

공간회귀모형을 이용한 대구경북 지역 단위면적당 아파트 매매가격 예측

이우정¹ · 박철용²

¹²계명대학교 통계학과

접수 2015년 3월 6일, 수정 2015년 3월 18일, 게재확정 2015년 3월 23일

요약

이 연구에서는 공간회귀모형 중 공간시차모형과 공간오차모형을 이용하여 대구·경북 지역 단위면적당 아파트 매매가격을 예측하였다. k -최근접이웃 (k -nearest neighbours)을 이용하여 공간가중행렬을 구축하였으며, 이를 이용해 2012년 3월의 단위면적당 아파트 매매가격에 대한 모형을 적합시켰다. 적합시킨 공간시차모형, 공간오차모형을 이용하여 2013년 3월의 단위면적당 아파트 매매가격을 예측하였으며 RMSE (root mean squared error), RRMSE (root relative mean squared error), MAE (mean absolute error)를 통해 두 모형의 성능을 비교하였다.

주요용어: 공간시차모형, 공간오차모형, 공간회귀모형, 아파트 매매가격.

1. 서론

우리나라 주택시장에서 가장 큰 비중을 차지하는 것이 아파트시장이다. 베이비붐 세대의 성장으로 젊은 연령층 비중의 증가, 핵가족화의 진전 등은 우리나라의 주택수요가 아파트 중심으로 변화하는 중요한 요인이 되었다. 경제성장으로 인한 중산층의 성장과 이들의 생활양식의 확산, 주거요구의 증대 등도 중요한 사회경제적 요인이 되었다 (Chun, 2002). 이러한 모든 요인들이 상호작용하여 현재 우리나라 주택시장에서 아파트시장은 가장 큰 비중을 차지하게 되었다.

2000년 총 주택 수 11,472,401호 중에서 아파트가 5,479,828호로 약 47.7%의 점유율을 보였으며 2010년 총 주택 수 14,677,419호 중에서 아파트가 8,576,013호로 58.4%의 점유율을 보였다. 총 주택 수 증가 대비 아파트 수의 증가율은 96.6%로 10년간 늘어난 주택 수의 대부분이 아파트 수의 증가로 인한 것임을 알 수 있다. 이는 우리나라에서 아파트 수요가 절대적으로 증가하고 있다는 것을 보여준다.

주택가격은 인접 지역 주택의 가격과 상호 영향을 주고받으며 공간자기상관이 발생하게 되는데 집단 주택방식인 아파트가 대표적인 주택유형인 우리나라에서는 주택가격에 공간자기상관이 발생할 가능성이 높다. 이러한 인접 지역 주택들이 가지는 상관관계를 모형에 적극적으로 반영하기 위해 공간회귀모형을 적용할 수 있을 것이다.

이 때 공간회귀모형에서 공간적 자기상관에 대한 정보를 가장 많이 담고 있는 것이 공간가중행렬이다. 공간가중행렬은 크게 인접성 측도와 거리 측도로 나뉘는데, 본 연구에서는 거리 측도 중 k -최근접이웃 (k -nearest neighbours)을 이용하였다. 공간가중행렬에 대해서는 2절에서 자세한 설명을 할 것이다.

¹ (704-701) 대구광역시 달서구 달구벌대로 1095, 계명대학교 통계학과, 석사과정생.

² 교신저자: (704-701) 대구광역시 달서구 달구벌대로 1095, 계명대학교 통계학과, 교수.

E-mail: cypark1@kmu.ac.kr

Park 등 (2003)은 공간회귀모형을 이용하여 아파트가격과 공간효과를 분석함으로써 공간효과를 고려하였을 때 아파트 가격 추정의 정확성이 제고될 수 있음을 보여주었다. Chung 등 (2012)은 보통최소제곱추정량 (ordinary least squares estimator)을 이용한 기존의 헤도닉 모형보다는 공간자기상관을 고려한 공간회귀모형들의 설명력이 높았다는 점에 기반하여 부산시 실거래 주택매매 가격을 공간회귀모형에 적합비교한 후 공간시차모형의 설명력이 가장 좋다고 하였다.

이 연구에서는 국토교통부의 2012년 3월, 2013년 3월의 아파트 실거래가격을 전용 면적으로 나눈 단위면적당 (제곱미터당) 아파트 매매가격을 분석하였다. 공간회귀모형 중에서도 공간시차모형과 공간오차모형을 이용하여 대구·경북의 단위면적당 아파트 매매가격을 예측하고 두 모형의 성능을 비교하고자 하였다. 공간시차모형과 공간오차모형에 대해서는 2절에서 자세한 설명을 할 것이다.

이 논문은 다음과 같이 구성되어 있다. 2절에서 공간가중행렬, 공간시차모형과 공간오차모형의 특징을 설명한다. 3절에서는 공간시차모형, 공간오차모형을 이용하여 구축된 모형을 통해 예측을 해보고 RMSE (root mean squared error), RRMSE (root relative mean squared error), MAE (mean absolute error) 등의 지표를 이용하여 두 모형을 비교하고자 한다. 마지막 4절에서는 이 연구의 결론을 내리고 추후 연구 과제를 제시한다.

2. 공간회귀모형

2.1. 공간가중행렬

공간가중행렬 (spatial weight matrix)은 공간중속성을 구조화하는 방법으로서 공간회귀모형의 성능을 좌지우지한다고 할 만큼 중요한 비중을 차지하고 있다. 공간가중행렬을 구축하기 위해서는 인근지역이 설정되어야 한다. 인근지역은 인접성 측도 (contiguity measure) 또는 거리 측도 (distance measure)에 따라 구성할 수 있다. 연구자가 거리에 대한 정보를 알고 있는 경우 인접성 측도 보다는 거리 측도를 이용하는 것이 모형의 성능을 높이는데 도움이 된다고 알려져 있다 (Anselin, 1988). 따라서 본 연구에서는 거리측도인 k -최근접이웃 (k -nearest neighbours)을 이용하였으며 k 가 3일 때 인근지역을 설정하였다. k -최근접이웃은 가장 가까운 k 개의 지역을 인근지역으로 설정하는 방법이다. 인근지역 설정 결과를 공간가중행렬로 만들어주기 위해 행이 인근지역을 설정하고자 하는 지역, 열이 인근지역으로 설정된 지역을 의미하는 행렬에 대해 횡단표준화 (row standardization)를 시행하였다. 횡단표준화를 할 경우 주변지역의 값이 한 지점에 평균적으로 얼마나 영향을 미치는지를 계량화 할 수 있다는 장점이 있다.

2.2. 공간시차모형

공간시차모형 (spatial lag model)은 공간적 자기상관을 또 하나의 설명변수로 동원하는 개념으로, 식 (2.1)과 같은 형태를 가진다. 여기서 \mathbf{y} 는 단위면적당 아파트 매매가격이고, \mathbf{X} 는 아파트 매매가격에 영향을 주는 설명변수이며, ρ 는 공간자기회귀계수, \mathbf{W} 는 공간가중행렬을 나타낸다.

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \beta \mathbf{X} + \epsilon, \quad \epsilon \sim N(0, \sigma^2 \mathbf{I}) \quad (2.1)$$

$$\begin{aligned} \mathbf{y} - \rho \mathbf{W} \mathbf{y} &= \beta \mathbf{X} + \epsilon \\ (\mathbf{I} - \rho \mathbf{W}) \mathbf{y} &= \beta \mathbf{X} + \epsilon \\ \mathbf{y} &= (\mathbf{I} - \rho \mathbf{W})^{-1} \beta \mathbf{X} + (\mathbf{I} - \rho \mathbf{W})^{-1} \epsilon \end{aligned} \quad (2.2)$$

공간시차모형은 일반선형모형에서 공간자기회귀계수와 공간가중행렬, 반응변수의 곱인 $\rho W y$ 가 설명변수로 추가된 형태이다 (Hur, 2007). 식 (2.1)의 $\rho W y$ 를 이항시켜 정리하면 식 (2.2)를 얻을 수 있는데 여기서 $(I - \rho W)^{-1}$ 항은 공간승수효과 (spatial multiplier)를 나타내는 것으로 공간 간의 상호작용 효과를 나타내며, 각 지점에서 다른 모든 지점들이 서로 연관되어 있음을 의미한다 (Anselin, 2001). 따라서 공간시차모형에서의 정확한 회귀계수는 $(I - \rho W)^{-1}\beta$ 이고 아파트 매매가격이 주변 지역의 변화 뿐 아니라 공간가중행렬을 통해 그 외 지역의 변화에도 영향을 받는다는 것을 의미한다. 공간회귀모형의 모수를 추정하는 여러 가지 방법이 있으나 본 논문에서는 최대우도법 (method of maximum likelihood)을 이용하였다. 공간회귀모형에서 보통최소제곱추정법을 사용하면 추정량의 불편성, 일치성, 효율성이 보장되지 않기 때문에 일반적으로 사용되는 보통최소제곱추정법이 아닌 최대우도법을 이용한 것이다.

2.3. 공간오차모형

공간오차모형 (spatial error model)은 모형에서 고려하지 않은 변수들 사이에 공간적 자기상관이 존재한다는 것을 가정한다. 따라서 공간시차모형과 달리 공간적 자기상관을 부차적인 것으로 간주하여 이를 오차항에 포함시켜 처리하려는 입장으로 (Stakhoych와 Bijmolt, 2008), 모형식은 식 (2.3)과 같이 정의된다.

$$y = \beta X + u$$

$$u = \lambda W u + \epsilon, \epsilon \sim N(0, \sigma^2 I) \tag{2.3}$$

$$(I - \lambda W)u = \epsilon$$

$$u = (I - \lambda W)^{-1}\epsilon \tag{2.4}$$

$$y = \beta X + (I - \lambda W)^{-1}\epsilon$$

오차항에 오차항의 공간자기회귀계수 λ , 공간가중행렬, 오차항의 곱인 $\lambda W u$ 가 추가된 형태를 가지며 식 (2.4)를 통해 공간오차모형은 공간시차모형과 달리 공간승수효과가 오차항에만 영향을 미친다는 사실을 알 수 있다. 모수 추정 방법으로는 2.2절의 공간시차모형과 마찬가지로 최대우도법을 이용하였다.

3. 자료분석

3.1. 모형적합

이 절에서는 2절에서 설명한 공간시차모형과 공간오차모형을 이용하여 모형을 구축해 보았다. 자료 분석에는 통계프로그래밍 언어인 R을 이용하였으며, 분석에 사용한 패키지는 spdep이다. 본 연구에서는 주어진 자료에 공간적 상관관계가 존재하는지를 검증하기 위하여 Moran's I 검정을 시행하였으며, 검정 결과 공간적 상관관계가 유의하게 나타난 2012년 3월의 자료를 이용하였다. 2012년 3월의 자료에 대해 공간가중행렬을 만들고 공간시차모형과 공간오차모형을 적합시켜 다음 년도의 값을 예측해 보고자 한다. 두 모형의 비교 지표로는 RMSE (root mean squared error), RRMSE (root relative mean squared error), MAE (mean absolute error)를 이용하였다.

먼저 분석에 사용한 변수들을 소개하면 Table 3.1과 같다.

Table 3.1 Variables used in the analysis

| Response variable | Variable name | Explanation |
|-----------------------|---------------|---|
| | MED_201203 | Apartment price per unit (m^2) (2012/3) |
| | GRDP_2011 | Gross regional domestic product |
| | POP_2012 | The number of registered resident persons |
| | PBANK_2012 | The number of financial institutions per 10000 persons |
| | SCOMP_2012 | The number of company with more than 1000 employees |
| | NWELF_2012 | The number of welfare facilities |
| | PPOOR_2012 | Proportion of recipients of national basic living subsidies |
| | PHANDI_2012 | Proportion of registered handicapped persons |
| | NCRIME_2012 | The number of crimes occurred |
| | NMOVE_2012 | The number of net migration persons |
| Exploratory variables | RESIAREA_2012 | Size of residential area |
| | CITYAREA_2012 | Size of downtown area |
| | NINDUST_2009 | The number of agricultural and industrial complexes |
| | FOREST_2012 | Proportion of woods and fields |
| | POLD_2012 | Proportion of the aged |
| | HOTEL_2012 | The number of tourist accommodations |
| | NDEAL_2012 | The number of apartment transactions |
| | DISE_2012 | Epidemic rates |
| | MARKET_2012 | The number of markets |
| | VISIT_2012 | The number of visitors |

반응변수인 단위면적당 아파트 매매가격은 각 시 또는 구의 동 별 실거래로 부터 자료의 단위면적당 (제곱미터당) 실거래가격을 구한 다음 각 시 또는 구별로 중앙값을 산출하여 대푯값으로 사용하였다. 분석하는 과정에서 반응변수인 MED_201203은 로그 변환을 취해 사용했다. 또한 설명변수 중에서도 PBANK_2012, PPOOR_2012, PHANDI_2012, POLD_2012는 로그 변환을 하였고, 나머지 설명변수들은 표준화를 시켜 사용하였다.

Table 3.2는 단위면적당 아파트 매매가격에 대해 공간시차모형과 공간오차모형을 이용하여 적합한 결과이다. 이 때 변수 선택 방법은 전진선택법을 사용하였다.

Table 3.2 Parameter estimates of spatial regression models

| | | Spatial lag model | Spatial error model |
|---------------|----------|-------------------|---------------------|
| intercept | Estimate | 10.763213 | 5.733498 |
| | p-value | <2.2e-16 | <2.2e-16 |
| GRDP_2011 | Estimate | -0.152182 | |
| | p-value | 6.102e-06 | |
| PHANDI_2012 | Estimate | -0.525817 | -0.410214 |
| | p-value | 0.035634 | 0.000003725 |
| FOREST_2012 | Estimate | -0.214029 | -0.103869 |
| | p-value | 8.882e-16 | 0.0001213 |
| NMOVE_2012 | Estimate | -0.193103 | -0.120436 |
| | p-value | 1.577e-14 | 0.000001906 |
| DISE_2012 | Estimate | 0.056918 | 0.081914 |
| | p-value | 0.002993 | 0.0008225 |
| HOTEL_2012 | Estimate | | 0.058623 |
| | p-value | | 0.0463622 |
| MARKET_2012 | Estimate | -0.085317 | |
| | p-value | 0.003616 | |
| CITYAREA_2012 | Estimate | 0.120087 | |
| | p-value | 8.290e-06 | |
| VISIT_2012 | Estimate | -0.063783 | |
| | p-value | 0.006364 | |
| POLD_2012 | Estimate | -0.312979 | |
| | p-value | 0.044316 | |
| RMSE | | 0.07651154 | 0.1127734 |
| RRMSE | | 0.01583442 | 0.02398665 |
| MAE | | 0.06243924 | 0.07861246 |

Table 3.3 Test results of spatial regression models

| | | Spatial lag model | Spatial error model |
|---------------|------------|-------------------|---------------------|
| LM | test value | 1.0377 | |
| | p-value | 0.30837 | |
| Breusch-Pegan | BP | 7.4242 | 2.0233 |
| | p-value | 0.593 | 0.8459 |
| BIC | | -17.9446 | -8.552061 |

Table 3.3의 결과를 보면 공간시차모형은 LM (Lagrange multiplier) 검정과 BP (Breusch-Pegan) 검정 결과가 유의하지 않으므로 오차항에 공간자기상관과 이분산이 존재하지 않는다. 공간오차모형은 BP 검정 결과가 유의하지 않으므로 오차항에 이분산이 존재하지 않는다. 두 모형의 적합 결과를 보면 공간시차모형의 BIC (Bayesian information criterion)는 -17.9446, 공간오차모형의 BIC는 -8.552061이다. RMSE, RRMSE, MAE값의 경우 공간오차모형보다 공간시차모형이 더 작은 값을 가진다. 따라서 두 모형 중 적합도가 높은 모형은 공간시차모형이며 유의한 변수로는 PHANDI_2012, FOREST_2012, DISE_2012, MARKET_2012, NMOVE_2012, GRDP_2011, CITYAREA_2012, VISIT_2012, POLD_2012가 있다.

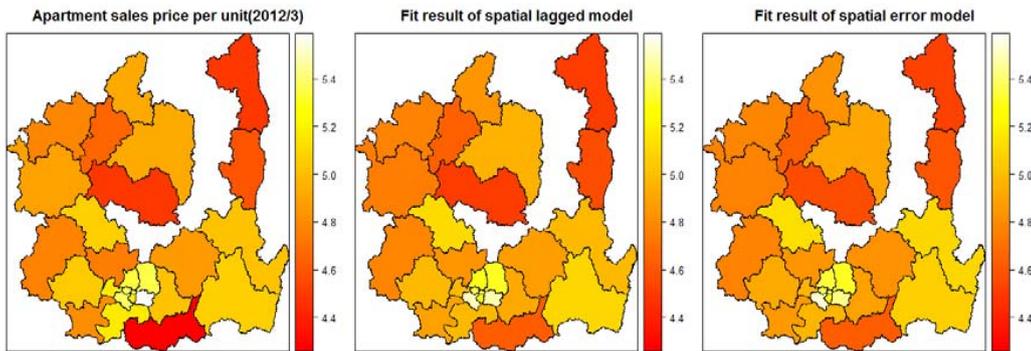


Figure 3.1 Model fitting results of spatial regression models for the apartment price per unit (m^2) in March, 2012

Figure 3.1은 2012년 3월에 대한 모형 적합 결과를 시각화한 것이다. 경상북도 청도군은 공간시차모형과 공간오차모형에서의 적합 값이 실제 값보다 작은 값으로 추정되었다. 이는 청도군의 인근 지역인 경상북도 경산시, 고령시의 적합 값이 실제 값보다 작은 값으로 추정되면서 그 영향을 받은 것으로 보인다. 또한 번두리에 위치해 적합한 인근 지역이 설정되지 못한 것으로 보인다. 경상북도 김천시와 칠곡군의 경우에도 적합도가 조금 떨어지는 것으로 보이는데 공간시차모형 보다는 공간오차모형에서 적합도가 좋았다.

3.2. 예측

2012년 3월에 대한 공간시차모형, 공간오차모형을 통해 1년 후인 2013년 3월의 단위면적당 아파트 매매가격을 예측한 결과는 다음의 Table 3.4와 같다. 예측 결과를 실제 값과 각 모형에 따라 시각화하면 Figure 3.2와 같다.

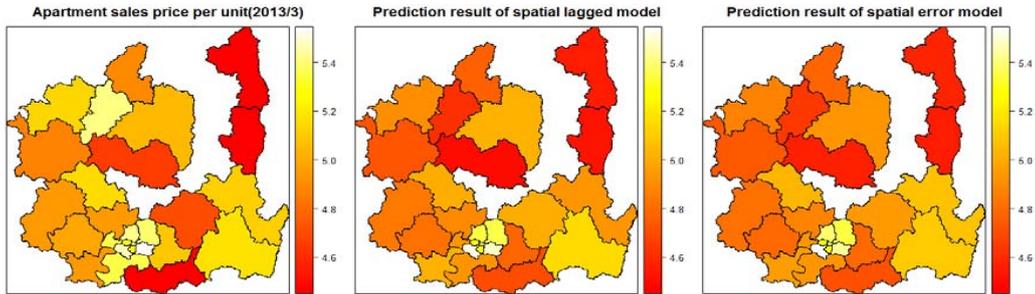


Figure 3.2 Prediction results for the apartment price per unit (m^2) in March, 2013 by spatial regression models

Table 3.4 Prediction results of spatial regression models

| | Area | Spatial lag model | | | Spatial error model | | |
|-------------|---------------|-------------------|------------|----------|---------------------|-----------|----------|
| | | actual | predicted | error | actual | predicted | error |
| Daegu | Jung-gu | 5.382992 | 5.374524 | 0.008468 | 5.382992 | 5.39316 | -0.01017 |
| | Dong-gu | 5.390805 | 5.312837 | 0.077968 | 5.390805 | 5.342354 | 0.048451 |
| | Seo-gu | 5.304573 | 5.229259 | 0.075314 | 5.304573 | 5.251663 | 0.05291 |
| | Nam-gu | 5.270231 | 5.31393 | -0.0437 | 5.270231 | 5.268484 | 0.001746 |
| | Buk-gu | 5.417101 | 5.378731 | 0.03837 | 5.417101 | 5.388376 | 0.028725 |
| | Suseong-gu | 5.546934 | 5.600263 | -0.05333 | 5.546934 | 5.436386 | 0.110548 |
| | Dalseo-gu | 5.46608 | 5.423145 | 0.042935 | 5.46608 | 5.541499 | -0.07542 |
| | Dalseong-gun | 5.357015 | 5.08935 | 0.267665 | 5.357015 | 4.924592 | 0.432423 |
| Gyeongbuk | Pohang-si | 5.11607 | 4.97329 | 0.14278 | 5.11607 | 5.070393 | 0.045677 |
| | Gyeongju-si | 5.180071 | 5.160942 | 0.019129 | 5.180071 | 5.090056 | 0.090015 |
| | Gimcheon-si | 4.944564 | 4.923193 | 0.021371 | 4.944564 | 4.822868 | 0.121696 |
| | Andong-si | 5.042665 | 4.942778 | 0.099886 | 5.042665 | 4.928218 | 0.114447 |
| | Gumi-si | 5.146374 | 5.008186 | 0.138188 | 5.146374 | 5.032814 | 0.113559 |
| | Yeongju-si | 4.892719 | 4.885468 | 0.007251 | 4.892719 | 4.787745 | 0.104974 |
| | Yeongcheon-si | 4.696274 | 4.894314 | -0.19804 | 4.696274 | 4.947708 | -0.25143 |
| | Sangju-si | 4.865857 | 4.817014 | 0.048842 | 4.865857 | 4.737562 | 0.128295 |
| | Mungyeong-si | 5.133264 | 4.884412 | 0.248851 | 5.133264 | 4.875701 | 0.257563 |
| | Gyeongsan-si | 5.081537 | 4.871202 | 0.210335 | 5.081537 | 4.792333 | 0.289204 |
| | Uiseong-gun | 4.65353 | 4.489406 | 0.164124 | 4.65353 | 4.56123 | 0.0923 |
| | Yeongdeok-gun | 4.460841 | 4.424603 | 0.036238 | 4.460841 | 4.542874 | -0.08203 |
| | Cheongdo-gun | 4.450054 | 4.352618 | 0.097436 | 4.450054 | 4.718744 | -0.26869 |
| | Goryeong-gun | 4.940702 | 4.773276 | 0.167425 | 4.940702 | 4.96436 | -0.02366 |
| | Seongju-gun | 4.982003 | 4.883358 | 0.098645 | 4.982003 | 4.793942 | 0.188061 |
| | Chilgok-gun | 4.940339 | 4.898599 | 0.04174 | 4.940339 | 4.867756 | 0.072582 |
| Yecheon-gun | 5.410014 | 4.666478 | 0.743536 | 5.410014 | 4.645328 | 0.764685 | |
| Uljin-gun | 4.471514 | 4.594047 | -0.12253 | 4.471514 | 4.57653 | -0.10502 | |
| | RMSE | | 0.1896103 | | 0.2250451 | | |
| | RRMSE | | 0.03621143 | | 0.04345912 | | |
| | MAE | | 0.1236193 | | 0.1511305 | | |

Table 3.4를 보면 실제 값과 예측 값이 얼마나 차이가 나는지 알 수 있다. 오차를 보면 공간시차모형의 경우 경상북도 예천군의 오차가 가장 큰 것으로 나타났고, 대구광역시 달성군, 경상북도 문경시 및 경산시도 비교적 오차가 큰 것으로 나타났다. 공간오차모형의 경우에도 마찬가지로 경상북도 예천군의 오차가 가장 큰 것으로 나타났고, 대구광역시 달성군과 경상북도 청도군의 오차가 공간시차모형에 비해 좀 더 커진 것을 알 수 있다. 마찬가지로 경상북도 경산시, 경상북도 문경시의 오차가 비교적 큰 것으로 나타났다.

Figure 3.2에 의하면 공간오차모형 보다는 공간시차모형의 예측력이 미세하지만 좋은 것으로 나타났다. 구체적으로 실제 값과 공간시차모형의 예측 값을 비교해보면 대체적으로 값이 작게 예측되었다. 특히 경상북도 예천군과 그 주변지역의 예측 값의 정확도가 크게 떨어지는데 이는 도청 이전으로 인하여 경상북도 예천군과 그 주변지역의 지가가 크게 상승함으로써 아파트 매매가격 또한 상승했기 때문으로 보인다.

전체적인 예측의 정확성을 좀 더 객관적으로 비교하기 위하여 RMSE, RRMSE, MAE를 이용하였다. Table 3.4를 보면 RMSE, RRMSE, MAE 모두 공간오차모형이 공간시차모형보다 큰 값을 가지는 것을 알 수 있다. 따라서 공간시차모형이 공간오차모형에 비해 예측의 정확도가 높다고 할 수 있다.

4. 결론

아직까지 공간회귀모형을 이용하여 아파트 매매가격을 예측한 사례가 그리 많지 않았다. 본 연구에서는 공간회귀모형 중에서도 가장 많이 쓰이는 공간시차모형과 공간오차모형을 이용하여 단위면적당 (제곱미터당) 아파트 매매가격을 예측하였다. 2012년 3월의 단위면적당 아파트 매매가격에 대해 모형을 적합시켰을 때 두 모형 모두 어느 정도 적합이 잘 되었으나, 공통적으로 경상북도 청도군에 대해서는 적합도가 크게 떨어졌다. 이는 주변 지역들로부터 받은 영향이 모형에 반영되어 주변 지역에 비해 매매가격이 낮은 청도의 적합도가 좋지 않게 나타난 것으로 보인다. 적합된 모형으로 예측을 했을 때 두 모형 모두 공통적으로 네 지역에서 예측력이 유독 떨어지는 모습을 보였다. 특히 경상북도 예천군의 예측력이 가장 떨어지는데 이는 도청 이전으로 인해 예천군의 지가가 상승하면서 아파트 매매가격도 상승한 것으로 보인다. 지가와 아파트 가격의 인과관계에 대한 대부분의 선행연구에서 지가 상승이 아파트 가격의 상승에 영향을 미친다는 결과를 제시하고 있었다. 하지만 우리나라에서 진행된 연구에서는 아직까지 대구·경북을 대상으로 한 연구가 없어 대구·경북의 아파트 매매가격 상승에 지가 상승이 영향을 미친다고 확신할 수는 없다. 공간시차모형과 공간오차모형을 비교하기 위해 RMSE, RRMSE, MAE를 사용하였으며 공간오차모형 보다는 공간시차모형의 예측력이 좋은 것으로 나타났다.

아파트 거래가 이루어지지 않거나 너무 적은 수의 거래가 이루어져 표본선정에서 탈락된 경북의 몇몇 군으로 인해 대구·경북의 모든 지역이 포함된 데이터를 얻을 수 없었기 때문에 정확한 공간가중행렬을 구축하는데 어려움이 있었다. 따라서 대구·경북의 모든 지역이 포함된 데이터를 얻을 수 있다면 공간가중행렬의 정확도가 향상되어 좀 더 나은 결과를 얻을 수 있으리라 생각된다.

References

- Anselin, L. (1988). *Spatial econometrics: Method and models*, Kluwer Academic Publishers, Dordrecht.
- Anselin, L. (2001). *Spatial externalities, spatial multipliers and spatial econometrics*, Regional Economics Applications Laboratory, Illinois.
- Chun, H. (2002). *A study on the living culture of apartment*, Master Thesis, Yonsei University, Seoul.
- Chung, K., Kim, S. and Lee, Y. (2012). A comparative study on the goodness of fit in spatial econometric models using housing transaction prices of Busan, Korea. *Journal of Korean Association of Geographic Information Studies*, **15**, 43-51.
- Hur, Y. (2007). A study on the estimation of house price in regard of spatial effects. *House Studies Review*, **15**, 5-23.
- Park, H., Jyoung, S. and Rho, T. (2003). An spatial econometric analysis of apartment sale prices and spatial effects. *Journal of Korea Planners Association*, **38**, 115-125.
- Stakhoych, S. and Bijmolt, T. (2008). Specification of spatial models: A simulation study on weights matrices. *Papers in Regional Science*, **88**, 389-408.

Prediction of apartment prices per unit in Daegu-Gyeongbuk areas by spatial regression models

Woo Jung Lee¹ · Cheolyong Park²

^{1,2}Department of Statistics, Keimyung University

Received 6 March 2015, revised 18 March 2015, accepted 23 March 2015

Abstract

In this study we predict apartment prices per unit in Daegu-Gyeongbuk areas by spatial lag and spatial error models, both of which belong to so-called spatial regression model. A spatial weight matrix is constructed by k -nearest neighbours method and then the models for the apartment prices in March, 2012 are fitted using the weight matrix. The apartment prices in March, 2013 are predicted by the fitted spatial regression models and then performances of two spatial regression models are compared by RMSE (root mean squared error), RRMSE (root relative mean squared error), MAE (mean absolute error).

Keywords: Apartment price, spatial error model, spatial lag model, spatial regression model.

¹ Master candidate, Department of Statistics, Keimyung University, Daegu 704-701, Korea.

² Corresponding author: Professor, Department of Statistics, Keimyung University, Daegu 704-701, Korea. E-mail: cypark1@kmu.ac.kr