://www.census.gov/>.
- U. S. Bureau of Labor Statistics(2004), "Proposed improvement in estimating and benchmarking state labor force estimates", OECD Short-term Economics Statistics Expert Group meeting in Paris, France.
- You, Y., Rao, J.N.K. and Gambino, J.(2003), "Model-Based Unemployment Rate Estimation for the Canadian Labour Force Survey: A Hierarchical Bayes Approach", *Survey Methodology*, 29, 25-32.