http://dx.doi.org/10.5392/JKCA.2012.12.02.461

공간적탐색기법을 이용한 부산 주택시장 다이나믹스 분석

Busan Housing Market Dynamics Analysis with ESDA using MATLAB Application

정건섭

부경대학교 행정학과

Kyoun-Sup Chung(kschung39@pknu.ac.kr)

요약

본 논문의 목적은 공간적탐색기법을 이용한 부산 주택시장 다이나믹스 분석으로써 MATLAB toolbox M-file을 이용하였다. 본 연구에서 사용된 자료는 2006년부터 2009년 2분기까지 공개된 부산지역 아파트 실거래가 64,530개 자료를 기준으로 법정동을 분류하여 각 평균값을 분석에 이용하였다. 주택시장분석에 많이 이용되는 헤도닉가격 모형은 도시주택경제 분야에서 주택시장 다이나믹스를 설명하는데 강력한 분석기법의 하나이다. 그럼에도 불구하고 전통적인 헤도닉가격 모형은 공간적자기상관의 영향력을 반영하지 않는 단점이 있다. 따라서 본 논문에서는 공간자기상관 관계를 반영한 다양한 공간계량모형, 예를 들어, 공간자기회귀모형(SAR), 공간오차모형(SEM), 일반공간모형(SAC) 등을 보통최소자승법을 이용한 전통적 헤도닉가격 모형과 비교하고자 한다. 이를 위해 결정계수(R^2), 분산(σ^2), 우도함수(Likelihood)의 값 등의 지표들을 이용하였다. 분석결과 공간자기상관을 고려한 공간계량모형이 전통적 헤도닉모형에 비해 높은 설명력을 보여주고 있다. 공간계량모형에서는 공간오차모형(SEM)과 일반공간모형(SAC)이 공간자기회귀모형(SAR) 보다 우수한 설명력을 보이고 있다.

■ 중심어 : | 공간적탐색기법 | 공간계량모형 | 공간자기상관 |

Abstract

The purpose of this paper is to visualize the housing market dynamics with ESDA (Exploratory Spatial Data Analysis) using MATLAB toolbox, in terms of the modeling housing market dynamics in the Busan Metropolitan City. The data are used the real housing price transaction records in Busan from the first quarter of 2006 to the second quarter of 2009. Hedonic house price model, which is not reflecting spatial autocorrelation, has been a powerful tool in understanding housing market dynamics in urban housing economics. This study considers spatial autocorrelation in order to improve the traditional hedonic model which is based on OLS(Ordinary Least Squares) method. The study is, also, investigated the comparison in terms of R^2, Sigma Square(σ^2), Likelihood(LR) among spatial econometrics models such as SAR(Spatial Autoregressive Models), SEM(Spatial Errors Models), and SAC(General Spatial Models). The major finding of the study is that the SAR, SEM, SAC are far better than the traditional OLS model, considering the various indicators. In addition, the SEM and the SAC are superior to the SAR.

■ keyword : | ESDA(Exploratory Spatial Data Analysis) | Spatial Econometrics Models | Spatial Autocorrelation |

* 이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 연구되었음(NRF-2011-371-B00037).

접수번호 : #111026-001

접수일자 : 2011년 10월 26일

심사완료일 : 2012년 01월 30일

교신저자 : 정건섭, e-mail : kschung39@pknu.ac.kr

I. 서 론

Anselin(1988)에 의해 학문적인 토대가 공고히 된 소위 공간계량경제학1)을[1] 바탕으로 이를 이용한 공간계량분석은 강력한 소프트웨어의(MATLAB, R 등) 발달과 더불어 최근 다양한 공간모형 부문에서 급속도로 확산되고 있는 실정이다. 특히 주택시장분석에 있어서 공간계량모형과 G.I.S(Geographic Information Systems)의 연계를 통한 주택시장분석은 이미 널리 사용되어지고 있는 것은 주지의 사실이다. 이와 더불어 MATLAB을 이용한 공간계량분석[2] 및 기초적인 기본 지도를 작성할 수 있는 G.I.S[3]의 도입 및 활용은 LeSage 및 Pace 교수 등에 의해서 일반인들도 쉽게 활용할 수 있게 되었다. MATLAB를 사용한 다양한 예제 프로그램의 Toolbox 제공으로 더욱더 그 적용 및 다양한 분석을 넓혀 놓았다.

LeSage(1998)에 의하면 공간종속성(spatial dependence)의 발생 원인은 크게 두 가지로 구분할 수 있다. 첫째는 자료수집의 공간단위에 의해 발생한다는 것이다. 연구자가 사용하게 될 자료의 지역 간 단위는 흔히 행정구역별 또는 경도-위도 좌표 등으로 구성된다. 그러나 이러한 지역 간 구분이 수행될 연구의 바람직한 구분 단위인지는 확인하기가 쉽지 않다. 따라서 수집된 자료의 지역 간 단위가 수집된 지역과 일치한 않는 경우에 발생된다. 둘째는 인구사회학적인 요인에 의해서 발생된다. 인간의 행위는 특정 공간에 고정된 채 이루어지는 것이 아니라 공간들을 이동하면서 이루어지는 행위가 대부분이다. 이러한 사회·인구학적 요인은 공간종속성(spatial dependence)2) 발생의 가장 중요한 이유인 동시에 필연적인 성격을 지니고 있다는 점에서 매우 중요하다고 할 것이다.

공간이질성(spatial heterogeneity)은 공간이 종속변수의 결정에 미치는 영향이 균일하게 나타나지 않음을 뜻한다. 즉, $y_i = f_i(X_i\beta_i + \varepsilon_i)$는 일반 선형회귀식인 $y_i = X_i\beta_i + \varepsilon_i$와는 달리 일반선형회귀식의 또 다른 함수관계를 형성하고 있음을 뜻한다. 즉, 흡사 매우 유사하게 설명변수를 가진 지역이라 할지라도 지역에 따라 다른 종속변수의 값을 가지게 됨을 뜻한다[4].

이와 같이 공간종속성(spatial dependence) 및 또는 공간이질성(spatial heterogeneity)의 개념이 포함 된 공간자기상관(spatial autocorrelation)은 모수의 추정에 있어서도 불편(unbiased) 추정량, 효율적(efficient) 추정량 및 일치(consistent) 추정량을 갖지 못하게 된다. 더욱더 큰 문제점은 이러한 현상들이 모형의 잘못된 선택(mis-specification), 이분성(heteroscedasticity), 또는 공간자기상관(spatial autocorrelation)에 의해서 발생하는지 그것을 구분하기가 현실적으로 매우 힘들다는 것이다.

주택시장의 특성 및 주택가격을 추정하는 방법으로 전통적으로 내재된 가치에 기초한 헤도닉 가격 모형(hedonic model)이 이용되어 왔다. 헤도닉모형은 Rosen(1974)[5]의 연구 이후 주택구성요소들의 이질성을 실증적으로 분석할 수 있는 적절한 수단이라고 일반적으로 인식되고 있다. Rosen(1974) 모형은 개별 주택구성요소에 대한 주택소비자들의 주관적 평가에 기초한 입찰가격(bid price)과 공급자의 공급가격(offer price)간의 상관관계에 대한 이론적 기반을 제공함으로써 주택특성가격을 분석하는 이론적 기초를 제공하고 있다[6].

많은 선행 연구들의 이론적 기반으로 삼고 있는 헤도닉가격모형(hedonic price model)은 주택가격함수 추정과 관련하여 다중공선성(multicollinearity)과 공간자기상관(spatial autocorrelation), 공간종속성(spatial dependence), 공간이질성(spatial heterogeneity)에 대한 문제점을 항상 내포하고 있다[7].

특히, 주택시장 분석에 있어 널리 사용되는 헤도닉(hedonic)분석 기법의 연구들 대부분이 OLS(ordinary least square)방법을 이용해 모수를 추정하고 있다{8}. 이러한 추정 방법은 잔차(residuals)는 독립적이며, 평균이 '0'이고 분산이 일정하며, 공분산이 '0'이라는 가정을 기초하고 있다. 하지만 공간적 자기상관(spatial

1) According to Anselin(1988), "The collection of techniques that deal with the peculiarities caused by space in the statistical analysis of regional science models is considered to be the domain of spatial econometrics."

2) $y_i = f(y_i)$ 단, $i = 1,2,\ldots n \quad i \neq j$

autocorrelation)이 발생할 경우 이러한 가정을 위반하게 되고, 이것으로 인해 왜곡된 추정결과를 낳는다[8]. 박헌수 등(2003)도 공간적 자기상관 등의 공간효과(spatial effect)를 제대로 반영하지 않을 경우 모수가 과대추정(over estimation)되기 때문에 유의할 필요가 있다고 주장했다[9].

따라서 주택가격을 추정하는데 있어 공간자기상관(spatial autocorrelation) 혹은 공간종속성(spatial dependence) 및 공간이질성(spatial heterogeneity)을 고려한 공간계량모형---공간자기회귀모형(SAR:Spatial Autoregressive Model), 공간오차모형(SEM: Spatial Errors Model)모형, 일반공간모형(SAC: General Spatial Model) 모형 등---이 기존 헤도닉가격함수(hedonic price function) 추정을 위해 사용되는 OLS방법과의 비교평가를 통해 어떤 추정방법이 보다 우수한 추청치를 제공하는지 살펴보는 것도 의미가 있다고 하겠다.

이를 위하여 다음의 II장에서는 몇 가지 선택된 공간계량모형에 대하여 간략하게 설명하고, III장은 공간계량모형의 필수요소 중에 하나인 공간가중행렬의 구성 및 대표적 사례에 대하여 언급하였다. IV장은 Moran's I를 중심으로 공간자기상관 검정 방법 등에 관하여 설명한다. 그리고 V장에서는 본 연구에 사용된 분석자료 및 변수에 대한 간략한 설명을 하였으며, 마지막 VI장은 모형의 결과 및 해석 그리고 간단하게나마 정책적 시사점을 도출하고자 한다.

II. 공간계량모형

공간계량모형에 앞서 기존의 전통적인 헤도닉 특성가격함수 모형을 간략하게 살펴보면 다음과 같다.

특성가격함수모형(HPM: Hedonic Price Models)은 주택가격 평가에 널리 사용되고 있는 방법으로 이론적 기초는 Rosen(1974)의 연구인 다중특성을 가진 단일재화시장에 대한 연구가 기초를 이루고 있다. 주택에 대한 헤도닉모형은 주택가격을 주택상품 혹은 서비스를 구성하고 있는 개별 특성들의 함수로 표시한 것이다. 일반적인 주택가격 함수는 식(1)과 같이 나타낸다.

$$P = h(S, N, L) \tag{1}$$

여기에서 P는 주택의 가격이고, S(구조적 변수들: Structural variables), N(근린적 변수들: Neighborhood variables), L(장소적·지역적 변수들: Locational variables)은 주택의 개별특성을 의미한다. 이러한 헤도닉 함수는 일반적으로 선형함수 형태로 아래 식(2)와 같은 기본모형을 갖는다.

$$y = \beta X + \epsilon \qquad \epsilon \sim N(0, \sigma^2 I_n) \tag{2}$$

공간계량모형은 일반선형모형에서 공간가중치행렬을 부가한 형태로 일반선형모형의 확장된 형태이다.

먼저 공간자기회귀모형(SAR)은 기본 헤도닉가격함수 모형에 설명변수 행렬을 첨가하여 준 형태로 구성된다. 연구자에 따라 공간시차모형(spatial lagged model) 또는 혼합자기회귀모형(mixed autoregressive-regressive model)로 명명되기도 하며, 시계열 분석과 유사한 형태를 띤다.

$$y = \rho W y + \beta X + \epsilon \quad \epsilon \sim N(0, \sigma^2 I_n) \tag{3}$$

여기서 W는 공간가중행렬이다. ρ는 공간자기회귀계수이고, X는 주택 속성별 변수의 벡터를 나타낸다. 다음 공간오차모형(SEM) 모형은 오차항에 공간자기상관을 포함하고 있는 모형으로 다음과 같이 정의된다.

$$\begin{aligned} y &= \beta X + u \\ u &= \lambda W u + \epsilon \qquad \epsilon \sim N(0, \sigma^2 I_n) \end{aligned} \tag{4}$$

이때 λ는 모형의 설명변수가 아닌 관측 불가능한 생략된 변수의 오차항에 대한 충격에 속한다. 다시 말하면 어느 특정지점의 주택가격은 채택된 설명변수 뿐만 아니라 생략된 변수의 함수라는 것이다[10].

SAC 모형은 SAR과 SEM을 혼합한 모형으로 가장 일반적으로 공간종속성이 공간자기회귀형태와 오차항 모두에 포함되어 있다.

$$y = \rho W_1 y + \beta X + u$$
$$u = \lambda W_2 u + \epsilon \qquad \epsilon \sim N(0, \sigma^2 I_n) \qquad (5)$$

III. 공간가중행렬

1. 공간가중행렬의 구성

본 연구에서의 공간가중치행렬(weight matrix)은 지역 간(구간)거리와 인접지역(contiguity) 여부를 기준으로 분석하고자 한다.3) 공간상에 분포해 있는 실체들의 관계를 정의하는 공간가중행렬(spatial weight matrix)은 주로 인접성척도(contiguity measure)와 거리척도(distance measure)를 기준으로 하고 있다[11].

공간인접성(spatial contiguity)을 수량화하는 방법들은 다음과 같이 요약할 수 있다. 우선 Linear Contiguity는 특정지역의 경계선에 공유된 다른 지역의 경계선이 완전히 공유된 경우 '1'의가중치를 부여하고(W_{ij}=1), 다른 경우에는 '0'의 가중치를 부여(W_{ij}=0)한다. Rook Contiguity는 상하좌우 경계선이 공유되면 '1'의가중치를 부여하고(W_{ij}=1), 그렇지 않으면 '0'(W_{ij}='0')을 부여한다. Bishop Contiguity는 두 지역의 꼭짓점이 연결된 경우 '1'의가중치를 부여하고 그렇지 않으면 '0'을 부여하는 형식이다. Queen Contiguity는 Rook, Bishop Contiguity을 종합한 것으로 두 지역이 한쪽 면이나 모서리를 공유하면 '1'의가중치를 부여하고(W_{ij}=1), 그렇지 않으면 '0'(W_{ij}=0)의 가중값을 부여하는 방법이다.4)

이때 공간가중행렬(spatial weight matrix)을 그대로 모형에 적용시킬 경우 해석상 중대한 오류가 발생한다.

따라서 이문제 해결을 위하여 공간가중행렬은 횡단표준화(row standardization)를 통해 이웃의 값이 한 지점에 평균적으로 얼마나 영향을 미치는지를 계량화 할 수 있다[12].

$$W^s_{ij} = W_{ij} / \sum_j W_{ij}, \quad (단, \sum_j W^s_{ij} = 1) \qquad (6)$$

인접성척도(contiguity measure)는 공간이 가지는 공간배열을 기준을 간접적으로 정의하는 반면 거리척도(distance measure)는 직접적으로 공간이 갖는 관계를 정의한다. 두 공간 실체의 거리를 공간가중행렬(spatial weight matrix)로 이용할 경우 거리가 멀어질수록 서로의 관계가 약화 될 수 있도록 식(7)과 같이 정의 한다.

$$W_{ij} = \frac{1}{d^{\alpha}_{ij}}, \ (\ \alpha=1,2,3,...) \qquad (7)$$

여기서 W_{ij}는 행렬 W의 i번째 열과 j번째 열의 요소를 의미하고, d_{ij}는 두 공간단위 i와 j의 거리를 나타낸다.

하지만 거리척도(distance measure)를 기준으로 공간가중행렬이 구성된 대부분의 연구에 있어 실질적인 거리를[7] 이용하기 보다는 두 공간간의 임계치를 설정하고, 두 공간단위의 거리가 임계치(cut-off point or bandwidth)보다 클 경우 두 주택간의 직접적인 연관성이 없는 것으로 아래 식(8)과 같이 가장 가중치 행렬을 구성하였다. 일부 연구에서는 구 단위 또 동 단위 자료를 동일 권역 내에 있는 것으로 가정하여 거리가중행렬을 구성하였다.

$$W_{ij} = 1 \text{ if } d_{ij} < D_{ij}, \ D_{ij}=2,3,4...(단위 ㎞) \qquad (8)$$

이외에도 근접한 이웃의 수(K-nearest neighbors)를 통해 가장 가까운 자료의 개수를 이용하거나 경제적 가중치(economic weight)로서 경제적인 차이를 이용하여 공간가중행렬을 구성할 수 있다.

공간가중행렬의 구성과 관련하여 기존연구를 종합해 볼 때 대부분의 연구가 관측 자료간의 거리를 기준으로 분석하였다. 하지만 이들 대부분의 연구가 임계값을 기준으로 공간자기상관의 존재를 가정하고 있어,

3) 본 연구에서의 실증분석은 Rook을 사용하여 분석하였다.

4) 공간인접성의 수량화

0	0	0
1	A	1
0	0	0

Linear

0	1	0
1	A	1
0	1	0

Rook

1	0	1
0	A	0
1	0	1

Bishop

1	1	1
1	A	1
1	1	1

Queen

자료) 최명섭 외,(2003:69) : 재정리

Dubin(1998)이 주장한 바와 같이 연구자의 자의적인 선택에 따른 문제점을 내포하고 있다[13].

공간계량모형의 분석에 있어 거의 모든 공간가중행렬을 이용한 관련 논문들은 인접성(contiguity)에 의한 방법을 사용하였고, 거리를 사용한 흔치 않은 논문들의 경우에도 실질적으로는 일정거리 이하는 '1', 그 이상의 거리는 '0'과 같은 공간가중치 방식(distance bandwidth)을 사용하고 있어 그 본질 면에서는 인접성 공간가중행렬과 큰 차이는 없다고 할 수 있다.

공간계량모형에 있어 한 지역의 주택가격의 변화가 인근지역에 주택가격에 영향을 미치는 공간적 자기상관을 파악하기 위해서는 공간적 효과를 나타내는 공간가중행렬의 정의하는 것이 매우 중요하다.

공간계량모형의 설정에 있어 공간가중행렬(spatial weight matrix)을 정의하는 다양한 방법들이 문헌에서 소개되고 있지만, 특정 공간가중행렬의 선택이 분석 결과에 어떤 영향을 미치는지에 대해서는 논의는 그다지 많지 않다.

그리고 대부분의 연구에서는 통계 package에서 제공하는 기준에 따라 여과 없이 공간가중행렬이 사용되고 있는 실정이다[14].

2. 공간가중행렬을 이용한 대표적 사례

공간가중행렬을 이용한 대표적 사례를 몇 가지 꼽자면 다음과 같다. 첫째, 인접성측도(contiguity measure)를 기준으로 Anselin(1988)은 범죄와 소득 및 주택가격과의 관계분석을 위해 Columbus 49개 지역의 인접성(contiguity)자료를 이용하여 공간가중행렬을 구성하였다. 이 때 공간인접성(spatial contiguity)의 수량화 방식으로는 Rook Contiguity방식이 사용되었다.

둘째, 근접성(nearest)을 기준으로 공간가중행렬을 구성하는 것으로 Pace(1997)은 i와 j지점간(3109×3109)의 유클리디안(Euclidean) 거리(d_{ij})를 기준으로 m번째로 가까운 관측치를 $d_{\max i}$와 비교를 통해 이것을 공간가중행렬을 구성, 표준화(standardized)하여 분석에 이용하였다. Dubin(1998) 역시 근접한 이웃의 수(1,2,3)을 기준으로 자료를 행의 합이 '1'이 되도록 횡단표준화 작업을 거쳐 공간가중행렬을 구성하였다.

셋째, 특정 임계값을 기준으로 공간자기상관의 가정하여 공간가중행렬을 구성하는 형태로 Can(1992), Dubin(1998) 뿐만 아니라 국내 김종원(2000)[15], 서경천・이성호(2001)[16], 허윤경(2007)[17] 등 많은 연구에서 이용되었다. 이러한 형태는 근접성(nearest)을 기준의 공간가중행렬의 구성과 개념적으로 유사한 형태로 어떠한 임계값을 선택하느냐에 따라 공간상관계수의 값과 모양(pattern)이 차이 나게 된다[13]..

넷째, 실질적인 거리를 기준으로 공간가중행렬을 구성한 것으로 Can(1992)의 경우[18] 식(7)의 α값이 1, 2인 경우를 구분하여 분석한 결과 α=2에서 Log Likelihood 값이 보다 우수한 것으로 결론지었다. Dubin(1998)은 α= 1, 2, 3으로 각각의 공간가중행렬을 구성하여 공간간의 상관관계를 분석한 결과 α값이 증가 할수록 영향력의 범위(band of influence)가 줄어들고, 공간 상관계수의 분산이 늘어나는 것으로 조사됐다.

국내연구의 경우 박헌수 외(2003), 이현석・박성균(2010)[19], 김성우 외(2010a,b,c,d)[20]등 몇몇 연구에서 이용되었다.

IV. 공간자기상관 검정

공간계량모형을 적용하기에 앞서 공간자기상관이 존재하는지를 살펴 볼 필요가 있다. 공간자기상관(spatial autocorrelation)을 검정하는 대표적인 방법으로는 Moran's I 검증, Geary's의 C Coefficient 등이 있다. 이외에도 MLE(Maximum Likelihood Estimator)의 모형 최적화검증으로 활용되는 Wald(W) 검정, Likelihood Ratio(LR), 그리고 Lagrange Multiplier(LM) 검정 방법들이 있다[4]. Tobler(1970)에 따르면 "모든 것은 다른 모든 것과 관계를 갖지만, 가까운 것이 먼 것보다 더 밀접한 관련성을 갖는다"고 한다[21]. 즉, 공간간에 거리나 인접성이 높을수록 더 큰 영향관계를 가진다. 공간상의 상호작용을 공간자기상관이라고 한다. 공간자기상관을 검증하는 대표적인 방법은 Moran's I 검증이 있

다.

Moran's I의 계수는 공간자기상관을 파악하기 위한 유용한 측정도구로 인접해 있는 공간단위들의 값을 비교하여 이 계수를 산출한다. 만일 인접한 공간단위들이 전체지역에 걸쳐 유사한 값을 갖는 경우 Moran's I의 계수가 높게 나타나는 반면, 서로 상이한 값들을 갖게 되면 Moran's I는 부정적 공간자기상관을 갖게 된다. Moran's I의 추정식은 다음과 같다[22].

$$I=\frac{n\sum\sum W_{ij}(X_i-\overline{X})(X_j-\overline{X})}{W\sum(X_i-\overline{X})^2} \qquad (9)$$

Wald(W), Likelihood Ratio(LR), Lagrange Multiplier (LM)의 추정식은 다음과 같다[4].

$$W=\lambda^2[t_2+t_3-(1/n)(t^{2_1})]\sim\chi^2_{(1)} \qquad (10)$$

여기서 $t_1=tr(WB^{-1}),\ t_2=tr(WB^{-1})^2,$
$t_3=tr(WB^{-1})'(WB^{-1}),$
$B=(I_n-\lambda W)$

$$LR=n[\ln(\sigma_0^2)-\ln(\sigma_1^2)]+2\ln|I_n-\lambda W|\sim\chi^2_{(1)} \qquad (11)$$

여기서 σ_0^2은 Spatial Dependence가 없다는 가정하의 잔차의 분산, σ_1^2은 Spatial Dependence가 있다는 가정하의 잔차의 분산이다.

$$LM_{error}=\frac{1}{T}\left[\frac{e'We}{\sigma^2}\right]\sim\chi^2(1)$$

$$\text{단 } T=tr(W+W')W \qquad (12)$$

Moran's I는 표준정규분포를, W, LR, LM은 $\chi^2(1)$을 따르고 있다.5)

V. 주택매매 실거래가를 이용한 실증분석

1. 분석자료

본 연구에 사용된 자료는 국토해양부 실거래가 홈페이지[23]에 2006년부터 2009년 2분기까지 공개된 부산지역 아파트 실거래가 64,530개 자료를 기준으로 법정동을 분류하여 각 평균값을 분석에 이용하였다.6) 분석에 사용된 단지의 위치는 다음의 [그림 1]과 같다.

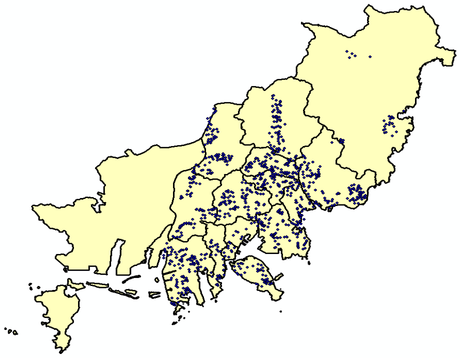

그림 1. 분석자료의 위치

2. 변수의 설정

본 연구에 사용된 변수들은 구체적으로 [표 1]과 같다.

표 1. 변수의 정의

변수		변수의 정의	비고
종속변수	실거래가격	국토해양부 아파트 실거래가격 (2006년~2009년)	만원
독립변수	주택면적	주택면적 (전용면적+기타공용면적)	평 (3.3058m^2)
	역세권	역세권여부 (0 = 도보 10분 이상 1 = 도보 10분 이내)	dummy
	재건축	재건축 추진여부 (0 = 재건축 비추진, 1 = 재건축 추진 중)	dummy
	총세대수	개별 아파트 총세대수	호
	경과년수	2010년 기준 아파트 경과년수 (2009년 입주 = 1, 2008년 입주 = 2,·)	년

분석 자료에 대한 기초통계량은 [표 2]와 같다. 아파

5) 따라서 Moran's I의 경우 임계값인 1.96보다 통계량이 클 경우, Lagrange Multiplier(LM) Error는 제시된 $\chi^2(1)$값 보다 통계량이 클 경우 귀무가설인 $H_0:\lambda=0$ 귀무가설(공간자기상관이 존재하지 않음)이 기각된다.

6) 전체 64,530의 자료를 법정동을 평균을 이용한 것은 분명 정보의 손실(lost information) 측면에서 좋은 방법이라 볼 수 없다. 다만 불편추정량의 입장에서는 수용 가능한 방법이라 여겨지며 특히 본 논문에서는 동별의 공간자기상관 영향력을 시각적으로 보이고자 하였다. 또한 전체 자료를 이용한 결과와도 비교해서 모수의 추정값 및 표준오차가 정확하게 일치하지 않지만 비슷한 수준으로 추정되었다.

트 실거래가 법정동별 3년간 평균값은 13,292만원으로 나타났다. 주택 면적은 평균 30.62평, 총세대수는 평균 579세대, 경과년수는 평균 14.42년을 보이고 있다. 법정동별 평균 단지의 최고층 높이는 18.70으로 나타났다. 더미변수를 사용한 역세권의 평균값은 0.55, 재건축 0.67으로 나타나고 있다.

표 2. 기초통계량

	n	평균	표준편차	최소값	최대값
실거래가	90	13,292.3	4,513,18	6,130	27,601
주택면적	90	30.62	4.90	18.56	52.0
역세권	90	0.55	0.42	0	1
재건축	90	0.67	0.15	0	1
총세대수	90	579.49	397.02	29	1,781
경과년수	90	14.42	5.58	5	34

경과년수의 경우 제곱과, 세제곱을 독립변수로 이용하였는데, 이는 아파트의 재건축에 따른 영향으로 경과연수가 18~35년 사이에 주택가격이 상승하는 경향을 [그림 2] 보이고 있기 때문이다[7].

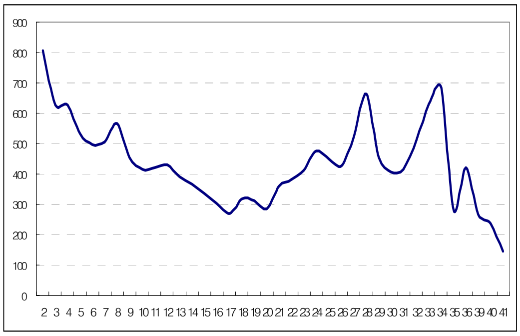

그림 2. 경과년수별 평균 평당 매매가격 곡선

다음은 MATLAB[7] 이용하여 각 동별 평균 주택매매가격을 [그림 3]과 같이 나타내어 보았다. 여기에서 볼 수 있듯이, 기장군을 포함한 서부산 지역 및 북구지역 동들이 적은 매매가격의 거래가 지배적인 반면 해운대, 남구 및 동래지역에 입지한 동들이 높은 매매가격이 형성됨을 알 수 있다.

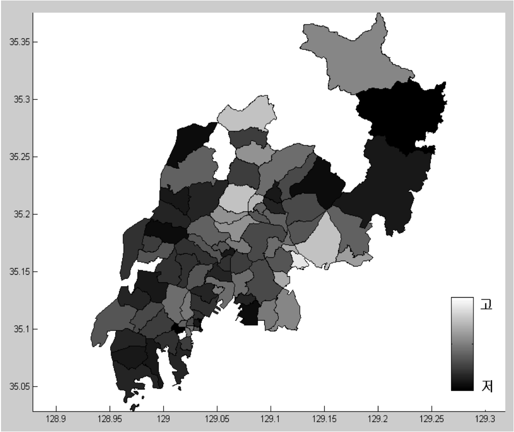

그림 3. 동별 평균 매매가격 분포도

다음의 [그림 4]는 Rook 가중행렬을 이용한 Moran's I의 분포도 이다. 대부분의 지역의 정(positive)의 공간자기상관관계를 보이고 있으며, 특히 해운대 지역 남구 및 동래구 지역의 동들이 높은 자기상관관계를 보이고 있음을 알 수 있다. 공간자기상관 검정통계량은 Moran I = 0.34, Moran I-statistic = 3.89, Marginal Probability = 0.0001으로 나타나고 있어 통계적으로 유의미한 정의 공간자기상관관계가 있음을 알 수 있다. 이밖에도 Wald, Log Likelihood, Lagrange Multiplier 모두 정의 공간자기상관계가 있음을 보이고 있다.[8]

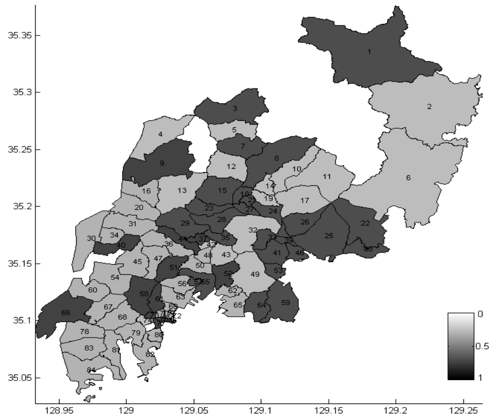

그림 4. Moran's I을 이용한 공간자기상관 분포도

7) 본 내용의 결과를 알 수 있도록 MATLAB의 coding 및 여기에서 사용된 자료는 저자에게(kschung39@pknu.ac.kr) 요구하면 받아 볼 수 있다.

8) Wald(39.18), LR(15.09), LM(12.81) 모두 $\chi^2_{(1)}$의 임계값 6.635보다 크게 나타나고 있다.

VI. 모형의 결과 및 해석

앞서 간략하게 언급하였듯이 어떠한 가중행렬(소위 weight matrix)을 사용하는가에 따라 모형의 추정값은 달라 질 수 있으며, 어떠한 공간가중행렬이 가장 적합한가에 대해서는 아직도 수많은 논의가 진행 중에 있다고 할 수 있다. 여기에서는 공간가중행렬에 따른 정확한 추정값의 예측 보다는 공간자기상관이 발견 되어질 때 다양한 공간계량모형의 적용 예를 간단하게 보여주는 것으로 한다.

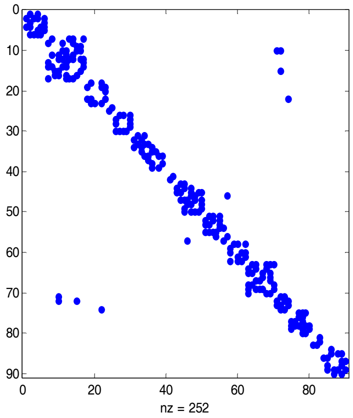

그림 5. 각 동의 공간인접행렬

앞의 [그림 5]는 각 동의 인접행렬을 그림으로 나타낸 것이다. 전체 8100개(90개동*90개동)의 이웃 조합에서 '0'이 아닌 252개동의 조합이 있음을 알 수 있으며 인접공간행렬(Weight Matrix by Rook)의 대부분[9]이 '0'(Sparse matrix)으로 되어 있음을 볼 수 있겠다.

다음의 [표 3]은 앞서 언급한 공간자기회귀모형(SAR: Spatial Autoregressive Model), 공간오차모형(SEM: Spatial Errors Model)모형, 일반공간모형(SAC: General Spatial Model) 모형들의 결과를 보여주고 있다. 보통최소자승법(OLS: Ordinary Least Squares)을 이용한 기존 헤도닉가격함수(hedonic price function) 모형도 공간모형들과의 비교를 위해 참고로 같이 추가하였다.

9) 8100 - 252 = 7848개의 '0'

표 3. Rook을 이용한 공간계량모형 추정결과

	OLS	SAR	SEM	SAC
constant	6277.81**	3840.63	4576.35*	4149.21***
주택면적	556.91 ***	567.57***	551.10***	560.54***
경과년수	-1996.12***	-2196.95***	-1804.72***	-1901.22***
경과년수^2	70.66 **	86.12***	69.66**	74.13***
경과년수^3	-0.59	-0.87	-0.69	-0.74
역세권	1544.38**	1416.65**	1719.46***	1681.38***
재건축	6262.35***	4949.77***	6007.44***	5902.08***
총세대수	3.17***	3.29***	3.07***	3.14***
ρ		0.20***		0.05
λ			0.41***	0.35***
R^2	0.7100	0.7150	0.7718	0.7676
sigma^2	6411967.16	5140026.09	4597152.66	4681111.36
Log Likelihood	-828.79	-792.68	-790.55	-790.35

주) *** p<0.01 ** p<0.05 * p<0.10

지면상 관계로 [표 3]의 추정결과 계수를 도무 설명할 수 는 없다. 그러나 해석방법은 유사하므로 OLS 모형과 SEM 모형의 선택된 몇 가지 계수만을 설명하고자 한다. 우선 OLS의 경우 주택면적이 1평 증가함에 따라 평당 약 557만원 가량이 증가하는 것으로 나타났고, 역세권의 경우 비역세권에 비해 1,544만원이 높은 것으로 나타났으며, 재건축은 비재건축 단지에 비해 6,262만원이 높은 것으로 추정되었다. SEM의 경우 주택면적이 1평 증가함에 따라 평당 약 551만원 가량이 증가하는 것으로 나타났고, 역세권의 경우 비역세권에 비해 1,719만원이 높은 것으로 나타났으며, 재건축은 비재건축 단지에 비해 6,007만원이 높은 것으로 추정되었다. 그러나 엄격하게 말해 역세권 및 재건축 변수는 가변수(dummy)가 아니기 때문에 정확한 해석이라고 할 수 없다. 다만 여기서는 해석의 편의를 위해 가변수화 하여 대략적으로 설명하고자 한다. 그리고 평균 경과년수에 대한 기울기의 변화률[10]은 OLS의 경우 -1823.8만원으로 연간 평균적으로 1824만원 SEM의 경우 -1663.9만원으로 연간 평균적으로 1664만원의 감가상각이 일어나는 것으로 추정되었다. 전반적으로 OLS모형

10) y=$\beta_1 Age^3 + \beta_2 Age^2 + \beta_3 Age$의 평균 기울기의 변화률은 $\frac{dy}{dAge}$=$3\beta_1 Age^2 + 2\beta_2 Age + \beta_3$

이 SEM모형과 비교해서 정(positive)의 공간자기상관 관계로 말미암아 과대추정 됨을 알 수 있다.

또한 SAR 및 SEM 모형에서, 공간영향력 및 공간자기상관을 나타내는ρ, λ 계수들이 통계적으로 유의미하게 나타나고 있다. 그러나 SAC 모형의 경우 λ만이 통계적으로 유의미하게 나타나고 있어 이는 SEM 모형이 SAR모형에 비해 다양한 simulation을 통해 설명력이 높음을 시사 한다고 할 수 있다(다음의 주석 12)을 참고).

본 연구에서는 전통적 헤도닉모형과 공간계량모형의 추정방법의 차이에 의해 기존 연구에서 일반적으로 사용되는 결정계수(R^2)를 참고로 하여 Sigma Square (σ^2) 및 Log Likelihood를 기준으로 각 모형별 적합도를 평가하였다.

결정계수(R^2)는 회귀식의 적합도를 평가함에 있어서 가장 많이 이용되는 것으로 독립변수가 종속변수를 설명할 수 있는 정도를 나타내는 것으로, 이런 의미에서 결정계수를 설명력이라고 부르기도 한다. 결정계수(R^2)는 그 값이 클수록 설명력이 증대하고, 회귀식의 활용가치가 높아지고 분석결과는 신뢰성을 더욱 인정받게 된다.

다음으로, 잔차제곱의합(ESS: error sum of squares), Sigma Square(σ^2)값을 비교하였다.

$$\text{ESS} = \sum (Y - \hat{Y})^2 \quad (13)$$

$$\text{Sigma Square} = ESS/(n-k) \quad (14)$$

여기서 Y는 실거래가, $\overline{Y}$은 Y의 평균값, $\hat{Y}$의 추정치, n은 표본의 개수, k는 추정계수의 개수 이다.

Likelihood 값은 관측된 표본 값들을 얻을 수 있는 가능성을 나타내는 계수를 말하며 보통 L로 표시하고, 수식은 아래와 같다.

$$L(y \ \rho, \sigma^2) = \frac{1}{(2\pi^2)^{(2/n)}} \ I_n - \rho W \quad (15)$$
$$\exp \ - \frac{1}{2\sigma^2}(y - \rho Wy)'(y - \rho Wy)$$

일반적으로 L자체보다는 L에 log값을 취해 극대화하곤 한다. 이는 log를 취해도 분석결과는 변하지 않을 뿐만 아니라 그 계산과정이 쉬워지기 때문이다. Log Likelihood 함수는 다음과 같다.

$$\ln L \propto -\frac{n}{2}\ln(y - \rho Wy)'(y - \rho Wy) \quad (16)$$
$$+ \ln \ I_n - \rho W$$

공간적으로 시차된 종속변수나 오차항이 포함된 경우 편의(bias)로 인해 OLS로 추정되어 질수 없고, 대신에 최우추정량(maximum likelihood)이 사용된다. 따라서 OLS의 적합도를 나타내는 R^2은 더 이상 공간모형과 비교될 수 없으며, Log Likelihood와 같은 우도함수가 공간모형의 적합도를 측정하는데 사용되어 진다. 여기서 가장 높은 Log Likelihood를 가진 모형이 더 좋은 모형으로 고려되어진다.

앞서의 [표 3]에서 볼 수 있듯이 보통최소자승법(OLS: Ordinary Least Squares)을 이용한 기존 헤도닉모형보다는 공간자기상관을 고려한 공간계량모형들이 전반적으로 우세하고 그 중에서도 공간오차모형(SEM: Spatial Errors Models)이 가장 적은 Sigma Square (σ^2)을 일반공간모형(SAC: General Spatial Models)이 가장 높은 Log Likelihood를 보이고 있어 OLS 및 SAR 모형과 비교해 더 우세한 것으로 나타나고 있다.[11]

특히, 공간오차모형(SEM)과 일반공간모형(SAC)의 비교 관점에서는 앞서의 Wald(W), Likelihood Ratio(LR), Lagrange Multiplier(LM)의 결과 및 SAC모형에서의 비유의한 ρ를 감안할 때 SEM 모형이 가장 우수한 모형으로 판단 할 수 있겠다.[12]

11) 물론 이러한 결과들은 소위 공간가중행렬(spatial weight matrix)의 방법에 따라 다른 결과가 나올 수 있다.

12) 이론적으로, Anselin(1988) 제시한 데로 Lagrange Multiplier test를 통하여 lag나 error모형을 선택하여 구분하는 것은 널리 알려진 사실이다. 그러나 대부분의 공간자기상관모형의 추정에 있어서, Robin Dubin(2009)에서도 언급하였듯이 ρ 및 λ의 유의성이 모두 비유의하면(insignificant) OLS 모형이, ρ 유의도 가 더 크면 SAR 모형이, λ의 유의도가 더 크면 SEM 모형이 설명력이 높음을 보여주고 있다. 이와 동시에 ρ 및 λ의 유의성이 모두 유의하면(significant) SAC 모형이 monte carlo simulation을 통해 다른 모형에 비해 설명력이 높음을 잘 보여주고 있다[24].

본 연구를 통한 공간계량모형의 적용으로 주택정책 수립에 필요한 몇 가지 시사점을 제시하면 다음과 같다.

우선 기존의 연구에서 많이 이용된 OLS방식의 주택가격의 추정보다 공간계량모형이 보다 우수한 것으로 나타났다. 우리나라에서 주택이 차지하는 자산의 비중과 국민경제에 미치는 영향을 감안할 때 OLS방법에 의한 추정보다는 다양한 공간계량모형들의 기법을 적용하여, 주택정책의 공간적 파급효과를 보다 자세하게 추정할 수 있을 것으로 판단되며 특히 국지적 지역 주택시장의 경우 그 적용성이 높다고 사료된다.

둘째 공간계량모형의 적용으로 자료간의 영향범위를 분석함으로써 부산 아파트 주택시장에 공간자기상관과 이분산성으로 지역적 하위시장이 존재함을 유추할 수 있다. 따라서 공간자기상관이 반영된 지역 하위시장의 범위를 규정함으로써 정책실효성을 높이는 계기가 될 수도 있을 것으로 기대한다.

공간계량모형의 경우 기존의 전통적인 OLS모형에 비해 우수한 예측력을 보이지만 대단위 공간가중행렬의 구성 및 최우도법(maximum likelihood)을 이용한 계산상의 복잡성 등을 내포하고 있다. 따라서 주택시장을 동질적인 하부시장으로 정확하게 세분화가 가능하다면, OLS모형이 공간계량모형에 비해 추정의 편리함은 물론이거니와 이론적으로도 보다 더 선호될 수도 있을 것이다.

마지막으로 헤도닉 모형의 지나친 단순화(under specified) 및 동별 평균 자료를 이용한 정보의 손실(lost information) 등은 본 논문의 한계점으로 판단된다.

참 고 문 헌

[1] L. Anselin, Spatial Econometrics: Method and Models. Dordrecht: Kluwer Academic Publishers, 1988.

[2] J. P. LeSage, Spatial Econometrics, http://www.spatial-econometrics.com/, 1998.

[3] R. K. Pace, "Performing Large Spatial Regressions and Autoregressions," Economics Letters, Vol.54, pp.283-291, 1997.

[4] 이성우, 윤성도, 박지영, 민성희, *공간계량모형응용*, 서울: 박영사, 2006.

[5] S. Rosen, "Hedonic Prices and Implicit Markets : Product Differentiation in Pure Competition," Journal of Political Economy, Vol.82, No.1, pp.35-55, 1974.

[6] 임재현, "주택특성가격이론의 발전 모색", 한국행정학보, Vol.32, No.1. pp.247-261, 1998.

[7] 김성우, 정건섭, "부산 아파트 실거래가를 이용한 전통적 헤도닉 모형과 공간계량모형간의 적합도 관한 비교 연구", 부동산학연구, Vol.16, No.3 pp.41-55, 2010c.

[8] 김성우, 정건섭, "견고한 공간계량경제모형의 추정-층수를 고려한 3차원 공간가중행렬을 이용하여", 주택연구, Vol.18, No.3, pp.73-92, 2010a.

[9] 박헌수, 정수연, 노태욱, "공간계량경제모형을 이용한 아파트가격과 공간효과분석", 국토계획, Vol.38, No.5, pp.115-125, 2003.

[10] 국토연구원, 공간분석기법, 한울아카데미: 438, 2004.

[11] 김성우, 정건섭, "주택정책을 위한 헤도닉모형 평가에 관한 연구: 공간계량모형을 중심으로", 정책분석평가학회보, Vol.20, No.3, pp.115-134, 2010b.

[12] 최명섭, 김의준, 박정욱, "공간종속성을 고려한 서울시 아파트 가격의 공간 영향력", 지역연구 Vol.19, No.3, pp.61-80, 2003.

[13] R. A. Dubin, "Spatial Autocorrelation: A Primer," Journal of Housing Research, Vol.7, pp.304-327, 1998.

[14] 박기호, " 근린가중행렬이 공간적 자기상관 추정에 미치는 형향 : 서울시를 사례로", 서울도시연구, Vol.5, No.3, pp.67-83, 2004.

[15] 김종원, "주택시장에서 공간자가상관의 검증 및 회귀계수의 추정", 경제학연구, Vol.48, No.2, pp.155-173, 2000.

[16] 서경천, 이성호, "공간적 자기회귀모델과 토지시

장분할에 의한 효율적 지가추정에 관한 연구", 국토계획, Vol.36, No.4, pp.77-94, 2001.

[17] 허윤경, "도시별 주택가격의 공간적 영향력 검증: 서울과 부산의 아파트 가격을 중심으로", 주택연구, Vol.15, No.4, pp.5-23, 2007.

[18] A. Can, "Specification and Estimation of Hedonic Housing Price Model," Regional Science and Urban Economics, Vol.22, No.3, pp.453-473, 1992.

[19] 이현석, 박성균, "공간자기상관을 고려한 권역별 등급별 오피스 임대료 결정요인분석", 국토계획, Vol.45, No.2, pp.165-177, 2010.

[20] 김성우, 정건섭, "공간계량모형에서의 실제거리를 반영한 공간가중행렬에 관한 연구-부산아파트 실거래가를 중심으로", 주택연구, Vol.18, No.4, pp.59-80, 2010d.

[21] W. Tobler, "A Computer Movie Simulating Urban Growth in The Detroit Region," Economic Geography, Vol.46, pp.234-240, 1970.

[22] 김광구, "공간자기상관의 탐색과 공간회귀분석의 활용", 정책분석평가학회보, Vol.13, No.1, pp.273-293, 2003.

[23] 국토해양부, http://rt.mltm.go.kr/

[24] R. A. Dubin, "Spatial lag and spatial revisited: some monte carlo evidence," Edited by Lesage and Pace. Advances in Econometrics, 2009, Vol.18, pp.75-98, 2009.

저 자 소 개

정 건 섭(Kyoun-Sup Chung) 정회원

- 1994년 5월 : The University of Texas at Dallas(정치경제학박사)
- 2002년 3월 ~ 2007년 8월 : 한서대학교 행정학과 부교수
- 2007년 9월 ~ 현재 : 부경대 행정학과 부교수
- 2008년 9월 ~ 현재 : 부경대 공공정책연구소장
- 2010년 1월 ~ 현재 : 한국비교정부학회 편집위원장

<관심분야> : 공간계량분석, 토지/주택정책, G.I.S. 및 행정계량분석