

## 소지역모형 추정기법을 활용한 전·월세 추정

이승수<sup>1</sup> · 박원란<sup>2</sup> · 정성석<sup>3</sup>

<sup>1</sup>통계청 고용통계과 · <sup>2</sup>통계개발원 연구기획실 · <sup>3</sup>전북대학교 통계학과

접수 2017년 2월 20일, 수정 2017년 3월 15일, 게재확정 2017년 3월 16일

### 요약

조사를 통한 자료생성은 시간과 경제적인 제약이 많고, 조사 방법 및 특성에 따라 자료의 질이 결정되며, 수집된 조사정보를 통계정보로 활용하기까지 오랜 시간이 소요된다. 이와 같은 어려움을 줄이고자 조사 표본설계 단위 보다 작은 지역 또는 다른 영역에 대한 자료를 기존에 조사된 자료 및 행정 자료를 이용하여 추정하는 소지역추정 통계방법 활용 연구는 꾸준히 진행되고 있다. 따라서 본 논문에서는 소지역추정기법을 이용하여 인간이 기본적인 삶을 영위하는데 반드시 필요한 필수재이며, 동시에 우리나라에서 투자재로서의 특징을 나타내는 주택과 관련하여, 요즘 새로운 주거형태를 차지하는 전세와 월세 지수에 모형기반 소지역추정기법을 적용하고자 한다. 적용된 소지역추정 모형은 회귀모형 추정법, 계층적 베이저안 추정법, 시-공간적 추정법이며, 분석결과 전세와 월세에서 시-공간적 추정모형이 가장 효율적인 것으로 나타났다.

주요용어: 모형기반 추정법, 소지역추정, 전세 및 월세

### 1. 서론

오늘날 사회는 과거에 비해 다양한 분야에서 수많은 자료에 대한 수요가 발생되며, 이러한 수요는 자료의 신규 생성 및 기(既) 생성자료의 정리·가공을 통해 충족되어지고 있다.

자료의 신규 생성기법인 조사 (survey)를 통한 자료생성은 수집을 계획하고 실행하는데 시간과 경제적인 제약이 많고, 조사 특성에 따라 설문지의 질문 내용과 양이 영향을 받아 조사자들의 무응답이 생성되고, 조사자의 역량에 따라 자료의 질이 결정된다. 더불어, 조사 공표단위가 소지역인 경우 대규모 표본을 대상으로 조사를 실시해야 하고, 수집된 조사정보가 통계 정보화되기까지 오랜 시간이 소요되는 한계가 있다. 즉, 조사통계는 생산과정의 예산, 조사가구의 비협조와 조사원에 따른 조사오차, 대규모 조사 후 자료를 집계해서 공표하기까지 시간, 정보력 손실 등의 문제들이 곳곳에서 발생되게 된다.

이와 같은 조사를 통한 어려움을 줄이고자 자료의 정리·가공을 위한 다양한 통계적 활용방법 연구가 진행되고 있다. 특히, 조사 표본설계 단위 보다 작은 지역 (area) 또는 다른 영역 (domain)에 대한 자료를 기존에 조사된 자료 및 행정자료를 이용하여 추정하는 소지역추정 (small area estimation)법에 관심은 증가되고 있다. 소지역추정 연구는 미국, 캐나다, 영국, 일본 등에서 정부기관과 전문학자들이 공동 연구를 통해 이론적 결과를 축적하고 정부 차원의 인구, 소득, 노동력, 농업, 보건통계 등 다양한 통계조사 분야에서 활용되고 있다. 우리나라에서는 Lee (2002), Kim 등 (2008), Kwon과 Kim (2010), Kim과

<sup>1</sup> (35208) 대전광역시 서구 청사로 189 (둔산동, 정부대전청사 3동), 통계청 고용통계과, 주무관.

<sup>2</sup> (35220) 대전광역시 서구 한밭대로 713 (월평동 282-1), 통계개발원 연구기획실, 서기관.

<sup>3</sup> 교신저자: (54896) 전북 전주시 덕진구 백제대로 567, 전북대학교 통계학과 (응용통계연구소), 교수.

E-mail: sschung@jbnu.ac.kr

Kim (2010), Kim (2011) 등이 실업자 수 추정에 관련된 고용 분야, Kwon (2011), Lee와 Kim (2015)의 가계소득 분야, Park (2015), An 등 (2015)은 보건 분야, Woo 등 (2014)은 주택매매가격 지수 주택분야 등에서 연구를 실시하였다.

그러나 우리나라에서 진행되는 소지역추정기법은 대다수가 시군별 실업률과 관련된 고용통계에 관한 연구로 국한되어, 외국과 같은 다양한 분야에 연구가 진행되지 못하는 게 현실이다. 이에 본 연구는 인간이 기본적인 삶을 영위하는데 반드시 필요한 필수재이며, 동시에 우리나라에서 투자재로서의 특징을 나타내는 주택과 관련하여 소지역추정을 적용하고자 한다. ‘90년대 일본 경제의 장기침체나 미국의 서브프라임 모기지 사태 등에서 볼 수 있듯이 주택시장 변동은 경제 전반에 매우 큰 영향을 줄 수 있는 지표로써 본 연구는 2014년 주거실태조사에서 주택 거주유형의 21.7%을 차지하는 전세와 22.1%을 차지하는 월세에 모형기반 소지역추정을 적용하고자 한다. 이에 대한 연구는 Park (2003)이 소지역추정기법 중 설계기반 (design-based) 추정법과 간접 (indirect) 추정법을 적용한 연구가 진행된 바 있다.

소지역 추정을 위한 기법은 설계기반 (design-based) 추정법, 간접 (indirect) 추정법, 모형기반 (model-based) 추정법 등으로 나눌 수 있으며, 일반적으로 모형기반 추정량이 설계기반 추정량에 비해 더 효율적이고, 대다수의 국외 국가에서도 모형기반 추정방법을 적용하여 소지역분석을 실시하고 있다. 모형기반 추정량은 회귀모형 추정법 (Fay-Herriot method; FH), 계층적 베이지안 추정법 (Hierarchical Bayesian method; HB), 시간과 공간의 특성을 반영한 시-공간적 추정법 (spatio-temporal method) 등이 일반적으로 이용된다.

본 논문은 위에서 제시한 회귀모형 추정법, 계층적 베이지안 추정법, 시-공간적 추정법에 전세와 월세 지수를 적용하여 전·월세 지수에 효율적 모형을 평가하고자 한다. 전세와 월세 지수는 한국감정원의 주택가격동향조사에서 아파트, 단독주택, 연립주택을 종합한 지수를 대상으로 분석했으며, 전·월세에 영향을 주는 설명변수는 전·월세와 주택가격과의 상관성을 고려하여, 주택가격에 영향을 주는 지역별 세대, 주택매매지수, 미분양 주택수를 사용하였다. 추정량 정확도는 참값에 대한 평균상대편향 (average relative bias; ARB), 평균 제곱상대편향 (average squared relative bias; ASRB), 평균절대편향 (average absolute bias; AAB), 평균제곱편차 (average squared deviation; ASD)를 이용하였다.

본 논문의 2절에서는 소지역추정법 모형의 이론적 내용을 살펴보고, 3절에서는 분석에 이용된 자료와 추정기법과 모형의 적합정도를 확인할 비교 척도 및 추정결과와 적합정도를 살펴보고, 마지막으로 4절에서 결론과 제언으로 구성하였다.

## 2. 소지역 추정기법

### 2.1. 소지역 추정법 (small area estimation)

현대 사회는 과거에 비해 다양한 분야에서 수많은 자료의 수요가 발생되며, 특히, 21세기에 들어오면 서 사회, 정치, 경제 등 다양한 분야의 현황을 파악하고 평가하기 위한 자료의 수요는 꾸준히 증가하고 있다.

이러한 수요는 사회전반에 걸친 자료의 신규 생성 방법과 기(既) 생성자료의 정리·가공 방법을 통해 충족되어진다. 자료의 신규 생성 주요방법인 조사 (survey)를 통한 자료생성은 조사 공표단위가 소지역인 경우 대규모 표본을 대상으로 조사를 실시하고, 수집된 조사가 통계정보로 가공되기까지 시간이 많이 소요되는 한계가 있다. 이러한 조사를 통한 소지역 자료 생성방법의 문제점을 줄이고자 자료의 정리·가공을 통해 “정보”와 “지식”을 얻는 다양한 통계 활용기법 연구가 현재 진행되고 있다. 특히, 조사 표본 설계 단위 보다 작은 지역 (area) 또는 다른 영역 (domain)에 세부·특성화된 자료 생성을 위해 기존에 조사된 자료 및 행정자료를 활용하는 “소지역추정 (small area estimation)”에 대한 연구는 꾸준히 진행되고 있다.

소지역추정 (small area Estimation) 기법은 원하는 소지역 (small area)의 자료를 얻고자 기존의 자료 생성의 표본 확대없이 해당 지역에 대한 통계나 소분류 (small domain)에 대한 통계를 간접적으로 추정하는 방법이다. 다시 말해, 이 기법은 표본규모를 조사목적에 맞춰 생성하지 않고, 행정업무에 의해 생성된 행정자료 및 기 생성된 조사자료 등을 보조정보로 활용해 신뢰수준을 갖춘 소지역단위 통계를 생산하는 방법이다 (Ghosh와 Rao, 1994).

소지역추정기법은 설계기반 (design-based) 추정법, 간접 (indirect) 추정법과 모형기반 (model-based) 추정법 등으로 나눌 수 있다.

설계기반추정량은 직접추정량으로서 조사된 자료만을 이용하거나 이외 가능한 그 지역의 행정자료나 센서스 자료를 보조정보로 이용하는 추정량이나, 해당 소지역에 배정된 표본의 크기가 작은 경우에는 추정량의 분산이 커져서 신뢰성이 떨어진다. 간접추정량은 해당 지역의 조사 자료 및 지역을 포함한 더 큰 지역의 조사 자료를 이용하여 추정량을 구하는 방법으로 지역간 변동을 설명하지 못한다는 단점이 있다. 모형기반 추정량은 보조정보들에 의해 설명되지 않는 지역간 변동을 포함하는 모형을 이용하는 추정량으로 모형 내에 표본오차, 지역간 변동, 시계열 등을 모두 고려할 수 있다는 장점이 있고, 다양한 보조정보를 사용하여 추정량의 정도를 높일 수 있다 (Rao (2003)).

모형기반 추정량은 회귀모형 추정법 (Fay-Herriot method; FH), 계층적 베이지안 추정법 (Hierarchical Bayesian method; HB), 시간과 공간의 특성을 반영한 시-공간적 추정법 (spatio-temporal method) 등이 알려져 있다. 위에 언급된 소지역추정기법 중 모형기반 추정법을 살펴보면 아래와 같다

## 2.2. 모형기반 추정법 (model based method)

모형기반 추정법은 소지역 내 포함하는 공변량 (covariate)을 통해 모형을 생성하고 적합된 모형을 통해 소지역의 원하는 값을 추정하는 방법으로, 소지역을 연결하고 있는 모형구조가 소지역 간의 복잡한 오차구조를 내포하여 추정의 정확도를 높일 수 있고 표본자료를 이용하여 모형의 타당성 및 유용성을 확인할 수 있으며, 자료의 형태가 연속형과 범주형 자료 및 시계열 자료와 같은 다양한 형태에 적용이 가능하여 현재 여러 방면으로 연구가 진행되고 있다.

모형기반 추정법은 조사단위의 개별 값으로 모형화한 단위수준모형 (unit level model)과 지역단위의 값으로 모형화한 지역수준모형 (area level model)으로 나누어진다.

### 2.2.1. 단위수준 모형 (unit level model)

단위수준 모형은 조사단위 (unit) 별로 자료값 ( $y_{ij}$ )과 이와 관련된 보조자료 ( $x_{ij}$ )를 이용하여 모형을 설정하는 방법으로, 지분오차회귀모형 (nested error regression model)을 이용하며 구조화된 설정모형은 아래와 같다 (Battese 등, 1988).

$$y_{ij} = x_{ij}^T \beta + v_i + e_{ij}, \quad v_i \sim iidN(0, \sigma_v^2), \quad e_{ij} \sim iidN(0, \sigma_e^2)$$

여기서  $i$ 는 소지역,  $j$ 는 소지역내의 조사단위를 의미하며,  $v_i$ 는 지역효과,  $e_{ij}$ 는 개인별 오차를 나타내고  $v_i$ 와  $e_{ij}$ 는 상호간의 독립을 가정한다.

위의 모형은 모집단을 대표하는 모형으로 표본 (sample)은 모집단과 동일시되며 표본오차는 존재하지 않는다고 가정된다.

모형을 통한 최고선형비편향 추정량 (Best Linear Unbiased Prediction; BLUP)은 다음과 같다.

$$\tilde{Y}_i^{BLUP} = 1/N_i \left( \sum_{j \in s_i} Y_{ij} + \sum_{j \in r_i} \tilde{Y}_{ij} \right)$$

$\tilde{Y}_{ij} = x_{ij}^T \tilde{\beta} + \tilde{v}_i$ ,  $\tilde{v}_i = \gamma_i(\bar{y}_{i,j \in s_i} - \bar{x}_{i,j \in s_i}^T \tilde{\beta})$ ,  $\gamma_i = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2 / n_i)$ 로  $v_i$ 의 BLUP이며,  $s$ 는 표본집단,  $r$ 은 비표본집단을 의미한다.

단위 (unit) 수준모형은 설명변수가 설명력이 높을 때, 단위수준의 자료를 분석한 직접추정량을 이용하는 지역 (area) 수준모형보다 원자료에 대한 정보를 직접 활용되므로 모형의 효율이 더 높은 것으로 알려져 있다.

### 2.2.2. 지역수준 모형 (area level model)

지역수준 모형은 지역과 관련된 직접추정량과 보조자료를 결합한 모형으로, 단위수준자료 (unit level data)를 얻을 수 없을 때 사용하며, 가구소득, 고용통계 등 대다수의 소지역분석모형이 지역수준 모형에 의해 연구가 진행되고 있다.

$$\hat{y}_i = x_i^T \beta + v_i + e_i, \quad i = 1, 2, \dots, m$$

여기서,  $\hat{y}_i$ 는 직접추정량,  $x_i$ 는 보조자료,  $i$ 는 소지역을 의미한다.

지역수준 모형은 Fay-Harriot모형이라 불리는 지역수준의 회귀모형 추정법과 지역수준모형의 여러가지 확장모형 (expanded area-level model), 계층적 베이즈 (Hierarchical Bayes estimation) 모형 등의 다양한 방법이 알려져 있다.

#### 2.2.2.1. 지역수준 기본모형 (basic area level model, Fay-Harriot model)

지역수준 기본모형은 Fay-Harriot모형이라고 불리우며, 소지역에 대한 직접추정량( $\hat{y}_i$ )과 소지역의 보조변수로 표현되는  $y_i$ 의 두 가지 성분들을 모형으로 연결하여 모형기반 추정량을 찾아낸다. 직접추정량-모수간의 모형과 모수-보조변수간의 선형회귀모형의 2단계 결합 형태를 단계별로 살펴보면 다음과 같다.

$$1\text{단계}) \quad \hat{y}_i = y_i + e_i, \quad e_i \sim N(0, \sigma_i^2), \quad i = 1, 2, \dots, m$$

$\hat{y}_i$ 는 모수  $y_i$ 의 직접추정량인 동시에 불편추정량이며,  $e_i$ 는 분산값  $\sigma_i^2$ 을 가진다.

$$2\text{단계}) \quad y_i = x_i^T \beta + v_i, \quad v_i \sim N(0, \sigma_v^2), \quad i = 1, 2, \dots, m$$

$x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ 는 지역단위 보조변수,  $v_i$ 는 모형오차로 분산  $\sigma_v^2$ 을 가지며 표본오차  $e_i$ 와는 독립임을 가정한다. 최종적인 1단계에 2단계가 결합된 아래의 모형이 회귀모형 추정법 (Fay-Harriot) 모형이다 (Rao와 Yu, 1994; Rao, 2003).

$$\hat{y}_i = x_i^T \beta + v_i + e_i, \quad i = 1, 2, \dots, m$$

$e_i$ 의 분산  $\sigma_i^2$ 는 알려져 있다고 가정되지만, 실제로는 알려지지 않아 주어진 자료를 이용하여 추정된다. 지역랜덤효과  $v_i$ 의 분산  $\sigma_v^2$ 이 모든 지역에 동일하다는 가정 하에  $\hat{y}_i$ 의 BLUP (best linear unbiased prediction) 추정량은 다음과 같다.

$$\hat{y}_i^{BLUP} = \gamma_i \hat{y}_i + (1 - \gamma_i) x_i^T \hat{\beta}$$

여기서,  $\hat{\beta}$ 는 자료에 의해 얻어진  $\beta$ 의 최소제곱추정량이고,  $\gamma_i = \sigma_v^2 / (\sigma_v^2 + \sigma_i^2)$ 이다.

특히, Fay-Herriot (1979) 모형은 가장 잘 알려진 지역단위 소지역 추정 모형으로 설계 모형 (design model)과 회귀분석 모형 (regression model)의 조합으로 다른 지역의 특성을 이용 (borrow strength)한다. Fay-Herriot 모형은 직접 조사로 추정된 추정값의 공변량의 선형성을 가정하여, 이 가정을 만족하지 않을 때는 소지역 모수 추정값의 편향이 발생한다고 알려져 있다.

### 2.2.2.2. 계층적 베이저안 모형 (Hierarchical Bayesian model)

계층적 베이저 추정량 (Hierarchical Bayes estimator)은 계층적 모형에 베이저안 모형 및 깁스 샘플러 (Gibbs sampler) 개념을 적용하여 얻어진다. 계층적 베이저안 모형은 관측자료를 어떤 모수를 조건부로 모형화하고, 모형화된 모수는 다시 초모수 (hyperparameter)인 부가적인 모수에 의해 모형화되는데, 여기서 먼저 초모수가 주어졌을 때 사전분포를 생각하고, 다음으로 초모수에 대해 초사전분포를 고려하여 모형화가 이루어진다 (Kim, 2005).

소지역 추정을 위한 계층적 베이저안 모형은 다음과 같이 정의된다.

$$\begin{aligned}\hat{y}_i|y_i &\sim N(y_i, \sigma_i^2) \\ y_i|\beta, \sigma_v &\sim N(x_i^T \beta, \sigma_v^2) \\ (\beta, \sigma_v^2) &\sim p(\beta)p(\sigma_v^2).\end{aligned}$$

베이저 모형은 모수에 대한 사전분포가 필요하다. 일반적으로 사전정보가 없을 때는 무정보적 사전분포 (noninformative prior)를 활용하고, 정보가 있는 경우는 정보를 활용할 수 있는 정보적 사전분포 (informative prior)를 사용한다. 본 논문에서는 무정보적 사전분포를 사용한다. 구체적인 형태는 아래와 같다.

$$p(\beta) \propto 1, p(\sigma_v^{-2}) = Ga(a, b)$$

여기서  $Ga(a, b)$ 는 평균이  $a/b$ 인 감마분포 (gamma distribution)를 나타낸다. 분석에서는  $a = b = 0.001$ 를 사용하였다. 계층적 베이저 모형의 완전 조건부 확률분포 (full conditional distributions)는 아래와 같으며, 깁스표집법 (gibbs sampler)를 사용하여 사후표본 (posterior sample)을 추출한다.

$$\begin{aligned}y_i|\hat{y}_i, \beta, \sigma_v &\sim N\left(\left(\gamma_i \hat{y}_i + (1 - \gamma_i)x_i^T \beta\right), \sigma_i^2 \gamma_i\right), \gamma_i = \sigma_v^2 / (\sigma_i^2 + \sigma_v^2) \\ \beta|\sigma_v^2, y, \hat{y} &\sim N\left(\left(X^T X\right)^{-1} X^T y, \sigma_v^2 (X^T X)^{-1}\right) \\ \sigma_v^{-2}|\beta, y, \hat{y} &\sim Ga\left(a + m/2, b + \sum_{i=1}^m \left(y_i - x_i^T \beta\right)^2 / 2\right),\end{aligned}$$

여기서  $\hat{y} = (\hat{y}_1, \dots, \hat{y}_m)^T$ ,  $y = (y_1, \dots, y_m)^T$ 이고  $X = [x_1 \dots x_m]^T$ 이다.

각 소지역  $i = 1, \dots, m$ 에 대한 계층적 베이저 추정량을 구하기 위하여 본 논문에서는 라오-블랙웰 추정량 (Rao-Blackwellized estimator)을 사용한다 (You 등, 2003). 라오-블랙웰 추정량을 구하기 위하여  $\{(\hat{y}^{(lk)}, \gamma^{(lk)}, \beta^{(lk)}, \sigma_v^{2(lk)})\}_{l=1:L, k=1:K}$ 를 깁스 표집법에 추출한 사후표본이라 하자. 단,  $L$ 은 동시에 생성된 마르코프 체인 (Markov chain)의 수를 나타내고  $K$ 는 추출된 사후표본의 개수를 나타낸다. 그때, 라오-블랙웰 추정량은 다음과 같다.

$$E[y_i|\hat{y}_i] = \frac{1}{LD} \sum_{l=1}^L \sum_{k=d+1}^{2d} \left[ \gamma_i^{(lk)} \hat{y}_i^{(lk)} + \left(1 - \gamma_i^{(lk)}\right) x_i^T \beta \right]$$

여기서  $d$ 는 소각 (burn-in)까지의 사후표본의 개수를 나타내고,  $D$ 는 소각 후의 생성된 사후표본의 개수를 나타낸다. 즉, 총 추출된  $K = 2d$ 개의 사후표본 중 절반을 소각하였다.

### 2.2.2.3. 시-공간적 모형 (spatio-temporal model)

지역모형에서 지역별 시점자료가 존재할 때, 공간과 시점을 동시에 고려한 시-공간적 모형을 생각할 수 있다. 여러 시-공간 모형 중 Marhuenda 등 (2013)가 제시된 모형은 Fay-Harriot모형을 기반으로 하

는 모형으로 시간적 특성은 자기상관모형 (autoregressive model, AR)을 이용하고 공간적 특성은 동시 자기상관모형 (simultaneous autoregressive model, SAR)을 이용한다. 구체적인 모형을 살펴보면 아래와 같다.

$$\hat{y}_i = x_i^T \beta + u_i + v + e_i, \quad e_i \sim N(0, \sigma_i^2), \quad t = 1, \dots, T, \quad i = 1, 2, \dots, m \quad (2.1)$$

여기서  $i$ 는 소지역,  $t$ 는 시점이며,  $u_i$ 는 공간 (지역)효과는 SAR(1)를 가정 하에 식 2.2),  $v$ 는 공간 (지역)-시간의 AR(1) 랜덤효과 가정 하에 식 2.3)로 표현된다.

$$u_i = \rho_1 W u_{i \neq d} + \epsilon, \quad |\rho_1| < 1, \quad \epsilon \sim N(0_m, \sigma_1^2 I_m) \quad (2.2)$$

$$v = \rho_2 v_{i, t-1} + \epsilon_2, \quad |\rho_2| < 1, \quad \epsilon_2 \sim N(0_m, \sigma_2^2) \quad (2.3)$$

$W_{m \times m}$ 는 소지역 간의 정보를 포함한 대각선이 0인 가중값 행렬이다.

위의 식 2.1)에서  $Z_1 = I_m \otimes 1_T$ ,  $Z_2 = I_m \times T$ 인  $Z = (Z_1, Z_2)$ 와  $u = (u_i, v)$ 을 이용하여 모형화하면 아래와 같이 표현할 수 있다.

$$\hat{Y} = X\beta + Zu + e, \quad (2.4)$$

여기서  $\hat{Y} = (\hat{Y}_1, \dots, \hat{Y}_m)^T$ ,  $e = (e_1, \dots, e_m)^T$ 이고  $X = [x_1, \dots, x_m]^T$ 이다.

위 식을 통해  $Y$ 의 분산과  $\beta$ 의 BLUE (best linear unbiased estimator)는  $V(Y) = ZV_u(Y)Z^T + V_e$ ,  $\hat{\beta}(Y) = [X^T V^{-1}(Y)X]^{-1} X^T V^{-1}(Y)\hat{Y}$ 과 같다.

### 3. 자료 및 분석

#### 3.1. 활용 자료

인간의 기본적인 삶을 영위하는데 필요한 필수재이며, 동시에 우리나라에서 투자재로서의 특징을 나타내는 주택은 일본 경제의 장기침체나 미국의 서브프라임 모기지 사태 등에서 보여지듯이 한 나라 경제에 중요한 부분을 차지한다.

최근 들어 우리나라는 ‘전·월세난’과 ‘주거안정’이란 용어가 이슈화되고, 주택을 더 이상 소유개념이 아닌 거주개념으로 인식하며 전국적으로 주택시장 구조에 변화가 일어나고 있다. 이는 전세가격의 급등으로 전세가격이 매매가격보다 높은 아파트가 나타나고, 금리인하로 임대인이 전세보다 월세를 선호하여 전세물량 부족현상이 나타나는 동시에 월세가격은 하향 안정되고, 1~2인 가구 증가와 젊은층, 서민층 중심으로 도심형 소형주택수요 증가 등과 주택보급률 증가로 주택가격 급등에 대한 기대감이 사라지는데 기인한다.

그러나, 현재 진행되는 주택관련 논의는 주택 매매가격 변동 및 예측 등의 주택매매시장에 집중되어, 소유가 아닌 거주개념으로의 주택을 생각하는 현재의 주택시장 변화를 반영하지 못하고 있다. 따라서 전·월세 시장의 변화가 일시적이지 않고, 장기적인 현상으로 주택시장의 변화를 가져올 수 있다면 전·월세의 구조적 파악을 위한 모형화는 의미가 클 것이라고 생각된다. 이에 본 연구는 이러한 주택시장의 구조적 변화에 가장 중요한 부분을 차지하는 전·월세 연구를 시장의 변화가 가장 크고, 변화를 주도하는 서울 25개 구를 대상으로 진행하고자 한다. 전·월세 지수는 입지와 접근성에 따라 선호도를 유발하고, 그에 따라 가격이 형성됨을 고려할 때, 이와 깊은 연관성을 지닌 전·월세의 소지역 추세를 살펴보는 것은 의미가 있을 것으로 판단된다.

소지역추정을 위해 사용된 자료는 한국감정원이 매달 발표하는 주택가격동향조사의 전·월세 지수를 활용하였다. 주택가격동향조사는 1986년에 국민은행에서 최초 작성되어, 현재는 한국감정원에서 전국

의 주택 매매, 전세, 월세가격의 변동과 시장동향을 조사·분석하여 주택 정책수립 등에 활용하고자 조사된다. 주택관련 지수는 전국 264개 시군구의 거래 가능한 아파트, 단독주택, 연립주택 (임대 제외)의 월간조사와 아파트 표본 일부에 조사되는 주간조사로 나뉘어 진행되는 현장조사에 실거래가격이나 유사표본의 실거래가격이 반영되어 결정된다. 그러나 주택유형 중 아파트를 제외한 단독 및 연립주택은 공표지역이 광역지자체로 한정되어 시군구 이하의 세분화된 자료가 없는 실정이다. 이에 본 논문은 전·월세 시장의 흐름을 모형화된 소지역분석에 적용하고자 한다.

모형화에 활용된 자료는 주택가격동향의 아파트, 단독, 연립주택 유형을 포괄한 전·월세 종합지수로, 전세지수가 2003.11월부터, 월세지수가 2015.6월부터 월별 발표되었고 전·월세지수의 상대표준오차는 2015.11월부터 공식 제공되었다. 더불어, 주택매매 지수를 소지역분석에 적용한 Woo 등 (2014)의 연구에서 2014.1월부터 2014.9월까지 9개월 자료를 활용한 연구를 참고하여 분석시점을 2015.11월부터 2016.10월까지 12개월 자료를 이용하였다.

모형을 위한 보조변수는 전·월세의 주택가격과의 상관성을 고려하여 주택가격 결정에 영향력이 있다고 알려진 지역별 주민등록세대, 주택 매매수, 미분양주택수를 사용하였다.

### 3.2. 분석 결과

모형을 통한 소지역추정을 위해 지역수준 기본모형 Fay-Harriot 모형, 계층적 베イズ 모형 (Hierarchical Bayes model), 시점과 공간의 변화를 반영한 시-공간모형 (spatio-temporal model) 3가지 추정법을 이용하였으며, 모형의 구현은 R (Isabel Molina & Yolanda Marhuenda, 2015)과 Winbugs을 이용하였다.

위 모형 중 계층적 베イズ 모형은 회귀계수  $\beta$ 에 관해 사전정보가 없으므로 사전분포로 무정보적 사전분포 (flat prior)를 사용하였고,  $\sigma_v^2$ 은 공액사전분포인 역감마분포를 이용하였다 (Rao, 2003). 역감마분포의 모수  $a, b$ 는 일반적으로 작은 값이 이용되는데, 본 연구에서는  $a = b = 0.001$ 을 이용하였다 (You 등, 2003).

Fay-Herriot 및 시-공간 모형 (spatio-temporal model)  $e_i$ 의 분산추정량  $\hat{\sigma}_i^2$ 은 실제자료의 표본오차를 이용했으며, 시-공간 모형의  $W_{m \times m}$ 은 구간의 근접지역 정보를 이용하였다. 추정된 결과의 평가는 참값에 대한 평균상대편향 (average relative bias; ARB), 평균 제곱상대편향 (average squared relative bias; ASRB), 평균절대편향 (average absolute bias; AAB), 평균제곱편차 (average squared deviation; ASD)를 이용하였고, 4개의 척도 값은 작을수록 적합한 모형임을 의미한다.

$$ARB = \frac{1}{m} \sum_{i=1}^m \left| \hat{y}_i - \hat{y}_i^{BLUP} \right| / \hat{y}_i$$

$$ASRB = \frac{1}{m} \sum_{i=1}^m \left( \hat{y}_i - \hat{y}_i^{BLUP} \right)^2 / \hat{y}_i^2$$

$$AAB = \frac{1}{m} \sum_{i=1}^m \left| \hat{y}_i - \hat{y}_i^{BLUP} \right|$$

$$ASD = \frac{1}{m} \sum_{i=1}^m \left( \hat{y}_i - \hat{y}_i^{BLUP} \right)^2$$

여기서  $\hat{y}_i$ 는 소지역  $i$ 의 조사 추정치이며,  $\hat{y}_i^{BLUP}$ 는 모형 추정치를 나타낸다.

설명변수 지역별 주민등록세대, 주택매매수, 미분양주택수와 종속변수의 상관성을 살펴본 결과, 전세와 3개의 설명변수 상관성은 1%하에 유의하였다. 월세와 주민등록세대는 1%, 미분양주택수는 5%하에

유의하였으나, 주택매매 수는 상관성이 유의하지 않았다. 모형기준 소지역분석은 변수간의 선형이 전제되어 월세는 주택매매 수를 설명변수에서 제외하였다.

Table 3.1은 추정모형을 비교한 결과로, 전세와 월세의 적합도를 살펴보면, 모형4인 시-공간적 모형이 ARB, ASRB, AAB, ASD 4개의 적합측도에서 가장 작은 값을 보여 효율적인 것으로 나타났다. 이외 모형의 적합성은 Fay-Herriot모형, 계층적 베이즈 모형 (Hierarchical Bayes model) 순으로 나타났다.

**Table 3.1** Accuracy of estimation about charter and monthly rent

	Charter				Monthly rent			
	ARB	ASRB	AAB	ASD	ARB	ASRB	AAB	ASD
model1 (FH)*	0.00037	0.00000	0.03837	0.00446	0.00023	0.00000	0.02279	0.00178
model2 (FH)	0.00028	0.00000	0.02929	0.00259	0.00023	0.00000	0.02290	0.00180
model3 (HB)	0.00105	0.00000	0.10960	0.04310	0.00137	0.00000	0.13769	0.04102
model4 (STFH)	0.00023	0.00000	0.02448	0.00168	0.00017	0.00000	0.01685	0.00095

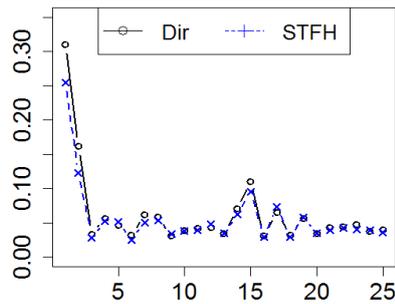
\* model1 (FH) includes time and area variable

Table 3.2는 모형 중 시-공간적 모형의 전·월세 모수 추정치이다.

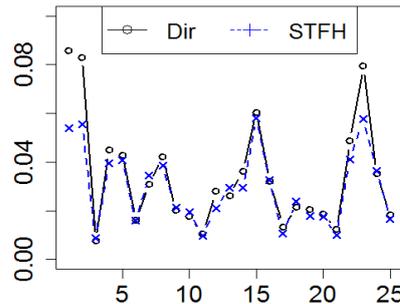
**Table 3.2** Estimation using spatio-temporal model

	charter		monthly rent	
	Coefficient	Standard error	coefficient	Standard error
Intercept	10280.95	1.8478462	99.603712	0.6341438
Household	0.0000077	0.0000106	0.0000035	0.0000036
House sale	-0.0000576	0.0001467	-	-
Unsold house	0.0015438	0.0010331	-0.0002017	0.0003951

Table 3.3은 시-공간적 모형의 2016년 10월의 주택가격동향의 조사값과 모형 추정치이며, Figure 3.1과 3.2는 전·월세의 조사값과 추정값의 CV (Coefficient of Variation)를 도시한 결과이다. Table 3.3의 결과를 살펴보면, 모든 구의 전·월세 조사값과 추정값은 차이는 크지 않았으나, CV는 몇 개 구를 제외한 대부분의 지역에서 전반적으로 작아져 효율성이 향상됨을 나타냈다.



**Figure 3.1** CV of charter



**Figure 3.2** CV of monthly rent

\* Dir: survey estimation, STFH: spatio-temporal estimation

\*\* Drawing 25 Gu in Seoul sorted by Hangeul alphabetical order

**Table 3.3** survey-estimation and CV (coefficient of variation)

	Charter				Monthly rent			
	Estimation		CV		Estimation		CV	
	Survey	Model*	Survey	Model*	Survey	Model*	Survey	Model*
Jongno	103.3	103.3	0.0473	0.0401	100.0	100.0	0.0795	0.0578
Jung	103.3	103.3	0.0376	0.0393	100.4	100.4	0.0354	0.0365
Yongsan	103.5	103.5	0.0432	0.0394	100.3	100.3	0.0122	0.0099
Seongdong	107.5	107.5	0.0312	0.0298	99.9	99.9	0.0320	0.0325
Gwangjin	105.1	105.1	0.0325	0.0255	99.8	99.8	0.0160	0.0160
Dongdaemun	106.3	106.3	0.0426	0.0399	100.4	100.4	0.0106	0.0097
Jungnang	104.4	104.4	0.0401	0.0363	99.6	99.6	0.0182	0.0167
Seongbuk	105.3	105.3	0.0647	0.0729	100.3	100.3	0.0134	0.0106
Gangbuk	104.0	104.0	0.0337	0.0280	100.4	100.4	0.0075	0.0086
Dobong	106.5	106.5	0.0385	0.0384	100.3	100.3	0.0176	0.0195
Nowon	108.0	108.0	0.0312	0.0343	99.8	99.8	0.0203	0.0213
Eunpyeong	104.3	104.3	0.0448	0.0424	99.8	99.8	0.0486	0.0410
Seodaemun	105.8	105.7	0.0702	0.0620	100.1	100.1	0.0361	0.0293
Mapo	106.0	106.0	0.0344	0.0352	100.1	100.1	0.0262	0.0295
Yangcheon	106.1	106.1	0.0564	0.0582	100.2	100.2	0.0205	0.0179
Gangseo	106.4	106.4	0.0569	0.0528	100.3	100.3	0.0451	0.0396
Guro	108.0	107.9	0.0622	0.0509	101.3	101.3	0.0309	0.0345
Geumcheon	104.1	104.1	0.0582	0.0532	100.2	100.2	0.0422	0.0386
Yeongdeungpo	106.2	106.2	0.0349	0.0355	100.0	100.0	0.0186	0.0174
Dongjak	105.4	105.4	0.0430	0.0486	100.7	100.7	0.0279	0.0210
Gwanak	105.8	105.8	0.0470	0.0514	101.0	101.0	0.0429	0.0408
Seocho	103.2	103.2	0.1105	0.0952	99.5	99.5	0.0604	0.0580
Gangnam	104.6	104.6	0.3102	0.2547	99.5	99.6	0.0857	0.0538
Songpa	103.5	103.5	0.0326	0.0290	101.2	101.2	0.0213	0.0237
Gangdong	102.7	102.8	0.1622	0.1229	98.7	98.9	0.0831	0.0554

\* Model: spation-temporal model

여기서, 주택가격동향의 조사값 전·월세 지수는 2015년 6월을 기준시점 (2015.6=100.0)으로 국제적으로 권고하는 지수 작성방법은 기하평균 방식 (제본스지수) 산식에 2016년 1월 기준 주택가격정보체계의 주택재고량을 가중치로 부여하여 산출된 결과이다.

#### 4. 결론 및 제언

서울지역의 구별 전·월세의 소지역추정에 적용하고자 Fay-Herriot 모형, 계층적 베이지 모형, 시-공간적 모형에 분석기법을 적용해 보았다. 분석 결과는 시간과 공간의 정보를 활용하여 모형을 설정하는 시-공간적 모형이 전·월세 추정에서 가장 효율적인 것으로 나타났다. 다음으로 적합도가 좋은 모형은 시간과 공간을 설명변수에 포함하지 않은 Fay-Herriot 모형, 시간과 공간변수를 설명변수로 사용한 Fay-Herriot 모형, 마지막으로 계층적 베이지 모형 (Hierarchical Bayes model)인 것으로 나타났다.

2016.10월 자료에 적합도가 가장 좋은 시-공간적 모형을 적합해 본 결과 조사자료의 상대표준오차 (CV) 보다 정확도가 향상됨을 알 수 있었다. 원 자료의 높은 정확도를 고려해 볼 때 이 향상 정도를 의미는 큰 것으로 생각되며, 향후 구보다 더 작은 단위의 자료를 활용하여 모형의 확대를 기대해 볼 수 있을 것이다.

본 연구는 새로운 소지역추정기법을 제안한 것은 아니다. 하지만, 시기적으로 우리나라에 새로운 주거형태를 반영하는 전·월세 실제 지수에 소지역 추정을 적용하므로 적합도가 향상된 모형을 찾아냈다는 사실은, 향후 주택시장의 읍면동단위까지의 자료 필요성을 고려할 때 의미가 있다고 생각되어진다. 그러나, 향후 본 연구에서 이용된 모형들 중 가장 적합도가 좋은 시-공간적 모형에 지역간 정보를 좀 더 효

과적으로 설정하는 방법, 계층화된 모형을 접합시키는 방법에 대한 연구는 더 진행될 필요가 있다고 판단되며, 제시된 모형들의 변수간 선형성 가정이 위배되었을 때 이 가정에 강건하다고 알려진 준모수적 모형 (semiparametric model)을 적용한 연구도 함께 진행되는 것도 의미 있을 것으로 판단된다.

## References

- An, D. S. and Han J. H and Yoon, T. H and Kim, C. H and Loh, M. S. (2015) Small area estimations for disease mapping by using spatial model. *Journal of the Korean Data & Information Science Society*, **26**, 101-109
- Battese, G. E., Harter, R. M. and Fuller, W. A. (1988). An error component model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, **83**, 28-36.
- Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269-277.
- Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science*, **9**, 55-76.
- Isabel Molina and Yolanda Marhuenda (2015). SAE : An R package for small area estimation. *The R Journal*, **7**, 55-93
- Kim, D. H. (2005) *Bayesian statistics using R and WinBUGS*, Freeacademy, Gyeonggi-do.
- Kim, J. and Kim, J. K. (2010). Small area estimation model for the urban unemployment statistics. *Journal of the Korean Statistical Society*, **17**, 37-347.
- Kim, S. T. and Ko S. N. and Kim S. D. (2008). An efficient approach to estimate labor statistics for small area - with a focus on unemployment rate. *Korean Journal of Industrial Relations*, **18**, 53-76.
- Kim, S. T. (2011). An efficient estimation of local area unemployment rate based on small area estimation. *The Korean Journal of Applied Statistics*, **24**, 1129-1138
- Kwon, S. P. and Kim, S. Y. (2010). *Estimate labor statistics for small area*. Statistics Korea, First half of 2010 Research Report, Daejeon.
- Kwon, S. P. and Kim, S. Y. and Lee, J. H. (2011). *A study on small area estimation method for regional statistics*, Statistics Korea, the second half of 2011 Research Report, Daejeon.
- Lee, G. O. (2002). *Development of municipality unemployment statistics by small area estimation method*, National Statistical Office Service Research Report, Statistics Korea, Daejeon.
- Lee, J. Y and Kim, D. H. (2015) Bayesian estimation of median household income for small areas with some longitudinal pattern. *Journal of the Korean Data & Information Science Society*, **26**, 755-762
- Marhuenda, Y., Molina, I. and Morales, D. (2013). Small area estimation with spatio-temporal Fay-Herriot models. *Computational Statistics and Data Analysis*, **58**, 308-325.
- Park, M. R. (2015). *Small area estimation based on mixed linear regression model using principal component score*, Master Thesis, Korea University, Seoul.
- Park, W. R. (2003). Estimation of rent change in small statistical unit. *Journal of The Korean Official Statistics*, **8**, 111-130.
- Rao, J. N. K. and Yu, M. (1994). Small-area estimation by combining time-series and cross-sectional data. *The Canadian Journal of Statistics*, **22**, 511-528.
- Rao, J. N. K. (2003). *Small area estimation*, Wiley, Hoboken, New Jersey.
- Woo, N. K., Kwon, B. J. and Park, B. J. (2014). Estimation of house price index using Bayesian inference method of small area model. *Korea Real Estate Analysts Association*, **2**, 1-6.
- 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*. *Statistics Canada*, **29**, 25-32.

## A case study of small area estimation about charter and monthly rent price index

Seung Soo Lee<sup>1</sup> · Won Ran Park<sup>2</sup> · Sung Suk Chung<sup>3</sup>

<sup>1</sup>Social Statistics Bureau, Statistics Korea,

<sup>2</sup>Statistical Research Institute, Statistics Korea

<sup>3</sup>Department of Statistics, Chonbuk University

Received 20 February 2017, revised 15 March 2017, accepted 16 March 2017

### Abstract

In this study we compared three models for small area estimation, Fay-Herriot, Hierarchical Bayes model and spatio-temporal model about charter, monthly rent price index. Charter, monthly rent price of Korea are important issue in these days. Because housing type rapidly changes from self to charter and monthly rent. The accuracy of the estimation was checked on four scales, that is ARB, ASRB, AAB, ASD. In this result, the spatio-temporal model among applied models has most optimal scales about small area estimation of charter and monthly rent index.

*Keywords:* Charter and monthly price index, Fay-Herriot model, Hierarchical Bayes model, small area estimation, spatio-temporal model.

---

<sup>1</sup> Official, Social Statistics Bureau, Statistics Korea, Government Complex-Daejeon, 189 Cheongsu-ro, Seo-gu, Daejeon 35208, Korea.

<sup>2</sup> Chief, Statistical Research Institute, Statistics Korea, Research Planning Division Statistical Center, Seo-gu, Daejeon 35220, Korea.

<sup>3</sup> Corresponding author: Professor, Department of Statistics, Chonbuk National University, Jeonbuk 54896, Korea. E-mail: sschung@jbnu.ac.kr