

Comparison of Small Area Estimations by Sample Sizes¹⁾

Jung-O Kim²⁾ and Key-Il Shin³⁾

Abstract

Model-based methods are generally used for small area estimation. Recently Shin and Lee (2003) suggested a method which used spatial correlations between areas for data set including some auxiliary variables. However in case of absence of auxiliary variables, Direct estimator is used. Even though direct estimator is unbiased, the large variance of the estimator restricts the use for small area estimation. In this paper, we suggest new estimators which take into account spatial correlation when auxiliary variables are not available. We compared Direct estimator and the newly suggested estimators using MSE, MAE and MB.

Keywords : Direct estimator; spatial correlation; SAR model.

1. 서론

최근 소지역 추정에 관한 연구가 활발히 진행되고 있다. 미국이나 캐나다 등에서는 센서스 중간 연도에 지방자치 정부의 요청에 따라 센서스 자료나 일반 행정자료를 이용한 소지역 추정을 활용하여 인구 추정 및 노동력 추정 등을 하고 있다. 최근 국내에서도 이계오 (2002)가 소지역 추정을 이용한 실업 통계에 관한 분석 보고서를 발표하는 등 소지역 추정에 관한 논문과 분석이 발표되고 있다. 소지역 추정법은 크게 자료기반 추정법과 모형기반 추정법으로 구별될 수 있으며 최근 모형기반 추정법에 관한 연구가 활발히 진행되고 있다. 신기일과 이상은 (2003)은 보조변수가 있는 경우 공간통계학을 이용한 소지역 추정법에 관하여 연구하였다. 이 논문에서 각 시군구 사이에는 공간 상관관계가 존재하는 것을 보였으며, 공간 상관관계를 분석에 이용할 경우 더 좋은 결과를 얻을 수 있음도 보였다. 또한 김재두 등 (2005)은 공간 통계뿐 아니라 시계열 분석 기법을 동시에 소지역 추정에 응용하였다. 이러한 모형기반 분석과는 달리 자료기반 소지역 추정법은 직접 추정법, 합성 추정법 그리고 복합 추정법이 있다. 이 중에서 합성 추정법과 복합 추정법은 보조 변수가 있는 경우에 사용할 수 있는 방법으로, 간단히 구할 수 있다는 장점은 있으나 모형기반 소지역 추정법이 더 우

1) This research was supported by the research fund of Hankuk University of Foreign Studies, 2006.

2) Graduate student, Department of Statistics, Hankuk University of Foreign Studies, Mohyun, Yongin, Kyonggi, 449-791, Korea.

3) Professor, Department of Statistics, Hankuk University of Foreign Studies, Mohyun, Yongin, Kyonggi, 449-791.

수한 결과를 주는 것으로 알려져 있어 그 사용이 제한적이다. 이상과 같이 보조 변수가 있는 경우 합성 추정법과 복합 추정법 그리고 모형기반 추정법이 사용될 수 있다.

그러나 얻어진 보조 변수가 전무한 경우에는 직접 추정법을 사용하는 것이 일반적인 방법이다. 직접 추정량은 불편 추정량이라는 하나 분산이 매우 큰 것으로 알려져 있어 사용에 어려움이 있다. 이러한 이유로 직접 추정량보다 더 우수한 추정량에 관한 연구가 필요한 실정이다. 이에 본 논문에서는 보조 변수가 존재하지 않는 경우에 사용할 수 있는 공간 상관관계를 이용한 소지역 추정량을 제안하였으며 또한 직접 추정량과 공간 상관관계를 이용한 추정량의 선형결합인 가중평균 추정량을 제안하였다. 본 논문에서 사용한 자료는 2001년 4월 경제활동인구조사에서 전국(제주도 제외) 29,899가구에서 얻어진 각 가구의 실업자수를 이용하였다. 이 자료는 신기일과 이상은(2003)의 논문에서 사용된 자료와 같은 자료이며 시군구 별로 공간 상관관계가 존재하는 것으로 알려져 있다. 본 논문에서도 Freeman-Tukey (1950)변환이 사용되었으며 자료의 히스토그램과 변환된 후 히스토그램은 신기일과 이상은(2003)을 참조하기 바란다.

2절에서는 직접 추정량과 공간 상관관계를 이용한 공간 추정량, 그리고 이들의 선형 결합인 가중평균 추정량을 살펴보았다. 3절에는 표본 수에 따른 각 통계량의 우수성을 모의실험을 통하여 비교하였으며 4절에 결론이 있다.

2. 사용된 추정법

서론에서도 언급하였듯이 모형 설정에 앞서 먼저 자료를 조사구 수, n_i 와 실업자 수, X_i 를 다음의 Freeman-Tukey (1950) 변환을 이용하여 변환하였다.

$$Y_i = (1000 \times X_i / n_i)^{1/2} + (1000 \times (X_i + 1) / n_i)^{1/2}$$

변환된 시군구별 자료가 분석에 사용되었다. 이 절에서는 분석에 사용된 직접 추정량, 공간 상관관계를 이용한 공간 추정량과 이들의 가중 평균인 가중 평균 추정량을 설명하였다.

2.1 직접 추정법 (direct estimation)

일반적으로 직접 추정량은 보조 정보가 없는 경우 사용되는 방법이며 본 논문에서도 그 지역에서 조사된 자료만을 이용하는 것을 고려하였다. 먼저 시도 기호를 a , 시군구 기호를 h 라 하자. 그러면 직접 추정량 Y_{ah}^{DE} 는 다음과 같다.

$$Y_{ah}^{DE} = \sum_{i=1}^{n_{ah}} w_{ahi} Y_{ahi}$$

여기서 n_{ah} 는 a 시도, h 시군구의 표본 수이며 w_{ahi} 는 가중치이다.

2.2 공간 추정량

본 논문에서 사용된 자료는 2001년 4월 경제활동인구조사에서의 실업자수 자료이며 신기일과 이상은 (2003)의 결과에 의하면 시군구 간에 공간 상관관계가 존재하는 것으로 알려졌다. 시군구 간의 공간 상관관계를 이용하기 위한 SAR 모형(Simultaneous Autoregressive Model)은 다음과 같다.

$$Z_{ah} = \rho_a S_{ah} + \epsilon_{ah}$$

여기서 $Z_{ah} = Y_{ah} - \bar{Y}_a$ 이고 $S_{ah} = \sum_{k \in N_a(h)} Z_{ak}$ 이다. 즉 S_{ah} 는 a 시도, h 시군구의 이웃 집합을 $N_a(h)$ 라 했을 때 이웃에 포함된 Z_{ak} 를 모두 더한 값이다. 이웃을 정하는 방법은 여러 가지가 있다. 본 논문에서는 여러 이웃을 정하는 방법을 실시하여 이 중에서 가장 좋은 결과를 주는 방법을 선택하여 사용하였다. 즉 본 논문에서 사용된 이웃은 서로 경계를 공유하는 시군구를 이웃으로 정하는 방법을 사용하였다. 이에 관한 내용은 Cressie (1993) 또는 김재두 등 (2005)을 참조하기 바란다. 따라서 추정량은 다음과 같이 정의된다.

$$\widehat{Y}_{ah}^{SP} = \bar{Y}_a + \hat{\rho}_a S_{ah} \tag{2.1}$$

본 논문에서는 \widehat{Y}_{ah}^{SP} 추정량을 간단히 공간 추정량이라 부르겠으며 보조 정보가 없을 때에도 사용할 수 있는 추정량임을 알 수 있다.

2.3 가중평균 추정량

보조 정보가 없는 경우에 사용하는 일반적인 방법이 직접 추정량이다. 직접 추정량은 불편 추정량이나 변동성이 크다는 단점이 있다. 또한 2.2 절에서 제안한 공간 추정량은 변동성이 작은 반면 편의가 큰 것으로 나타났다. 즉 공간 추정량은 자료기반 소지역 추정량의 하나인 합성 추정량과 같은 성질을 갖고 있는 것이다. 따라서 직접 추정량과 합성 추정량을 선형 결합하여 우수한 특성을 갖는 복합 추정량을 만든 것처럼 직접 추정량과 공간 추정량을 선형 결합함으로써 직접 추정량의 변동성과 공간 추정량의 편의를 동시에 줄일 수 있는 새로운 추정량을 만들 수 있을 것이다. 즉 제안된 새로운 추정량인 가중평균 추정량은 다음과 같이 정의 된다.

$$\widehat{Y}_{ah}^{NE} = \alpha_{ah} \widehat{Y}_{ah}^{DE} + (1 - \alpha_{ah}) \widehat{Y}_{ah}^{SP}$$

일반적으로 α_{ah} 는 \widehat{Y}_{ah}^{NE} 의 MSE 를 최소로 하는 값으로 정해지며 \widehat{Y}_{ah}^{NE} 의 MSE 를 최소로 하기 위하여 MSE 를 구하면 다음과 같다.

$$\begin{aligned} MSE(\widehat{Y}_{ah}^{NE}) &= \alpha_{ah}^2 MSE(\widehat{Y}_{ah}^{DE}) + (1 - \alpha_{ah})^2 MSE(\widehat{Y}_{ah}^{SP}) \\ &\quad + 2\alpha_{ah}(1 - \alpha_{ah})E(\widehat{Y}_{ah}^{DE} - Y_{ah})(\widehat{Y}_{ah}^{SP} - Y_{ah}) \end{aligned}$$

위의 식을 최소로 하는 α_{ah} 를 구하면 다음과 같다.

$$\alpha_{ah} = \frac{MSE(\widehat{Y}_{ah}^{SP}) - E(\widehat{Y}_{ah}^{DE} - Y_{ah})(\widehat{Y}_{ah}^{SP} - Y_{ah})}{MSE(\widehat{Y}_{ah}^{SP}) + MSE(\widehat{Y}_{ah}^{DE}) - 2E(\widehat{Y}_{ah}^{DE} - Y_{ah})(\widehat{Y}_{ah}^{SP} - Y_{ah})}$$

이제 $E(\widehat{Y}_{ah}^{DE} - Y_{ah})(\widehat{Y}_{ah}^{SP} - Y_{ah})$ 가 무시될 정도로 작다고 가정하면, 최적의 가중치

$$\alpha_{ah} = \frac{MSE(\widehat{Y}_{ah}^{SP})}{MSE(\widehat{Y}_{ah}^{SP}) + MSE(\widehat{Y}_{ah}^{DE})} \quad (2.2)$$

를 얻게 된다. 이에 관한 자세한 내용은 Schaible (1978) 또는 Rao (2003)를 참조하기 바란다. 이와 같이 두 개의 추정량을 결합하여 새로운 추정량을 만든 것을 직접 추정량과 합성 추정량을 결합하여 복합 추정량을 만든 것 뿐 아니라 합성 추정량과 사후 추정량을 결합하여 만든 복합 추정량도 Falorsi 등 (1994)에 의해 제안되었다. 또한 이때의 가중값, α_{ah} 의 계산도 (2.2)와 같은 식을 사용하고 있다.

3. 모의실험

이 절에서는 2절에서 언급한 직접 추정량, 공간 추정량 그리고 가중평균 추정량을 비교하였다. 김재두 등 (2005)에서는 2001년 4월 경제활동인구조사에서 얻어진 29,899 가구 중에서 시계열이 유지되는 자료만을 사용하여 분석하였는데 이때의 모집단의 수는 약 16,000 이었으며 이중에서 8,000개를 추출하여 분석에 사용하였다. 본 논문에서는 2001년 4월 경제활동인구조사에서 얻어진 29,899 가구를 모집단이라 가정하고 표본의 수에 따라 추정량의 우수성을 살펴 보기 위해 $n = 3,000, 5,000, 10,000, 15,000$ 그리고 20,000 가구를 계통 추출하였다. 따라서 2.1절의 직접추정량, \widehat{Y}_{ah}^{DE} 에서 가중치 $w_{ahi} = N/n = 29,899/n$ 이 된다. 일반적으로 표본설계에서는 단순히 계통추출법을 사용하지 않고 상황에 맞는 정교한 방법이 사용되고 있다. 그러나 본 논문은 공간 통계모형이 소지역 추정에 사용될 수 있다는 것을 보이기 위한 것이므로 본 논문에서는 계통추출법을 사용하였다. 김재두 등 (2005)에서와 같이 Freeman-Tukey 변환을 실시하였으며 모의실험을 간단히 하기 위하여 (2.1)식에서, a 시도에서 얻어진 시도 평균 \bar{Y}_a 와 $\hat{\rho}_a$ 대신에, 신기일과 이상은 (2003)에서처럼 전국을 하나의 소지역이라 생각하고 얻어진 전국 평균 \bar{Y} 와 $\rho_a = \rho$ 를 이용하였다. 이를 자세히 설명하면 다음과 같다. 먼저 구해진 210개 시군구의 직접 추정량, \widehat{Y}_{ah}^{DE} 의 전국 평균, \bar{Y} 를 구하고 $\widehat{Y}_{ah}^{DE} - \bar{Y} = Z_{ah}$ 를 구한다. 구해진 210개의 Z_{ah} 와 이 지역의 이웃으로 정해진 Z_{ah} , $k \neq h$ 들의 합 S_{ah} 를 구한 후 210개의 자료 쌍 (Z_{ah}, S_{ah}) 를 구한다. 이제 210개의 자료를 절편이 없는 회귀 모형에 적합시켜 ρ 를 구한다. ρ 가 구해지면 (2.1)식에 의해 공간 추정량이 구해진다.

다음으로 제안된 추정량의 비교를 위해 세 개의 비교 통계량이 사용되었다. 즉 h 를 시군구라 할 때 비교 통계량은 $MSE(h)$, $MAE(h)$ 그리고 $MB(h)$ 가 사용되었다. $MAE(h)$ 의 결과는 $MSE(h)$ 와 큰 차이를 보이고 있지 않아 지면 관계상 $MSE(h)$ 와 $MB(h)$ 결과만을 표로 만들었다. 사용된 통계량은 다음과 같다.

$$MSE(h) = \frac{1}{R} \sum_{r=1}^R (Y_h - \widehat{Y}_{hr})^2$$

$$MAE(h) = \frac{1}{R} \sum_{r=1}^R |Y_h - \widehat{Y}_{hr}|$$

$$MB(h) = \frac{1}{R} \sum_{r=1}^R (Y_h - \widehat{Y}_{hr}) = Y_h - \frac{1}{R} \sum_{r=1}^R \widehat{Y}_{hr}$$

여기서 Y_h 는 h 시군구의 실제값을, \widehat{Y}_{hr} 는 r 번째 추정값을 의미하며 이러한 과정을 $R=1,000$ 번 반복하였다.

위에서도 언급하였듯이 가중평균 추정량을 사용하기 위해서는 α_{ah} 를 구해야 한다. α_{ah} 는 (2.2)식을 사용하여 구하나 실제 자료 분석에서는 $MSE(\widehat{Y}_{ah}^{SP})$ 와 $MAE(\widehat{Y}_{ah}^{DE})$ 를 구하는 것이 간단하지 않다. 붓스트랩 등의 방법을 사용하거나 적당한 숫자를 골라 사용하기도 한다. 또한 시군구별로 다른 가중치를 사용하는 것이 아니라 시도별 또는 전국 단위로 하나의 가중치를 구해서 사용하는 것이 일반적인 방법이다. 본 논문에서는 3.1 절에 시도 및 시군구에 상관없이 $\alpha_{ah}=1/2$ 인 경우를 살펴보았다. 또한 본 논문처럼 모집단이 있는 경우에는 (2.2)식을 이용하여 α_{ah} 의 값을 추정할 수 있으므로 3.2절의 모의실험에서는 구해진 추정값을 이용하여 얻어진 결과를 분석하였다.

3.1 $\alpha_{ah}=1/2$ 인 경우의 모의실험

먼저 <표 1>과 <표 2>에 서울 지역에서 얻어진 $MSE(h)$ 와 $MB(h)$ 결과를 수록하였다. 다른 시도의 경우는 각각을 표로 만들지 않고 각 시도 별로 직접 추정량과 공간 추정량(Y^{DE} 와 Y^{SP})을 비교하여 우수한 시군구의 개수를 표로 작성하였다. 같은 방법으로 직접 추정량과 가중평균 추정량(Y^{DE} 와 Y^{NE}), 그리고 공간 추정량과 가중평균 추정량(Y^{SP} 와 Y^{NE})을 비교하였다. 이를 <표 3>부터 <표 5>까지 정리하였다.

<표 1>을 살펴보자. 먼저 $n=3,000$ 인 경우 직접 추정량은 제안된 두 추정량에 비해 매우 큰 MSE 값을 갖는 것을 볼 수 있다. <표 3>에서도 알 수 있듯이 직접 추정량, Y^{DE} 이 공간 추정량, Y^{SP} 보다 좋은 지역은, 중랑구, 구로구, 영등포구, 송파구 등 네 지역에 불과하고, Y^{DE} 가 Y^{NE} 보다 좋은 지역은 영등포구 한 지역뿐이다. 또한 세 통계량을 비교하였을 때에도 영등포구 지역만이 직접 추정량, Y^{DE} 이 가장 우수한 것으로 나타났다. 따라서 $n=3,000$ 인 경우 직접 추정량을 사용하는 것 보다 공간 추정량 또는 가중평균 추정량을 사용하는 것이 바람직하다. 그러나 자료의 수가 5,000개에서 20,000로 증가하는 경우 Y^{DE} 의 MSE 는 빠른 속도로 감소하는 것을 확인할 수 있다. 이에 반하여 Y^{SP} 의 MSE 는 대부분이 감소하고 있으나 일부 구는 오히려 증가하는 것을 볼 수 있다. 특히 Y^{DE} 가 우수한 지역인 중랑구, 구로구, 영등포구, 송파구 등 네 지역에서 Y^{SP} 의 MSE 는 감소하지 않는 것을 볼 수 있다.

<표 1> 서울지역에서 각 소지역 추정량의 MSE 비교

시 도	3000개 추출			5000개 추출			10000개 추출			15000개 추출			20000개 추출		
	Y^{DE}	Y^{SP}	Y^{NE}	Y^{DE}	Y^{SP}	Y^{NE}	Y^{DE}	Y^{SP}	Y^{NE}	Y^{DE}	Y^{SP}	Y^{NE}	Y^{DE}	Y^{SP}	Y^{NE}
종로구	19.4	1.6	6.7	27.9	3.1	6.8	5.4	9.1	2.9	1.5	15.9	3.9	1.3	14.1	3.8
중구	32.6	7.1	9.5	11.6	8.2	4.9	1.0	5.5	1.7	1.2	3.9	1.4	0.5	3.3	0.9
용산구	77.2	64.9	33.1	68.1	52.3	32.4	10.7	33.2	11.0	3.8	26.0	7.9	2.5	23.9	6.9
성동구	74.8	9.2	21.9	14.3	4.8	5.4	10.9	1.2	3.7	2.5	1.1	1.4	2.3	0.9	0.8
광진구	55.7	6.8	16.1	21.6	3.2	6.1	14.7	0.3	3.9	1.6	0.6	0.3	2.1	0.7	0.6
동대문구	83.8	35.5	34.9	18.4	36.1	12.1	8.5	24.6	9.2	7.4	11.8	4.8	4.6	11.5	3.8
중랑구	140.0	264.1	99.9	130.6	267.8	101.1	46.6	243.3	69.4	12.5	224.2	59.7	14.6	217.2	56.8
성북구	167.5	35.9	62.0	69.7	31.4	23.7	50.0	18.0	12.6	2.5	8.9	2.5	3.6	7.8	2.2
강북구	60.6	16.5	21.5	21.6	12.6	8.8	6.7	7.9	3.6	6.7	6.3	2.8	1.1	5.3	1.6
도봉구	114.2	2.8	29.4	39.4	2.1	10.0	9.9	1.5	3.3	9.0	1.2	1.8	3.4	0.9	1.3
노원구	193.5	6.2	58.3	43.1	3.5	10.3	12.7	2.0	4.1	6.7	4.1	2.1	4.7	6.4	3.2
은평구	51.4	1.8	14.2	24.3	2.6	6.6	17.1	1.0	3.1	5.6	0.8	2.1	1.6	0.6	0.7
서대문구	72.0	9.0	22.7	78.6	4.9	23.5	22.2	0.7	5.8	20.9	1.1	7.3	4.5	0.7	1.1
마포구	123.6	104.0	60.8	77.7	89.0	41.0	13.1	59.3	19.1	6.1	52.8	13.3	2.8	47.3	12.8
양천구	132.5	78.4	53.3	95.1	88.6	47.2	38.0	88.9	32.3	12.6	88.3	26.7	3.2	86.7	22.9
강서구	59.7	3.6	14.4	17.4	4.4	5.9	11.3	3.0	2.9	3.9	2.6	1.9	5.3	1.6	2.1
구로구	85.7	134.0	56.1	75.2	158.3	59.6	30.1	143.5	42.8	6.3	134.8	36.0	2.8	130.4	34.0
금천구	68.3	40.7	28.3	16.4	37.6	13.4	9.0	27.0	9.2	5.0	23.1	6.8	1.9	21.6	5.9
영등포구	8.3	47.9	15.3	3.9	65.9	17.4	1.7	122.9	31.2	0.8	155.1	39.3	0.3	167.2	42.0
동작구	79.7	7.6	19.0	48.6	8.1	12.6	18.1	3.9	7.5	5.3	3.6	1.9	3.2	1.4	1.3
관악구	55.8	21.7	18.2	84.7	20.1	25.2	21.1	8.9	7.5	6.0	4.3	1.9	3.2	2.3	1.4
서초구	110.8	27.1	38.5	24.8	24.4	14.3	9.9	11.8	6.2	3.6	7.4	3.3	2.1	5.9	1.9
강남구	48.1	3.8	14.1	55.0	4.6	16.8	25.7	9.0	7.7	6.4	6.5	4.1	7.4	6.7	2.6
송파구	33.1	43.3	16.3	12.3	37.5	10.4	7.2	39.5	10.1	1.9	44.0	11.1	1.1	42.5	10.6
강동구	162.9	47.1	53.6	85.3	60.0	35.4	34.6	67.0	24.5	24.7	68.8	24.4	10.0	69.5	20.1

이러한 현상은 추가적인 정보를 이용하여 추정의 정확도를 높이는 추정 방법(회귀 추정법, 베이지안 추정법) 등에서 일반적으로 발생할 수 있다. 즉 추정법에 사용된 기본적인 가정과 자료에서 얻어진 특징이 서로 다를 때 나타나는 결과이다. 본 논문에서는 같은 경계선을 공유하는 경우 이웃이라 정했고 이웃의 자료와 기준이 되는 시군구의 자료 사이에 큰 차이가 존재하기 때문에 이러한 결과가 나타났다고 할 수 있다. 이에 관한 자세한 내용은 Ghosh (2006) 등을 참조하기 바란다. 따라서 보다 세심한 이웃 결정이 필요하다고 하겠다. 결국 자료의 수가 20,000인 경우 <표 5>에서도 알 수 있듯이 16개 지역에서 Y^{DE} 가 Y^{SP} 보다 그리고 Y^{DE} 가 Y^{NE} 보다 우수한 것을 알 수 있다. 또한 세 통계량을 동시에 비교해 보면 12개 지역에서 Y^{DE} 가 가장 우수한 결과를 주고 있음을 확인 할 수 있다. 이는 자료가 충분한 경우인 일반 표본조사에서 직접 추정량의 우수성을 보여주는 결과라 할 수 있다. 물론 Y^{NE} 의 경우 $\alpha_{ah} = 1/2$ 를 사용하지 않고 다른 가중치를 사용한다면 다른 결과를 얻게 될 것이다. 다음으로 편

의를 알아보기 위한 비교 통계량인 MB 결과를 확인 하도록 하자.

<표 2> 서울지역에서 각 소지역 추정량의 MB 비교

시 도	3000개 추출			5000개 추출			10000개 추출			15000개 추출			20000개 추출		
	Y^{DE}	Y^{SP}	Y^{NE}	Y^{DE}	Y^{SP}	Y^{NE}	Y^{DE}	Y^{SP}	Y^{NE}	Y^{DE}	Y^{SP}	Y^{NE}	Y^{DE}	Y^{SP}	Y^{NE}
종로구	-0.02	-0.14	-0.08	0.05	-1.10	-0.52	0.09	-2.80	-1.36	-0.04	-3.63	-1.83	0.05	-3.73	-1.84
중구	-0.06	2.64	1.29	-0.03	2.86	1.41	-0.01	2.33	1.16	0.00	1.96	0.98	0.02	1.79	0.90
용산구	-0.69	8.04	3.67	0.30	7.16	3.73	-0.15	5.62	2.74	0.04	4.97	2.51	0.11	4.87	2.49
성동구	0.02	2.91	1.47	0.08	1.80	0.94	-0.02	0.08	0.03	-0.01	-0.64	-0.32	-0.02	-0.78	-0.40
광진구	0.09	2.43	1.26	-0.10	1.65	0.78	-0.10	0.35	0.12	0.03	-0.50	-0.24	-0.09	-0.70	-0.39
동대문구	0.30	5.65	2.97	-0.29	5.66	2.69	0.08	4.27	2.17	-0.08	3.35	1.63	-0.07	3.15	1.54
중랑구	0.33	16.06	8.20	0.35	16.31	8.33	0.19	15.47	7.83	0.05	14.96	7.50	-0.13	14.73	7.30
성북구	0.08	5.44	2.76	-0.16	5.44	2.64	-0.19	3.96	1.88	0.04	2.76	1.40	-0.03	2.52	1.25
강북구	0.24	3.93	2.09	0.20	3.54	1.87	0.06	2.79	1.43	-0.07	2.49	1.21	0.01	2.30	1.15
도봉구	-0.27	0.98	0.36	-0.08	1.31	0.62	0.09	1.10	0.59	0.05	0.83	0.44	-0.06	0.84	0.39
노원구	-0.45	0.45	0.00	-0.05	0.85	0.40	0.09	-0.72	-0.32	0.04	-1.80	-0.88	0.08	-2.18	-1.05
은평구	-0.21	0.76	0.27	0.41	0.69	0.55	0.12	-0.07	0.03	-0.02	-0.43	-0.22	-0.02	-0.54	-0.28
서대문구	-0.20	2.65	1.22	0.20	1.96	1.08	-0.12	0.31	0.10	-0.05	-0.38	-0.21	0.01	-0.47	-0.23
마포구	0.45	10.02	5.24	-0.05	9.19	4.57	-0.08	7.61	3.77	0.13	6.85	3.49	0.04	6.81	3.42
양천구	-0.31	8.80	4.25	0.44	9.36	4.90	-0.02	9.41	4.70	0.01	9.35	4.68	0.07	9.30	4.68
강서구	-0.20	1.57	0.68	0.22	1.88	1.05	0.04	1.47	0.76	0.03	1.29	0.66	0.04	1.16	0.60
구로구	0.10	11.44	5.77	-0.01	12.52	6.25	0.08	11.87	5.98	0.00	11.46	5.73	0.06	11.39	5.72
금천구	0.29	6.36	3.32	-0.05	6.12	3.03	0.02	5.19	2.60	-0.09	4.79	2.35	0.00	4.64	2.32
영등포구	-0.07	-6.66	-3.36	0.12	-7.81	-3.85	-0.01	-10.91	-5.46	0.02	-12.38	-6.18	-0.01	-12.89	-6.45
동작구	-0.04	2.20	1.08	-0.11	2.45	1.17	0.03	1.48	0.75	-0.09	1.17	0.54	-0.03	1.02	0.50
관악구	-0.10	4.39	2.15	0.07	4.19	2.13	0.07	2.43	1.25	0.10	1.49	0.79	0.01	1.33	0.67
서초구	-0.03	5.02	2.49	0.20	4.81	2.50	-0.14	3.22	1.54	0.11	2.40	1.25	0.04	2.32	1.18
강남구	-0.14	0.04	-0.05	0.06	-0.16	-0.05	-0.09	-1.62	-0.86	-0.08	-2.24	-1.16	0.16	-2.40	-1.12
송파구	0.44	-6.39	-2.98	0.06	-5.92	-2.93	0.12	-6.15	-3.02	-0.05	-6.51	-3.28	-0.01	-6.47	-3.24
강동구	-0.01	6.78	3.39	-0.02	7.70	3.84	0.01	8.15	4.08	0.07	8.27	4.17	0.18	8.33	4.25

$n = 3,000$ 인 경우 노원구와 강남구 등 두 개 지역을 제외하고 MB를 기준으로 하였을 때 Y^{DE} 가 Y^{SP} , Y^{NE} 에 비해 우수한 결과를 주고 있음을 알 수 있다. 기본적으로 Y^{DE} 는 불편 추정량인 것으로 알려져 있는 반면, Y^{SP} 는 이웃 정보를 이용하여 분석하는 방법이고 만약 이웃 값이 추정하려는 시군구와 같은 특징을 갖고 있지 않으면 편의가 발생하게 되므로 이러한 결과는 어느 정도 예상된 것이라 할 수 있다. 결국 몇 개의 구에서는 자료의 수가 증가하여도 Y^{SP} 의 MB는 줄어들지 않는 특징을 보이고 있다.

<표 3> MSE를 기준으로 한 각 시도별 우수 통계량 개수

		3000개 추출		5000개 추출		10000개 추출		15000개 추출		20000개 추출	
서울	Y^{DE} vs Y^{SP}	4	21	7	18	14	11	15	10	16	9
	Y^{DE} vs Y^{NE}	1	24	1	24	9	16	11	14	12	13
부산	Y^{DE} vs Y^{SP}	4	12	5	11	9	7	9	7	13	3
	Y^{DE} vs Y^{NE}	2	14	3	13	5	11	5	11	13	3
대구	Y^{DE} vs Y^{SP}	1	7	2	6	5	3	6	2	8	0
	Y^{DE} vs Y^{NE}	0	8	0	8	2	6	5	3	6	2
인천	Y^{DE} vs Y^{SP}	1	8	4	5	6	3	7	2	8	1
	Y^{DE} vs Y^{NE}	1	8	1	8	2	7	4	5	6	3
광주	Y^{DE} vs Y^{SP}	0	5	1	4	3	2	3	2	4	1
	Y^{DE} vs Y^{NE}	0	5	1	4	1	4	2	3	4	1
대전	Y^{DE} vs Y^{SP}	0	5	0	5	1	4	2	3	4	1
	Y^{DE} vs Y^{NE}	0	5	0	5	0	5	2	3	2	3
울산	Y^{DE} vs Y^{SP}	1	4	1	4	4	1	3	2	4	1
	Y^{DE} vs Y^{NE}	0	5	1	4	2	3	2	3	3	2
경기	Y^{DE} vs Y^{SP}	3	23	3	23	10	16	16	10	16	10
	Y^{DE} vs Y^{NE}	1	25	1	25	4	22	7	19	12	14
강원	Y^{DE} vs Y^{SP}	5	12	6	11	7	10	10	7	11	6
	Y^{DE} vs Y^{NE}	4	13	5	12	6	11	8	9	8	9
충북	Y^{DE} vs Y^{SP}	0	10	1	9	3	7	6	4	6	4
	Y^{DE} vs Y^{NE}	0	10	0	10	1	9	2	8	4	6
충남	Y^{DE} vs Y^{SP}	2	13	3	12	7	8	10	5	11	4
	Y^{DE} vs Y^{NE}	2	13	2	13	3	12	5	10	9	6
전북	Y^{DE} vs Y^{SP}	5	8	5	8	6	7	9	4	11	2
	Y^{DE} vs Y^{NE}	4	9	4	9	5	8	6	7	7	6
전남	Y^{DE} vs Y^{SP}	8	9	8	9	9	8	11	6	12	5
	Y^{DE} vs Y^{NE}	7	10	7	10	7	9	9	8	10	7
경북	Y^{DE} vs Y^{SP}	12	9	15	6	18	3	19	2	20	1
	Y^{DE} vs Y^{NE}	11	10	13	8	15	6	17	4	18	3
경남	Y^{DE} vs Y^{SP}	6	13	7	12	10	9	15	4	18	1
	Y^{DE} vs Y^{NE}	6	13	6	13	6	13	8	11	11	8

<표 4> MB를 기준으로 한 각 시도별 우수 통계량 개수

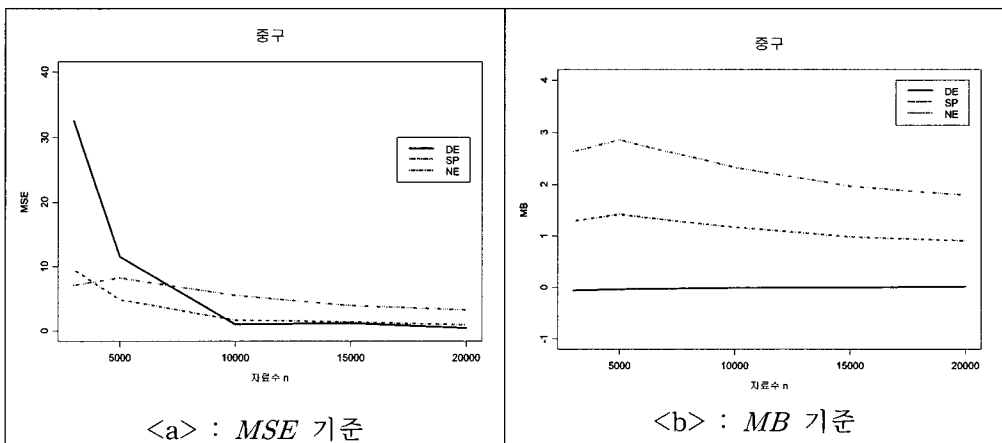
		3000개 추출		5000개 추출		10000개 추출		15000개 추출		20000개 추출	
서울	Y^{DE} vs Y^{SP}	23	2	25	0	24	1	25	0	25	0
	Y^{DE} vs Y^{NE}	23	2	24	1	23	2	25	0	25	0
부산	Y^{DE} vs Y^{SP}	15	1	16	0	16	0	16	0	15	1
	Y^{DE} vs Y^{NE}	14	2	15	1	16	0	15	1	15	1
대구	Y^{DE} vs Y^{SP}	8	0	8	0	8	0	8	0	8	0
	Y^{DE} vs Y^{NE}	8	0	8	0	8	0	8	0	8	0
인천	Y^{DE} vs Y^{SP}	9	0	9	0	9	0	9	0	9	0
	Y^{DE} vs Y^{NE}	9	0	9	0	9	0	9	0	9	0
광주	Y^{DE} vs Y^{SP}	5	0	5	0	5	0	5	0	5	0
	Y^{DE} vs Y^{NE}	5	0	5	0	5	0	4	1	5	0
대전	Y^{DE} vs Y^{SP}	5	0	5	0	4	1	5	0	5	0
	Y^{DE} vs Y^{NE}	4	1	5	0	4	1	4	1	5	0
울산	Y^{DE} vs Y^{SP}	5	0	5	0	5	0	5	0	5	0
	Y^{DE} vs Y^{NE}	4	1	5	0	5	0	5	0	5	0
경기	Y^{DE} vs Y^{SP}	24	2	25	1	25	1	26	0	26	0
	Y^{DE} vs Y^{NE}	23	3	24	2	25	1	26	0	26	0
강원	Y^{DE} vs Y^{SP}	17	0	17	0	16	1	17	0	17	0
	Y^{DE} vs Y^{NE}	17	0	17	0	16	1	17	0	17	0
충북	Y^{DE} vs Y^{SP}	9	1	10	0	10	0	10	0	10	0
	Y^{DE} vs Y^{NE}	9	1	9	1	10	0	10	0	10	0
충남	Y^{DE} vs Y^{SP}	15	0	15	0	15	0	14	1	15	0
	Y^{DE} vs Y^{NE}	15	0	15	0	15	0	13	2	15	0
전북	Y^{DE} vs Y^{SP}	13	0	13	0	13	0	13	0	13	0
	Y^{DE} vs Y^{NE}	12	1	12	1	13	0	13	0	13	0
전남	Y^{DE} vs Y^{SP}	17	0	16	1	17	0	16	1	17	0
	Y^{DE} vs Y^{NE}	16	1	16	1	17	0	16	1	16	1
경북	Y^{DE} vs Y^{SP}	21	0	21	0	21	0	21	0	21	0
	Y^{DE} vs Y^{NE}	20	1	20	1	21	0	20	1	21	0
경남	Y^{DE} vs Y^{SP}	18	1	19	0	19	0	19	0	19	0
	Y^{DE} vs Y^{NE}	16	3	19	0	19	0	19	0	18	1

<표 5> MSE와 MB를 기준으로 한 각 시도별 우수 통계량 개수

		3000개 추출		5000개 추출		10000개 추출		15000개 추출		20000개 추출	
		Y^{SP}	Y^{NE}	Y^{SP}	Y^{NE}	Y^{SP}	Y^{NE}	Y^{SP}	Y^{NE}	Y^{SP}	Y^{NE}
서울	MSE	15	10	11	14	7	18	4	21	4	21
	MB	1	24	0	25	0	25	0	25	0	25
부산	MSE	7	9	7	9	5	11	5	11	2	14
	MB	0	16	0	16	0	16	0	16	1	15
대구	MSE	5	3	3	5	1	7	0	8	0	8
	MB	0	8	0	8	0	8	0	8	0	8
인천	MSE	6	3	3	6	3	6	1	8	0	9
	MB	0	9	0	9	0	9	0	9	0	9
광주	MSE	4	1	3	2	1	4	0	5	0	5
	MB	0	5	0	5	0	5	0	5	0	5
대전	MSE	4	1	4	1	2	3	1	4	1	4
	MB	0	5	0	5	1	4	0	5	0	5
울산	MSE	4	1	2	3	1	4	1	4	1	4
	MB	0	5	0	5	0	5	0	5	0	5
경기	MSE	20	6	15	11	8	18	5	21	2	24
	MB	2	24	1	25	1	25	0	26	0	26
강원	MSE	9	8	9	8	6	11	6	11	3	14
	MB	0	17	0	17	0	17	0	17	0	17
충북	MSE	7	3	7	3	4	6	2	8	3	7
	MB	1	9	0	10	0	10	0	10	0	10
충남	MSE	8	7	7	8	3	12	1	14	2	13
	MB	0	15	0	15	0	15	0	15	0	15
전북	MSE	8	5	7	6	3	10	2	11	0	13
	MB	0	13	0	13	0	13	0	13	0	13
전남	MSE	7	10	8	9	5	12	3	14	2	15
	MB	0	17	0	17	0	17	1	16	0	17
경북	MSE	7	14	3	18	2	19	1	20	1	20
	MB	0	21	0	21	0	21	0	21	0	21
경남	MSE	12	7	9	10	3	16	2	17	1	18
	MB	1	18	0	19	0	19	0	19	0	19

<표 3>에서 <표 5>는 전국을 시도별로 나누는 후, 시군구 별로 각 통계량을 비교하여 우수한 지역의 수를 표로 만들었다. 먼저 <표 3>의 경우 직접 추정량 Y^{DE} 와 공

간 추정량 Y^{SP} 그리고 직접 추정량, Y^{DE} 와 가중평균 추정량, Y^{NE} 를 MSE 를 기준으로 비교하였다. $n=3,000$ 이고 서울의 경우를 보면 4개 구에서 Y^{DE} 가 Y^{SP} 보다 우수하고 1개 구에서 Y^{DE} 가 Y^{NE} 보다 우수하다. 같은 방법으로 부산에서는 4개 지역에서 Y^{DE} 가 Y^{SP} 보다 우수하고 2개 지역에서 Y^{DE} 가 Y^{NE} 보다 우수하다. 즉 대부분의 시군구에서 Y^{DE} 에 비해 Y^{SP} 와 Y^{NE} 가 우수한 것을 확인 할 수 있다, 그러나 자료의 수 $n=5,000$ 과 $10,000$ 이 되면서 Y^{DE} 가 우수한 시군구의 수가 증가하는 것을 확인 할 수 있으며 $n=20,000$ 인 경우에는 Y^{DE} 가 다른 두 추정량에 비해 우수한 것을 확인 할 수 있다. 다음으로 편의에 관해 정리한 <표 4>를 살펴보자. 편의의 경우 시도에 상관없이 그리고 자료 수에 상관없이 Y^{DE} 가 다른 두 추정량에 비해 우수한 것을 알 수 있다. 다음으로 <표 5>는 본 논문에서 제안한 두 추정량을 비교한 결과이다. 소지역 추정의 경우는 자료의 수가 작은 경우에 해당되므로 $n=3,000$ 또는 $n=5,000$ 인 경우를 살펴보자. MSE 를 기준으로 살펴보면 Y^{SP} 가 Y^{NE} 에 비해 약간 우수하거나 거의 비슷한 결과를 주고 있다. $n=3,000$ 인 경우에 Y^{SP} 가 약간 더 우수한 것을 확인 할 수 있으나 $n=5,000$ 인 경우 두 추정량은 거의 같은 결과를 주고 있다. 이에 반하여 MB 를 기준으로 하였을 경우 Y^{NE} 가 Y^{SP} 에 비해 전 시도에서 우수한 것을 알 수 있다. 물론 소지역 추정과 무관한 경우인 자료의 수가 큰 경우, $n=20,000$, Y^{NE} 는 Y^{SP} 에 비해 매우 우수한 결과를 주고 있다. 그러나 Y^{NE} 는 Y^{DE} 보다 나쁜 결과를 주고 있으므로 두 통계량의 비교는 큰 의미가 없다.



<그림 1> 자료의 크기에 따른 서울 중구의 소지역 추정량 비교

<그림 1>은 서울 중구의 소지역 추정에 관한 설명이다. <a>는 MSE 에 관한 설명이다. Y^{DE} 의 MSE 는 자료 수 n 이 증가 할 때 매우 빠른 속도로 감소하다가 $n=10,000$ 이상에서는 감소가 둔화 된 것을 볼 수 있다. 이에 반하여 Y^{SP} 는 $n=3,000$ 에서 가장 작은 MSE 를 갖고 있었으나 n 이 증가 할 때 그 감소가 매우 느

린 것을 보여 주고 있다. Y^{NE} 는 $n=5,000$ 일 때 세 추정량 중에서 가장 작은 MSE 값을 갖고 있으며 이후로 매우 안정적인 결과를 주고 있다. 를 보면 MB 에 관한 설명으로 Y^{DE} 의 편의는 “0” 근처 값을 계속 유지 하고 있는 반면 Y^{SP} 는 가장 큰 값을 갖고 있음을 알 수 있으며 자료가 커짐에도 불구하고 “0” 으로 수렴하지 않음을 알 수 있다. 3.1절의 모의실험에서는 Y^{NE} 가 두 추정량의 평균으로 계산되었기 때문에 Y^{NE} 의 편의는 두 추정량의 편이의 평균으로 계산된 것을 알 수 있다.

3.2 최적의 α_{ah} 를 이용한 서울 지역에서의 모의실험

3.1절에서 얻어진 $MSE(\hat{Y}_{ah}^{DE})$ 와 $MSE(\hat{Y}_{ah}^{SP})$ 를 이용하여 α_{ah} 를 구하고 이를 <표 6>에 작성하였다.

<표 6> 서울지역 α_h 와 최적 α_h 에 대한 Y^{NE} 의 MSE 와 MB

시 도	α_{ah}					3000 ($\alpha_n=0.326$)		5000 ($\alpha_n=0.469$)		10000 ($\alpha_n=0.681$)		15000 ($\alpha_n=0.845$)		20000 ($\alpha_n=0.907$)	
	3000	5000	10000	15000	20000	MSE	MB	MSE	MB	MSE	MB	MSE	MB	MSE	MB
종로구	0.076	0.100	0.628	0.914	0.916	4.0	-0.16	6.0	-0.62	3.0	-0.91	1.2	-0.60	1.2	-0.34
중구	0.179	0.414	0.846	0.765	0.868	6.4	1.78	4.7	1.54	1.0	0.73	1.0	0.28	0.4	0.16
용산구	0.457	0.434	0.756	0.872	0.905	35.9	5.27	31.6	3.77	8.0	1.68	3.6	0.77	2.5	0.50
성동구	0.110	0.251	0.099	0.306	0.281	12.9	1.99	4.8	0.90	5.4	0.07	2.0	-0.10	1.7	-0.05
광진구	0.109	0.129	0.020	0.273	0.250	9.5	1.68	5.8	0.96	7.2	0.05	1.0	-0.06	1.8	-0.06
동대문구	0.298	0.662	0.743	0.615	0.714	28.8	3.90	13.9	3.07	7.4	1.46	5.7	0.46	3.8	0.19
중랑구	0.654	0.672	0.839	0.947	0.937	131.1	10.77	101.6	8.57	43.4	4.92	14.4	2.31	13.7	1.29
성북구	0.176	0.311	0.265	0.781	0.684	43.4	3.65	25.5	3.01	22.2	1.31	1.7	0.41	2.7	0.25
강북구	0.214	0.368	0.541	0.485	0.828	15.7	2.78	8.0	1.78	4.0	0.87	4.8	0.43	1.0	0.22
도봉구	0.024	0.051	0.132	0.118	0.209	14.4	0.53	8.1	0.65	4.8	0.41	5.9	0.14	2.9	0.09
노원구	0.031	0.075	0.136	0.380	0.577	29.6	0.24	8.4	0.38	7.4	-0.20	4.3	-0.22	3.8	-0.10
은평구	0.034	0.097	0.055	0.125	0.273	7.2	0.38	5.4	0.28	7.1	-0.15	4.5	-0.01	1.3	-0.04
서대문구	0.111	0.059	0.031	0.050	0.135	13.4	1.69	21.6	1.01	10.7	0.20	14.8	-0.02	3.5	-0.02
마포구	0.457	0.534	0.819	0.896	0.944	60.6	6.57	42.4	4.86	13.7	2.47	4.6	1.03	3.0	0.65
양천구	0.372	0.482	0.701	0.875	0.964	51.6	5.90	46.9	5.07	25.6	2.78	12.7	1.51	3.4	0.84
강서구	0.057	0.202	0.210	0.400	0.232	6.6	1.01	5.0	0.95	5.3	0.46	3.3	0.22	4.3	0.07
구로구	0.610	0.678	0.827	0.955	0.979	70.1	7.72	59.6	6.46	26.8	3.61	8.0	1.77	3.4	0.99
금천구	0.373	0.696	0.750	0.822	0.919	26.7	4.38	14.5	3.22	7.2	1.70	4.1	0.69	1.7	0.36
영등포구	0.852	0.944	0.986	0.995	0.998	23.8	-4.50	20.9	-4.24	13.2	-3.48	4.4	-1.90	1.7	-1.19
동작구	0.087	0.143	0.177	0.404	0.304	9.6	1.50	11.6	1.19	10.3	0.52	3.6	0.13	2.8	0.11
관악구	0.280	0.192	0.297	0.417	0.418	14.2	2.88	23.6	2.10	10.1	0.89	4.0	0.25	2.6	0.15
서초구	0.197	0.496	0.544	0.673	0.738	28.0	3.26	14.5	2.64	6.6	1.01	3.0	0.39	1.8	0.15
강남구	0.073	0.077	0.259	0.504	0.475	7.5	0.04	14.6	-0.02	12.0	-0.67	4.9	-0.37	5.6	-0.30
송파구	0.567	0.753	0.846	0.959	0.975	20.9	-4.23	11.3	-3.13	6.4	-1.99	2.1	-1.01	1.2	-0.59
강동구	0.224	0.413	0.659	0.736	0.874	39.9	4.41	34.8	4.02	22.7	2.79	20.9	1.41	9.1	0.76

다음으로 서울의 각 구에서 얻어진 α_{ah} 의 평균을 구하였으며 이를 이용하여 자료의 수가 3,000에서 20,000일 경우의 Y_{ah}^{NE} 의 MSE 와 MB 를 구하였다. 이결과는 <표 6>에 나와 있다. 구별로 얻어진 α_{ah} 를 사용할 경우의 결과가, MSE 를 기준으로 했을 때, 서울의 평균을 사용했을 경우보다 우수하게 나오는 것은 당연할 것이다. 그러나 시도 또는 전국 값을 사용하는 것이 일반적인 방법이므로 이 결과를 표에 작성하였다. 이제 <표 6>의 결과와 <표 1>과 <표 2>를 비교하자. $n=3,000$ 의 결과를 살펴보면 MSE 의 경우 세 지역, 중랑구, 구로구 그리고 영등포구를 제외한 지역에서 $\alpha_{ah} = 1/2$ 을 사용한 결과보다 서울의 평균 $\alpha_{ah} = 0.326$ 를 사용한 것이 더 우수한 것으로 나타났다. 이 세 지역을 살펴보면 최적의 가중치가 $\alpha_{ah} \geq 0.5$ 이상이다. 따라서 0.5보다 작은 값을 사용하였기 때문에 공간 추정량의 영향력이 커져 나쁜 결과가 나오고 있다.

다음으로 서울이외의 다른 시도에서도 모의실험을 통하여 각 시도의 가중치 α_a 를 구하였고 이를 <표 7>에 작성하였다. <표 7>을 살펴보면 주어진 자료의 수에 따라 시도별로 차이가 있다. 그러나 일반적인 뚜렷한 경향을 보이고 있다. 즉 자료의 수 $n=3,000$ 인 경우 가장 큰 가중치가 강원도의 $\alpha_a = 0.387$, 그리고 가장 작은 곳은 전북으로 $\alpha_a = 0.086$ 이다. 그러나 $n=20,000$ 인 경우 가장 큰 경우가 전남으로 $\alpha_a = 0.968$ 그리고 가장 작은 곳이 전북으로 $\alpha_a = 0.564$ 이다. 따라서 자료의 수가 증가할수록 Y_{ah}^{DE} 의 영향력을 크게 하는 것이 바람직함을 보여주고 있다.

<표 7> 시도별, 자료수별 가중치 α_a

시 도	α_a				
	3000	5000	10000	15000	20000
서울	0.326	0.469	0.681	0.845	0.907
부산	0.386	0.500	0.667	0.852	0.944
대구	0.276	0.514	0.712	0.762	0.905
인천	0.212	0.355	0.509	0.709	0.880
광주	0.174	0.311	0.507	0.695	0.884
대전	0.146	0.111	0.230	0.528	0.746
울산	0.153	0.359	0.789	0.817	0.941
경기	0.207	0.287	0.482	0.640	0.716
강원	0.387	0.380	0.560	0.830	0.849
충북	0.249	0.470	0.643	0.841	0.876
충남	0.149	0.235	0.418	0.819	0.877
전북	0.086	0.084	0.160	0.295	0.564
전남	0.423	0.588	0.831	0.897	0.968
경북	0.265	0.577	0.816	0.898	0.962
경남	0.154	0.220	0.358	0.675	0.806

이제 서울에서 가중치 α_a 에 따른 MSE 와 MB 의 변화를 살펴보자. 이 결과는 <표

8>에 정리되어 있다. 먼저 자료의 수 $n=3,000$ 인 경우를 보면 가중치 α_u 에 상관없이 Y^{DE} 보다 Y^{NE} 가 우수한 것을 알 수 있다. 이러한 결과는 $n=5,000$ 인 경우에도 해당된다. 따라서 자료수가 $n=10,000$ 인 경우에는 가중치에 따라 상황이 달라지는 것을 알 수 있다. 따라서 Y^{NE} 를 소지역 추정량으로 사용할 때, 자료수가 $n=10,000$ 인 경우에는 가중치 α_u 를 잘 선택하여야 한다. 자료수가 $n \geq 10,000$ 인 경우에도 가중치의 선택이 중요하나 자료수가 많은 경우이므로 소지역 추정법과 다른 기법이 적용되어야 할 것으로 보인다. Y^{SP} 와 Y^{NE} 의 비교는 MB 를 기준으로 했을 때 Y^{NE} 가 우수한 것을 알 수 있으나 MSE 를 기준으로 했을 경우에는 α_u 에 따라 우열이 달라지고 있다. 그러나 작은 α_u 를 사용하게 된다면 Y^{DE} 의 영향이 줄어들어 일반적으로 Y^{NE} 가 Y^{SP} 보다 우수한 결과를 주고 있음을 알 수 있다.

<표 8> 서울에서 α_u 를 변화 시켰을 때의 우수 통계량 개수

		3000		5000		10000		15000		20000		3000		5000		10000		15000		20000	
		Y^{DE}	Y^{NE}	Y^{DE}	Y^{NE}	Y^{DE}	Y^{NE}	Y^{DE}	Y^{NE}	Y^{DE}	Y^{NE}	Y^{SP}	Y^{NE}	Y^{SP}	Y^{NE}	Y^{SP}	Y^{NE}	Y^{SP}	Y^{NE}	Y^{SP}	Y^{NE}
M B	1/4	25	0	25	0	25	0	25	0	25	0	0	25	0	25	0	25	0	25	0	25
	1/3	23	2	25	0	25	0	25	0	25	0	2	23	0	25	0	25	0	25	0	25
	1/2	23	2	24	1	23	2	25	0	25	0	1	24	0	25	0	25	0	25	0	25
	2/3	23	2	23	2	24	1	25	0	25	0	2	23	0	25	1	24	0	25	0	25
	3/4	21	4	23	2	23	2	25	0	25	0	2	23	0	25	0	25	0	25	0	25
M S E	1/4	2	23	6	19	11	14	14	11	15	10	10	15	5	20	6	19	3	22	1	24
	1/3	1	24	5	20	10	15	13	12	14	11	12	13	8	17	7	18	3	22	1	24
	1/2	1	24	1	24	9	16	11	14	12	13	15	10	11	14	7	18	4	21	4	21
	2/3	1	24	1	24	3	22	8	17	12	13	17	8	12	13	11	14	4	21	7	18
	3/4	1	24	1	24	1	24	5	20	10	15	18	7	13	12	11	14	6	19	7	18

4. 결론

본 논문에서는 보조 정보가 없는 경우에 사용되는 직접 추정량의 문제점을 극복할 수 있는 새로운 추정량들을 제안하였다. 제안된 추정량은 MSE 를 기준으로 직접 추정량과 비교되었으며 비교 결과 제안된 추정량의 결과가 매우 우수한 것으로 확인되었다. 다만 MB 를 기준으로 했을 경우 직접 추정량이 우수한 것으로 나타났다. 그러나 소지역 추정의 경우 특정 시군구에 해당되는 자료의 수가 매우 작기 때문에 직접 추정량을 사용할 경우 매우 큰 MSE 를 갖게 되어 그 결과를 그대로 사용하는 것은 문제가 될 수 있다. 따라서 MSE 를 기준으로 추정량의 우수성을 판단하는 것은 타당하고 모의실험 결과 제안된 통계량의 우수성이 입증되었으므로, 보조 자료가 없을 때에는 직접 추정량을 사용하는 것보다 본 논문에서 제안한 두 통계량, 공간 추정량 Y^{SP}

와 가중평균 추정량 Y^{NE} 를 사용할 것을 제안한다.

참고문헌

- [1] 김재두, 신기일, 이상은 (2005). 공간 시계열 모형을 이용한 소지역 추정, 응용통계 연구, 18권, 627-637.
- [2] 신기일, 이상은 (2003). Model-Data Based Small Area Estimation. *The Korean Communications in Statistics*, Vol. 10, 637-645.
- [3] 이계오 (2002). 소지역 추정법에 관한 시군구 실업 통계 개발, 통계청 최종 보고서.
- [4] Cressie, N. (1993). *Statistics for Spatial Data*, John Wiley and Sons, Inc., New York.
- [5] Freeman, M.F. and Tukey, J.W. (1950). Transformation Related to the Angular and the Square Root. *Annals of Mathematical Statistics*, 607-611.
- [6] Falorsi, P.D. Falorsi, S. and Dusso A. (1994). Empirical Comparison of Small Area Estimation Methods for the Italian Labour Force Survey. *Survey Methodology*, Vol. 20, 171-178.
- [7] Ghosh, M. Maiti, T. and Roy, A. (2006). Influence Functions and Robust Bayes and Empirical Bayes Small Area Estimation. Malay Ghosh 초청세미나. 조사통계연구회.
- [8] Rao, J.N.K. (2003). *Small Area Estimation*, John Wiley and Sons, Inc, New York.
- [9] Schaible, W.L. (1978). Choosing Weights for Composite Estimators for Small Area Statistics, Proceedings of the Section on Survey Research Methods. *American Statistical Association*, 741-746.

[Received May 2006, Accepted July 2006]