

부동산 하부시장 구획: 헤도닉 모형의 개선을 중심으로

이창로* · 엄영섭** · 박기호***

Submarket Identification in Property Markets: Focusing on a Hedonic Price Model Improvement

Chang Ro Lee* · Young Seob Eum** · Key Ho Park***

요약 : 부동산 가격을 추정하기 위한 헤도닉 모형(hedonic model)의 적용에서 가장 중요한 사안은 모형의 정확한 구성과 하부시장의 구획이라 할 수 있다. 모형의 구성에 대해서는 비교적 활발한 개선 노력이 있었으나 하부시장 구획은 상대적으로 큰 관심을 받지 못하였다. 그러나 부동산 가격형성 과정의 공간적 범위 파악이 선행되지 않으면 헤도닉 모형의 적용 결과는 그 정확성이 저하될 수밖에 없다. 본 연구는 헤도닉 모형의 성능 개선에 초점을 두고, 서울시 25개 자치구 중 상대적으로 이질적인 부동산 집단으로 구성된 강남구와 비교적 균일한 부동산 집단으로 이루어진 중랑구를 사례지역으로 하여 하부시장 구획을 시도하였다. 먼저 하부시장 구획을 위한 투입변수로 혼합 GWR(Mixed GWR) 모형에서 산출된 가변 회귀계수(variable coefficients)를 사용하였다. 헤도닉 모형의 회귀계수는 부동산을 구성하는 속성항목(attributes)의 잠재가격(shadow price)으로 해석할 수 있기 때문이다. 다음으로 공간적으로 연결된 하부시장을 구획하기 위해 최소신장트리(minimum spanning tree)에 기반한 SKATER 알고리즘을 사례지역에 적용하였다. 마지막으로 다수준 모형(multi-level model)을 적용하여 구획된 하부시장 결과의 적정성을 검토하였다. 검토 결과, 중랑구는 하부시장이 존재하지 않음을, 강남구는 간선도로를 중심으로 한 5개의 하부시장으로 구분하는 것이 합리적임을 확인하였다. 간선도로와 같은 도시의 인프라는 하부시장 구획에 있어 지금까지 큰 주목을 받지 못한 변수였으나 본 연구를 통해 그 중요성이 실증적으로 확인되었다.

주요어 : 헤도닉 모형, 하부시장, 혼합 GWR(Mixed GWR) 모형, SKATER 알고리즘, 다수준 모형

Abstract : Two important issues in hedonic model are to specify accurate model and delineate submarkets. While the former has experienced much improvement over recent decades, the latter has received relatively little attention. However, the accuracy of estimates from hedonic model will be necessarily reduced when the analysis does not adequately address market segmentation which can capture the spatial scale of price formation process in real estate. Placing emphasis on improvement of performance in hedonic model, this paper tried to segment real estate markets in Gangnam-gu and Jungrang-gu, which correspond to most heterogeneous and homogeneous ones respectively in 25 autonomous districts of Seoul. First, we calculated variable coefficients from mixed geographically weighted regression model (mixed GWR model) as input for clustering, since the coefficient from hedonic model can be interpreted as shadow price of attributes constituting real estate. After that, we developed a spatially constrained data-driven

* 서울대학교 지리학과 박사과정(Ph.D Student, Department of Geography, Seoul National University), k25125@kab.co.kr

** 서울대학교 지리학과 석사과정(Student in the Master's course, Department of Geography, Seoul National University), eys257@gmail.com

*** 서울대학교 지리학과 교수 및 국토문제연구소 겸무 연구원(Professor, Department of Geography, Seoul National University, and Researcher, Institute for Korean Regional Studies), khp@snu.ac.kr

methodology to preserve spatial contiguity by utilizing the SKATER algorithm based on a minimum spanning tree. Finally, the performance of this method was verified by applying a multi-level model. We concluded that submarket does not exist in Jungnang-gu and five submarkets centered on arterial roads would be reasonable in Gangnam-gu. Urban infrastructure such as arterial roads has not been considered an important factor for delineating submarkets until now, but it was found empirically that they play a key role in market segmentation.

Key Words : hedonic model, submarket, mixed GWR, SKATER algorithm, multi-level model

1. 서론

부동산 가격의 정확한 추정은 금융기관의 담보 대출, 건설업자의 부동산 개발, 정부의 과표 산정이나 공시지가 결정 등 다양한 분야에서 그 수요가 증대되고 있다. 부동산 가격 추정에 일반적으로 활용되는 헤도닉 모형(hedonic model)의 적용에서 가장 중요한 사안은 모형의 정확한 구성과 하부시장의 구획이라 할 수 있다(Wheeler *et al.*, 2013). 모형의 정확한 구성에 대해서는 통상회귀모형(Ordinary Least Squares; OLS)의 한계를 극복하기 위한 다양한 모형들(공간회귀모형, 다수준 회귀모형, 베이지안 위계모형 등)이 개발 및 활용되는 등 활발한 연구 실적이 있었으나 하부시장의 구획에 대해서는 상대적으로 관심이 적은 것이 일반적 경향이었다. 그러나 하부시장이 존재함에도 이를 간과하거나 하부시장 구획이 부정확한 경우 헤도닉 모형 적용 결과의 정확성은 저하될 수밖에 없다.

지금까지 하부시장 구획은 주어진 행정구역을 그대로 사용하거나 인근지역의 질을 나타내는 몇몇 변수(주택의 평균시세, 주민소득, 도심과의 거리 등)를 이용하여 하위 행정구역을 몇 개의 커다란 상위 행정구역으로 묶는 방법(원제무 외, 2009; Watkins, 2002), 또는 부동산 관련 전문가가 사전에 설정한 하부시장을 받아들이는 방법(Bourassa *et al.*, 2007) 등이 주로 사용되었다.

헤도닉 모형 적용에 있어 가장 중요한 사안 중의 하나임에도 불구하고 하부시장 구획에 대한 활발한 연구가 진행되지 않은 이유는 다음 두 가지 어려움 때문

으로 풀이된다. 첫째는 명확한 경계를 가지는 하부시장 개념에서부터 그러한 경계는 사실상 존재하지 않는다는 견해(Paez *et al.*, 2008)에까지 하부시장 정의 자체에 대한 일치된 견해가 존재하지 않는다. 둘째는 하부시장 구획을 위한 방법론이 매우 다양하게 제시되었지만, 어떠한 방법이 가장 효율적인 접근인지에 대해 현재까지 논의가 진행 중이다. 즉, 정의와 방법론 모두에 대해 합의된 바가 없어 연구실적이 미진한 편이며, 뒤집어 해석하면 앞으로 추가 연구가 필요한 부분이기도 하다.

여전히 모호하기는 하지만 광의의 개념으로 ‘상호간에 대체가능한 부동산 집단 밀집지역’을 하부시장으로 정의할 수 있으며(Islam & Asami, 2009), 특정 하부시장에 소재하는 부동산 집단은 다른 하부시장에 소재하는 부동산 집단과의 대체가 상대적으로 어려운 특징을 가지고 있다. 이러한 하부시장 구획에서 ‘One Model For Multi-use’라는 개념은 성립하기 어렵다. 연구 맥락과 모형의 사용 목적에 따라 하부시장은 달리 구획될 수밖에 없으며, 또 그렇게 해야만 유용하게 활용될 수 있다(Bourassa *et al.*, 2003).

본 연구의 목적은 부동산 가격의 정확한 추정에 있으므로 헤도닉 모형의 성능이 가장 우수하게 나타날 수 있는 측면에 초점을 두고 하부시장 구획을 시도하였다. 부동산 하부시장은 활용 목적에 따라 특정 지역을 최대한 세부적으로 분류할 수도 있고, 되도록 적은 수(2~3개)로 묶을 수도 있다. 헤도닉 모형의 경우, 시장을 세부적으로 분류할수록 해당 하부시장에 존재하는 표본 수의 부족으로 가격 예측력이 약화되는 반면, 하부시장 수가 적어질수록 지역별 이질성이 평균화되어 가격 예측력이 약화될 수 있다. 따라서 본 연

구에서는 가장 정확한 가격 추정이 가능한 최적의 하부시장 수를 찾고 구획하는 데 중점을 두었다.

본 연구는 정부가 매년 발표하는 공시지가(公示地價) 중 부동산 가격 전문가인 감정평가사가 현장조사 및 정밀평가를 통해 가격을 결정하는 표준지(標準地) 공시지가 자료를 활용하여 부동산 하부시장을 구획하였다. 본 연구는 비교적 좁은 공간 스케일에 해당되는 서울시 두 개 자치구(강남구, 중랑구)를 분석지역으로 선정하였다. 시군구 단위는 공시지가 대량평가 모형(‘비준표’ 모형)의 작성 단위이자 부동산 평가 실무에 있어 분석 가능한 최대의 시장 범위로 인식되고 있다. 따라서 본 연구 결과는 시군구 내부에 이질적인 하부시장이 존재하는지 여부, 존재한다면 어떠한 구획 방법이 가장 합리적인지에 대한 단초를 제공할 수 있을 것으로 예상되며, 나아가 공시지가의 정확성 제고 및 감정평가 결과의 신뢰성 향상에 일조할 것으로 기대된다.

2. 선행연구 고찰

1) 하부시장 구획을 위한 방법론

여러 문헌에서 제시된 하부시장 구획 방법은 크게 다음의 다섯 가지로 분류 가능하다(그러나 아래의 분류가 모든 방법론을 망라한 것은 아니다).

① 헤도닉 모형(Hedonic model)

공간적 연접성 같은 지리적 위치 관계를 명시적으로 고려하지 못하는 한계가 있지만, 해석 및 측정이 용이하여 헤도닉 모형에 기반한 하부시장 구획은 가장 널리 사용되는 방법이다(Islam & Asami, 2009). Goodman(1978)이 제시한 방법은 현재까지 폭넓게 활용되고 있는데, 헤도닉 모형에서 산출된 계수는 부동산을 구성하는 속성항목(attributes)의 잠재가격(shadow price)으로 해석할 수 있기 때문에 이러한 헤도닉 가격계수를 후속 모형에 투입하여 하부시장을 구획하는 방법이다. 가격계수 뿐 아니라 헤도닉 모형

에서 산출된 잔차를 기준으로 분석한 사례도 흔히 찾아볼 수 있다. Tu *et al.*(2007)은 헤도닉 모형에서 산출된 잔차를 기준으로 잔차들 간의 상관 정도를 군집 분석의 투입변수로 이용하여 주택 하부시장을 구획하였다. 헤도닉 가격계수 또는 잔차가 하부시장별로 상이할 경우, 구획된 하부시장 결과는 적정한 것으로 해석할 수 있다.

② 다변량 분석기법(Multivariate analysis)

주성분 분석(principal component analysis), 요인분석(factor analysis), 다차원척도분석법(multidimensional scaling) 등도 하부시장 구획의 한 축을 구성한다. 이 방법은 대부분 최초의 설명변수를 대상으로 주요한 요인들(factors)을 추출한 후, 이 요인들을 분석하여 하부시장을 구획하는 것이다. Bourassa *et al.*(1999)은 최초 설명변수에서 주성분분석을 통해 몇 개의 요인들을 추출한 후, 이 요인들을 군집분석의 투입변수로 활용하여 하부시장을 구획한 바 있다.

③ 이산선택모형(Discrete choice model)

주택을 선택할 때 세 가지 주요한 선택은 거주 형태(소유, 임대 등), 주택 유형(규모, 신축연도 등 구조적 속성) 및 거주 지역으로 요약할 수 있으며 이산선택모형 연구는 거주 형태의 선택에 관한 것이 다수를 차지하고 있다(Yates & Mackay, 2006). 주택의 하부시장 구분과 관련된 연구는 주택 유형 및 거주 지역의 선택에 관한 것으로 볼 수 있으며, 이러한 연구에서는 우선 주택 유형 및 거주 지역을 선택하고(하부시장의 구분), 이렇게 구분된 하부시장 내에서 특정 개별주택을 선택하게 된다는 2단계 의사결정 과정을 모형화한다. 따라서 하부시장의 구분은 주택 유형과 관련된 변수(단독·다가구 등 주택의 종류, 면적, 건물구조 등) 및 거주 지역과 관련된 변수(도심과의 거리, 학군, 지역의 소득수준 등)를 기준으로 이루어지게 된다. 주택 선택과 관련된 다양한 대안 중 하나를 최종 선택하게 되므로 자연스럽게 다항로짓모형(multinomial logit model)을 자주 활용하며, 여러 대안 사이의 상관관계를 고려한 위계로짓모형(nested multinomial logit model) 등도 흔히 이용된다. 이산선택모형을 적

용한 해외 연구의 경우 국내 부동산 시장과 달리, 다른 무엇보다 특정 민족이나 계층의 밀집경향이 하부 시장 구분에 있어 가장 중요한 요인인 것으로 재차 확인되고 있다(Bayer *et al.*, 2003).

④ 공간통계기법(Spatial statistics method)

공간통계와 GIS의 활용은 비교적 새로운 경향에 해당한다. 대체로 이 분야의 분석과정은 3가지 절차로 요약된다. 먼저 베리오그램(variogram), 크리깅(kriging) 등으로 대표되는 지리통계적 방법(geostatistical method)을 이용하여, 하부시장을 구분하지 않은 상태의 잔차, 보다 정확하게 잔차의 공간적 구조를 추정한다. 둘째, 추정한 잔차의 공간적 구조에 따라 개별 주택(또는 필지)들을 군집분석을 통해 그룹화한다. 마지막으로 해당 그룹을(필요한 경우) 보다 상위의 집단으로 통합하여 하부시장을 최종 구획한다. Bourassa *et al.*(2007)은 베리오그램을 이용한 Lattice Model(CAR Model, SAR Model)이 기존의 OLS 모형보다 하부시장 구획 성능이 우수함을 밝힌 바 있다.

⑤ 그 밖의 기법

현재 하부시장 구획에 대한 연구는 기존의 방법을 정교화하려는 노력(예로 헤도닉 모형에서 선형의 함수 형태가 아닌 비선형 함수 형태의 적용, 모형의 오차항을 보다 세분화하여 공간 효과와 관련된 오차항을 추론하고 이를 지도화하는 시도 등)과 비교적 새로운 기법을 시도하려는 노력(신경망 분석(neural network), 자기 조직화 지도(self-organizing map)와 GIS를 병용한 접근, 퍼지 군집분석(fuzzy clustering) 등)으로 나뉘어 진행되고 있다(Islam & Asami, 2009).

본 연구에서는 구득 가능한 설명변수의 종류가 많지 않아 다변량 분석기법을 활용할 실익이 적고, 설명변수도 대부분 부동산의 구조적 특성(structural attributes)을 나타내는 항목(도로접면, 형상 등)이며 인근지역의 질을 나타내는 변수가 거의 없어 이산선택 모형을 효율적으로 적용하기 어렵다. 또한 공간통계 기법의 경우 잔차를 대상으로 분석을 수행하는 경우가 많은데, 모형 설정 오류(model mis-specification) 등으로 인한 효과가 모두 잔차에 반영되는 바, 이러한

잔차를 대상으로 신뢰성 있는 분석을 수행하기 어려울 수 있다. 따라서 본 연구에서는 비교적 이론 정리가 잘 되어 있는 헤도닉 모형을 적용하되, 필지 간 거리 등 공간상의 위치를 명시적으로 고려할 수 있는 방법을 적용하였다.

2) 공간적 연접성을 고려한 하부시장 구획

하부시장들 간의 위치 관계, 즉 공간적 연접성(spatial contiguity)이 하부시장 구획의 필수적 전제조건인지에 대해서는 오래전부터 논쟁이 있어 왔다. 상당수의 연구는 지리적 거리에 기초하여 공간적으로 연접한 지역을 하부시장으로 정의하였으나(Michaels & Smith, 1990; Helbich *et al.*, 2013), 일부 연구는 유사한 부동산 속성을 가지고 있어 잠재 구매자에게 대체 가능한 부동산이 존재하는 지역이면 족할 뿐, 공간적 연접성은 필요하지 않다고 보는 견해도 있다(Rothenberg *et al.*, 1991).

Goodman & Thibodeau(2007)는 공간적 연접성을 제약조건으로 부과한 방법과 부과하지 않은 방법을 비교하였으며, 그 결과 공간적 연접성을 부과하지 않은 방법은 부과한 방법만큼이나 우수함을 제시하였다. 이들은 공간적 연접성을 제약식으로 부과할 경우 계산시간이 기하급수적으로 늘어나며, 이러한 계산 비용을 고려할 때 공간적 연접성을 부과하지 않는 것이 보다 효율적이라고 주장하였다. 그러나 이들 주장의 이면에는 공간적 연접성을 부과하는 것이(계산 비용을 고려하지 않을 경우) 보다 우수한 방법이라는 전제가 깔려 있는 것으로 풀이된다.

또한, 상호 간에 대체 가능한 부동산들은 지리적 위치와 깊은 관련을 맺을 수밖에 없다(Wu & Sharma, 2012). 동일한 지역에 소재하는 부동산들은 다양한 지역요인들을 공유하기 마련이다. 예를 들어 학교와 같은 교육의 질, 사회복지서비스의 수준, 문화·편의 시설의 활용 가능성, 대중교통시설에의 접근성, 더 나아가 주민들이 지역에 대해 가지는 애착심(장소 정체성) 등 동일한 지역에 소재하는 부동산들은 이러한 요인들을 공유하게 된다. 따라서 공간적 연접성은 부동산 하부시장 구획에서 빼놓을 수 없는 요소라고 할

수 있다.

공간적 연결성 조건이 하부시장 구획의 주요한 선결조건임에도 이러한 제약조건을 명시적으로 고려한 사례는 흔치 않다. 이러한 시도가 적었던 이유는 ① 공간적 연결성을 효율적으로 반영할 수 있는 모형 내지 알고리즘 개발이 미진하였고, ② 그러한 모형 등이 존재하더라도 계산량이 방대하여 실무에 쉽게 적용하기 어려웠기 때문인 것으로 풀이된다.

하부시장 구획은 과업의 속성상 군집분석을 자주 활용하게 되는데, 군집분석과정에서 공간적 연결성을 고려한 방법은 크게 세 가지 경향으로 나누어볼 수 있다. 첫째는 군집분석을 두 개의 단계로 나누어 수행하는 것이다. 우선 속성정보를 투입변수로 하여 전통적인 군집분석을 수행한다. 이후 하나의 군집으로 묶였으나 공간적으로 연결하지 않은 개체들은 다시 나누어 서로 다른 군집으로 분화한다. 단순하다는 장점이 있으나 공간적 연결성을 직접적으로 통제하지 못하므로, 자료의 공간적 자기상관성(spatial autocorrelation)이 약할 때 적절할 수 있다(Haining *et al.*, 2000).

두 번째는 속성정보 외에 위치정보를 추가적인 투입변수로 하여 군집분석을 수행하는 것이다. Martin(1998) 등의 활용 사례가 있으나 공간적으로 연속된 군집을 얻기 위해 반복적으로 분석을 수행하여야 할 뿐 아니라, 위치정보를 별도로 고려하기보다는 군집분석을 위한 기존 투입변수에 2개의 변수(예를 들어 UTM 좌표계의 X 좌표값, Y 좌표값)를 추가하는 것에 그쳐 위치정보와 속성정보를 동일한 수준으로 취급하는 한계가 있다.

세 번째는, 군집분석 수행과정에서 공간 개체들 간의 근린관계를 명시적으로 고려하는 방법이다. Openshaw(1977)가 제시한 AZP(Automatic Zoning Procedure)가 대표적인 방법으로, 목적함수(objective function)를 최소화할 수 있도록 n 개의 개체를 k 개의 지역으로 임의분할하되, 공간적 연결성을 제약식으로 부과하는 것이다. 그러나 이 방법은 계산량이 방대하여 중간 크기 이상의 데이터에 대해 장시간이 소요되는 단점이 있어, 데이터 크기가 일정량을 초과할 경우 실무에 적용하기 곤란하다. 이러한

단점을 보완하여 계산의 효율성을 높이기 위한 방법이 SKATER(Spatial 'K'luster Analysis by Tree Edge Removal) 알고리즘이다(Assunção *et al.*, 2006).

본 연구에서는 공간적 연결성을 명시적으로 고려하되, 계산시간의 제약을 거의 받지 않는 SKATER 알고리즘을 이용하여 하부시장 구획을 시도하였다.

3. 사례 분석

1) 사례지역의 선정 및 분석 절차

본 연구에서는 2013년 표준지 공시지가를 검토하여 서울시 25개 자치구 중 지가의 이질성이 가장 큰 지역과 낮은 지역으로 판단되는 강남구와 중랑구를 사례지역으로 선정하였다¹⁾. 표준지 중 자료의 수가 많지 않고 분석에 포함시킬 경우 오히려 분석 결과를 왜곡시킬 가능성이 높은 농경지, 임야 등은 제외하였으며, 그림 1은 두 개 구 자료의 분포 현황을 보여준다. 강남구가 중랑구에 비해 지가의 이질성이 크다는 사실은 그림 2의 베리오그램을 통해서도 쉽게 확인할 수 있다²⁾. 동일한 하부시장 방법론을 적용하였을 때, 지가의 이질성이 큰 지역과 작은 지역의 하부시장 구획 결과를 비교하는 것은 의미가 있을 것으로 판단되어 상기 두 지역을 선정하였다.

이하에서는 2013년 표준지를 기준으로 공시지가 발표시 함께 공개되는 여러 특성(용도지역, 면적, 도로조건 등)을 설명변수로 하여 헤도닉 모형(OLS)을 구성한 후, 지리적 가중회귀모형(Geographically Weighted Regression; GWR)을 통해 하부시장 구획에 필요한 투입변수를 선별하였다. 이후 공간적 연결성을 제약식으로 부과한 군집분석 알고리즘을 적용하여 하부시장을 구획하였다. 마지막으로 구획된 각 하부시장을 제2수준의 변수로, 도로조건과 같은 개별 필지 특성을 제1수준의 변수로 구성한 다수준 모형(Multi-level model)을 적용하여 구분된 하부시장의 적정성을 검토하였다³⁾. 그림 3은 본 연구의 분석 절차를 간략하게 제시한 것이다.

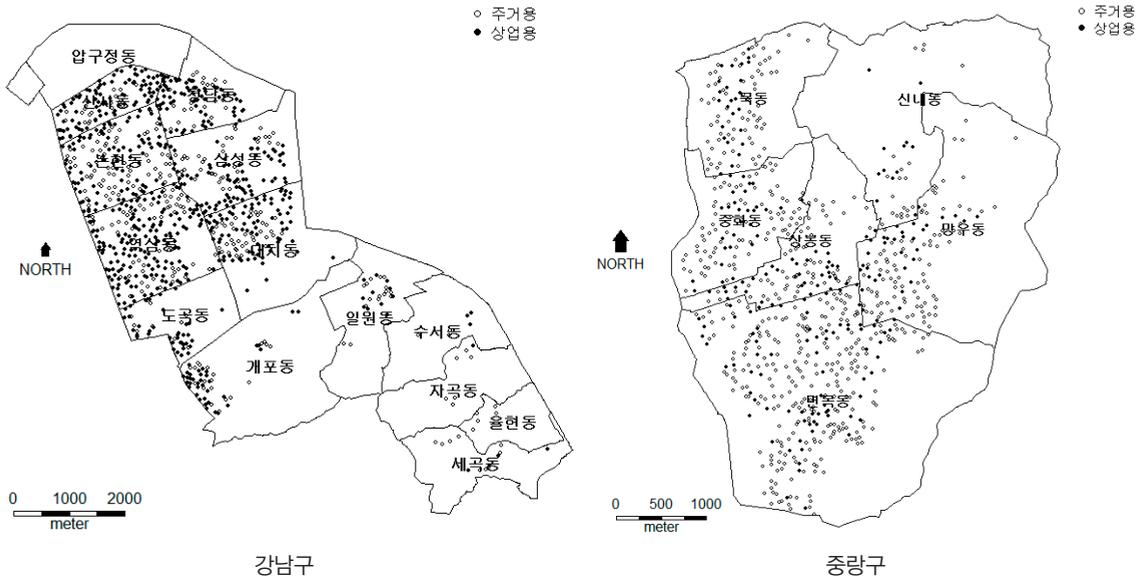


그림 1. 표준지 자료의 분포 현황

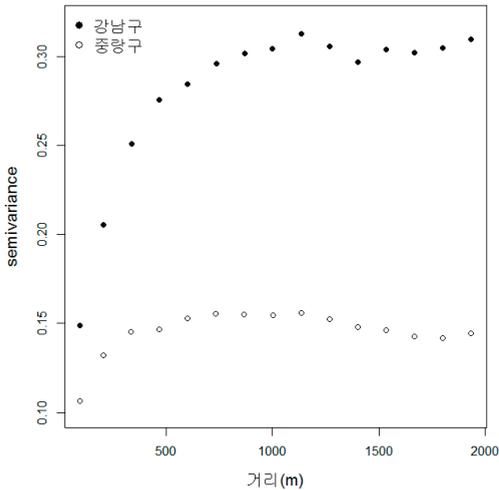


그림 2. 지가를 기준한 베리오그램

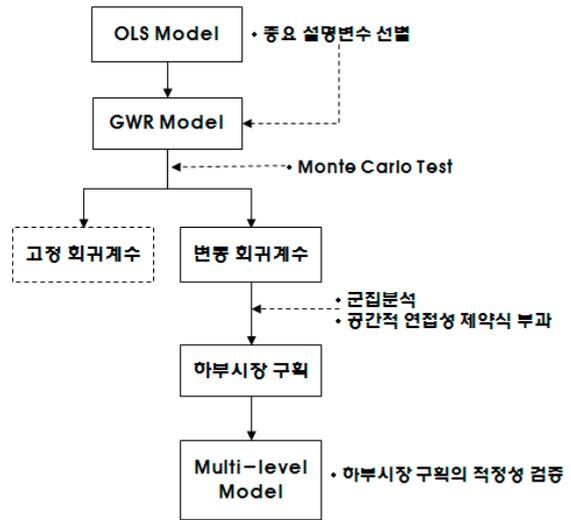


그림 3. 분석 절차

2) 하부시장 구획을 위한 투입변수 선별

본 연구에서는 하부시장 구획을 위한 투입변수로 GWR 모형을 통해 산출된 가변 회귀계수(variable coefficients)를 사용하였다. 이미 많은 선행연구(Goodman, 1978; Helbich *et al.*, 2013)에서 헤도닉 모형의

회귀계수가 해당 부동산의 잠재가격을 나타내는 것으로 해석하여 회귀계수를 이용한 하부시장 구획을 시도하였고, 본 연구에서도 이러한 접근을 따랐다.

GWR 모형을 적용하기 전에, 표준지 공시지가 발표시 함께 공개되는 특성항목들을 중심으로 중요 설명변수들을 검토하였다(OLS 모형 적용⁴⁾). 최초에는

8개 항목(면적, 용도지역, 용도지구, 경사도, 유해시설(철도)과의 거리, 이용상황, 형상, 도로접면)이 설명변수에 포함되었으나 통계적 유의성 등을 고려한 최종 설명변수는 4개로 압축되었다(용도지역, 이용상황, 형상 및 도로접면). GWR 모형 적용 전 OLS 모형을 통해 중요 변수를 선별한 이유는 GWR 모형의 약점으로 자주 지적되는 다중공선성(multi-collinearity)을 완화하고(Wheeler & Tiefelsdorf, 2005), 공개되는 표준지 항목 중 지방에서는 의미가 있으나 서울지역에서는 중요도가 떨어지는 변수(경사도 등)를 제거하기 위함이다. 표 1은 최종 선별된 설명변수, 그리고 각 OLS 모형의 적합도를 보여준다. 용도지역의 경우

기준범주인 2종 일반주거지역 대비 일반상업지역이나 근린상업지역의 회귀계수는 양(+)으로, 자연녹지지역이나 개발제한구역의 회귀계수는 음(-)으로 나타나는 등 모형의 전반적인 적합 결과는 일반적인 직관과 부합하는 것으로 보인다.

GWR 모형은 토지가격을 설명하는 여러 변수 중 공간적 이질성이 강한 변수를 파악하기 위한 매우 효율적인 도구이다(Fotheringham & Rogerson, 2009, p.243-244). GWR 모형은 계수 추정을 위해 전체 데이터의 일부부만 사용하는 일종의 국지 모형(local model)이라 할 수 있다. 다시 말해, moving-window regression과 유사하게 각 관찰치를 중심으로 관찰치

표 1. 중요 설명변수의 선별(OLS 모형 적합)

설명변수		강남구(Adj. R ² = 0.80, n = 960)			중랑구(Adj. R ² = 0.89, n = 789)		
		회귀계수	표준오차	t-value	회귀계수	표준오차	t-value
상수항		14.99	0.14	110.75	14.12	0.04	391.98
용도 지역	1종전주	-0.29	0.04	-6.36			
	1종일주	-0.04	0.04	-1.07	-0.09	0.03	-3.35
	3종일주	0.11	0.02	5.25	0.04	0.01	2.99
	준주거	-0.01	0.06	-0.20	0.24	0.02	14.52
	일반상업	0.49	0.03	16.94	0.37	0.03	12.17
	근린상업				0.48	0.03	14.58
	자연녹지	-1.25	0.15	-8.27			
	개발제한	-0.43	0.11	-3.98	-0.32	0.09	-3.49
이용상황(상업용)		0.48	0.03	18.32	0.44	0.02	18.76
형상(자루형 외)		0.23	0.07	3.22	0.16	0.03	5.00
도로 접면	광로	0.49	0.12	4.06	0.39	0.03	13.12
	중로	0.21	0.12	1.72	0.31	0.03	10.79
	소로	0.08	0.12	0.69	0.22	0.02	10.23
	세로	-0.01	0.12	-0.09	0.11	0.02	6.75

용도지역	1종전용주거지역, 1종일반주거지역, 2종일반주거지역*, 3종일반주거지역, 준주거지역, 일반상업지역, 근린상업지역, 자연녹지지역, 개발제한구역
이용상황	주거용, 상업용
형상	자루형**, 자루형 외
도로접면	광로(폭 25m 이상), 중로(폭 12~25m), 소로(폭 8~12m), 세로(폭 8m 미만 중 자동차 통행 가능), 세로 (폭 8m 미만 중 자동차 통행 불가능)

* 밑줄은 기준범주(reference category)를 의미(가장 빈도가 높거나 해석이 용이한 항목을 기준범주로 설정).

** 출입구가 좁고 기다란 띠 모양으로 이루어진 토지를 '자루형'이라 하며, 건축 등 토지이용에 있어 가장 큰 제약을 받음

인근에 소재하는 데이터만을 고려하여 회귀모형을 적합하고, 이후 관찰치를 옮겨가며 같은 과정을 반복한다. 해당 관찰치와 가까운 관찰치가 모형 적합에 보다 큰 영향을 미치도록 거리조락함수와 유사한 커널함수(kernel function)를 이용하여 각 데이터에 부여할 가중치를 결정한다. 그러나 GWR 모형 결과는 커널함수의 형태보다는 대상 관찰치에서 얼마나 떨어진 관찰치까지 계산에 포함시킬 것인지, 즉 대역폭(bandwidth)에 보다 큰 영향을 받는 것으로 알려져 있다(Fotheringham *et al.*, 2002). 본 연구에서는 거리(km)를 기준한 고정 대역폭(fixed bandwidth)보다는 자료의 불균등 분포를 고려한 가변 대역폭(adaptive bandwidth)을 사용하였다⁵⁾.

GWR 모형은 기본적으로 모든 설명변수의 회귀계수가 공간상에서 가변적임을 가정한다. 그러나 일부 변수는 공간상에서 가변적인 값을 가지며, 다른 변수는 그렇지 않은 경우(즉 기존의 전역 모형(global model)이 적절하다면) 혼합 GWR(Mixed GWR) 모형이 보다 적합하다. 즉, 혼합 GWR 모형은 공간적 이질성이 통계적으로 유의하지 않은 회귀계수는 일정한 값으로 유지시키고, 여타 회귀계수들은 공간상에서 가변적인 값을 갖도록 한다. 혼합 GWR 모형의 산식은 다음과 같이 표현할 수 있다.

$$y_i = \sum_{j=1}^k a_j x_{ij} + \sum_{l=1}^m b_l(u_i, v_i) x_{il} + e_i \quad (1)$$

위 식에서 y_i 는 종속변수(자연로그를 취한 공시지가 단가), a_j 는 공간상에서 일정한 값을 가지는 고정 회귀계수, $b_l(u_i, v_i)$ 은 지리적 위치 (u_i, v_i) 에서 가변적인 값을 가지는 가변 회귀계수, e_i 는 오차항을 나타낸다. 회귀계수가 가지는 공간적 이질성의 통계적 유의성 여부는 Brunson *et al.*(1998)이 제시한 일종의 몬테카를로 검정을 활용하여 판단하였다. 이 검정의 귀무가설 및 대립가설은 다음과 같다.

$$\begin{aligned} H_0 : \beta_{ij} &= \beta_j \forall i \\ H_1 : \beta_{ij} &\text{가 모두 동일하지는 않다 } \forall i \end{aligned} \quad (2)$$

β_j 는 j 번째 설명변수의 회귀계수를 나타내며(고정

회귀계수), β_{ij} 는 j 번째 설명변수의 관찰치 i 에서의 회귀계수를 나타낸다(가변 회귀계수). 회귀계수의 변동성을 측정하는 통계량은 다음과 같다.

$$v_j = \sum_i^N (\beta_{ij} - \beta_j)^2 / N \quad (3)$$

위에서 N 은 자료의 개수, β_j 는 모든 관찰치 i 에 대해 계산한 j 번째 회귀계수의 평균값을 나타낸다. 따라서 v_j 의 값이 작을수록 해당 회귀계수는 고정된 것으로 보는 것이 합리적이다. 본 연구에서는 몬테카를로 시뮬레이션 기법을 이용하여 v_j 의 값을 반복 계산한 후, 실제 관찰된 v_j 값의 순위를 p-value로 해석하였다(Brunson *et al.*, 1998). 예를 들어 100번의 시뮬레이션 결과, 실제 관찰된 v_j 의 값이 4번째로 큰 값이었다면 p-value는 0.04가 된다. 각 회귀계수마다 100번의 시뮬레이션을 실시하여 공간적 변동성의 통계적 유의성을 확인하였으며, 그 결과는 표 2와 같다.

회귀계수의 값이 통계적 유의성이 있을 만큼 공간적 변화가 클 경우, 이는 하부시장의 존재를 의미하는 것으로 해석할 수 있다(Helbich *et al.*, 2013). 표 2를 보면 강남구의 경우에는 1종일반주거지역, 3종일반주거지역, 일반상업지역, 이용상황 및 형상의 회귀계수 변동성이 큰 것으로 나타났고(p-value 0.05 기준, 이하 동일), 중랑구는 3종일반주거지역, 준주거지역, 세로의 회귀계수 변동성이 큰 것으로 확인되었다. 따라서 표 2의 결과에 기초하여 p-value 0.05 이하인 변수는 가변 회귀계수로, 이외의 변수는 고정 회귀계수로 모형을 구성한 후, 혼합 GWR 모형을 적용하였으며 그 결과는 표 3과 같다.

표 3의 고정 회귀계수를 보면, 표 1의 OLS 모형에서 유의성이 없었던 계수를 제외하고는 대부분 부호와 절대크기가 OLS 모형의 것과 일치한다. 반면 가변 회귀계수를 보면, 세부 지역별로 상당한 차이가 있음을 알 수 있다. 예를 들어 강남구의 이용상황 계수(주거용 대비 상업용의 격차)를 보면 상업용이 주거용보다 항상 가격수준이 높게 형성됨을 알 수 있는데, 작게는 1.4배, 크게는 1.8배까지 차이가 나고 있다⁶⁾. 강남구 이용상황 계수값을 연속표면으로 표현한⁷⁾ 그림 4(a)를 보면 '전형적인' 강남 상업지대에 해당하는

북쪽의 신사동, 청담동, 논현동에서 주거용 대비 상업용 토지의 가격 격차가 가장 크게 발생하고 있으며, 주로 아파트 및 다세대주택으로 이루어진 남쪽의 도곡동, 개포동 일대는 가격 격차가 상대적으로 작은 편이다.

또한 강남구의 용도지역 계수 중 1종 일반주거지역 계수를 보면, 대부분 기준범주인 2종 일반주거지역보다 가격 수준이 열세한 것으로 나타났지만, 일부 지역에서는 두 개 용도지역 간 가격 수준 차이가 거의 없거나 근소하지만 1종 일반주거지역의 가격 수준이 보다 높은 곳도 있음을 알 수 있다. 그림 4(b)를 보면 강남구 남서쪽의 도곡동 및 개포동 일대가 그러한 지역에 해당하는데, 이 지역 일대의 1종 일반주거지역은 주로 양재천 도로변을 따라 띠 모양으로 지정되어 있는 반면, 2종 일반주거지역은 양재천 도로변 후면에 사각형 형태로 지정되어 있다. 따라서 도로변에 접하면서 비교적 경관이 우수한 지역이 1종 일반주거지역

표 2. GWR 모형 시뮬레이션(100회) 결과(p-value)

설명변수		강남구	중랑구
상수항		0.88	0.08
용도 지역	1종 전용주거지역	0.09	-
	1종 일반주거지역	0.03	0.43
	3종 일반주거지역	0.02	0.04
	준주거지역	0.47	0.01
	일반상업지역	0.00	0.87
	근린상업지역	-	0.96
	자연녹지지역	0.63	-
	개발제한구역	0.98	0.98
이용상황(상업용)		0.00	0.26
형상(자루형 외)		0.00	0.51
도로 접면	광로	0.28	0.35
	중로	0.41	0.16
	소로	0.70	0.72
	세로	0.60	0.05

표 3. 혼합 GWR 모형 적합 결과

설명변수	강남구					중랑구					
	최소값	1사분위	중위수	3사분위	최대값	최소값	1사분위	중위수	3사분위	최대값	
상수항	14.94					14.12					
용도 지역	1종전주	-0.27									
	1종일주	-0.12	-0.10	-0.09	-0.06	0.04	-0.09				
	3종일주	0.00	0.02	0.03	0.07	0.18	-0.04	0.00	0.03	0.09	0.17
	준주거	0.06					0.16	0.19	0.25	0.30	0.34
	일반상업	0.10	0.26	0.45	0.57	0.75	0.37				
	근린상업						0.48				
	자연녹지	-1.13									
	개발제한	-0.35					-0.31				
이용상황(상업용)	0.35	0.42	0.46	0.52	0.59	0.45					
형상(자루형 외)	0.17	0.25	0.27	0.29	0.32	0.16					
도로 접면	광로	0.59					0.38				
	중로	0.30					0.31				
	소로	0.15					0.22				
	세로	0.05					0.09	0.10	0.11	0.12	0.14
가변대역폭	170개(전체 관찰치의 18%)					83개(전체 관찰치의 11%)					

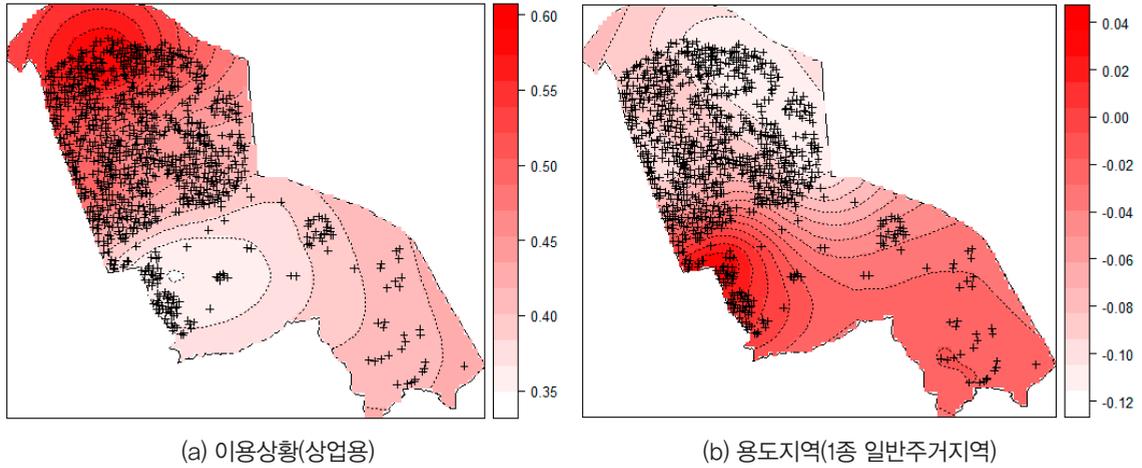


그림 4. 회귀계수 연속표면(강남구)

으로 지정되어 용적률 등 건축법상의 제약에도 불구하고 이 지역의 가격수준은 2종 일반주거지역과 유사하거나 오히려 높게 나타난 것으로 풀이된다.

중랑구의 경우에도 강남구와 유사한 해석이 가능하며, 예를 들어 세로 토지의 경우 기준범주인 세불 토지보다 작게는 1.09배, 크게는 약 1.15배까지 가격 수준이 높게 형성됨을 알 수 있다.

다음 절에서는 상기와 같이 추출된 가변 회귀계수를 투입변수로 하여 하부시장을 구획하고자 한다.

3) 공간적 연접성을 고려한 하부시장 구획

하부시장 구획을 포함한 대부분의 군집분석에서 가장 많이 지적받는 사항은 적절한 군집 수의 결정이라 할 수 있다. 적정 군집의 수를 결정하는 방법은 연구 맥락에 따라 연구자가 주관적으로 설정하는 경향이 일반적이었으나 이러한 주관성을 최소화하기 위해 효율적인 통계적 지표(소위 ‘stopping rule’)를 개발 및 적용하려는 시도 역시 꾸준히 이루어지고 있다. 현재까지 30여 개가 넘는 지표가 제시되었으나 (Legendre & Legendre, 2012, p.389), 일관되게 우월한 성능을 보이는 지표는 없는 것으로 보인다.

본 연구에서 하부시장의 수는 단순하지만 직관

에 부합되는 과정을 통해 결정하였다. 즉, 각 군집의 평균값과 해당 군집 내 개별 관찰치 값들의 차이인 SSW(Sum of Squares Within the clusters)값을 기준으로 정하였으며, 수식은 다음과 같다(Tan *et al.*, 2006, p.539).

$$SSW = \sum_i \sum_{x \in C_i} (x - m_i)^2 \tag{4}$$

위 식에서 x 는 i 번째 군집 C_i 에 속한 관찰값을 나타내며, m_i 는 i 번째 군집 C_i 의 평균값을 의미한다. SSW는 특정 군집 내에서의 평균값과 개별 관찰치 값들의 차이, 즉 비유사성(dissimilarity) 정도를 측정하는 지표로서 값이 낮을수록 군집의 평균값과 개별 관찰치 값 사이의 차이가 적어 동질적인 관찰치들로 구성된 군집이 이루어진 것으로 해석할 수 있다. 그러나 이러한 SSW값은 지표의 특성상 군집이 전혀 이루어지지 않은 최초의 상태에서 그 값이 가장 크고, 군집의 수가 많아질수록 지속적으로 하락할 수밖에 없다. 따라서 본 연구에서는 SKATER 알고리즘 적용과정에서 다양한 군집의 수를 검토하되, SSW 값이 비교적 가파르게 하락하다 이후 완만하게 하락하는 경계지점을 적정 군집의 수로 파악하였다. 그림 5는 이러한 SSW 값의 변화를 보여주고 있으며, 강남구 및 중랑구 모두 5개 정도의 군집이 적절한 것으로 판단된다⁸⁾.

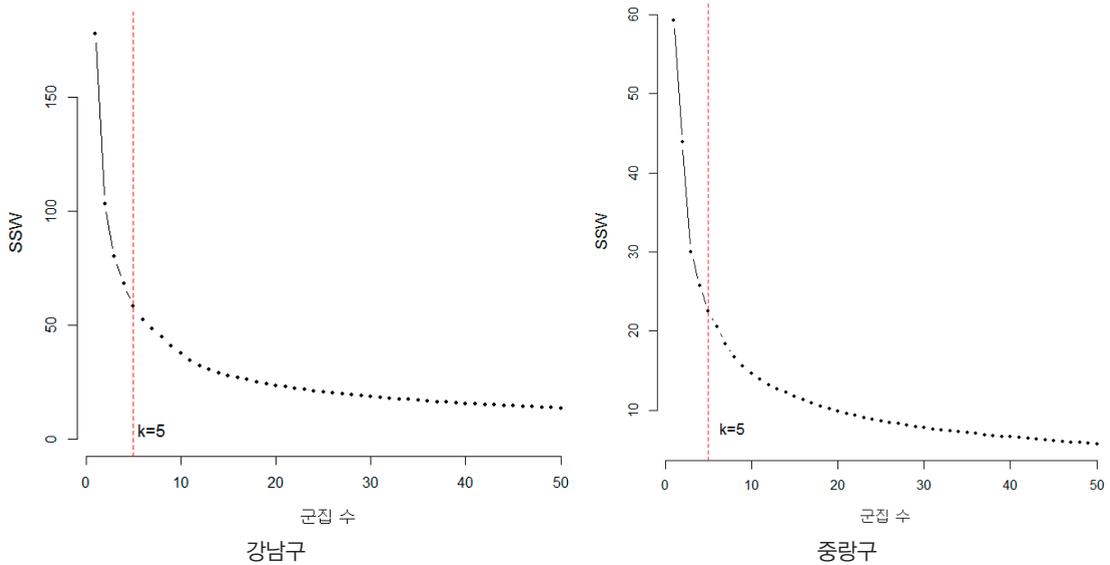


그림 5. SSW값의 변화

본 연구에서는 혼합 GWR 모형으로부터 산출된 가변 회귀계수를 투입변수로 하여 군집분석을 수행하되, 공간적 연결성을 명시적으로 고려할 수 있도록 Assunção *et al.* (2006)가 제시한 SKATER(Spatial 'K'luster Analysis by Tree Edge Removal) 알고리즘을 활용하였다. SKATER 알고리즘은 공간상에 존재하는 개체들을 그룹화하여 연속된 동일지역을 구획하는 방법으로 최소신장트리(minimum spanning tree; MST)의 구성을 기본으로 한다. 최소신장트리(MST)가 동일 군집으로 묶인 개체들의 유사성을 완벽하게 보장할 수 없다는 비판(Guo, 2008)이 있기는 하지만 SKATER는 계산 절차가 상대적으로 효율적이며 중간 내지 대용량 데이터를 처리하기에 적합한 도구이다. 반면 AZP 내지 이를 확장시킨 알고리즘은 계산량이 방대하여 데이터 수가 상대적으로 적은 면 데이터(areal data)의 군집화에 주로 활용되는 경향이 있다. 본 연구에서 사용한 데이터 양(700~1,000개)에 비추어 볼 때 SKATER는 공간적 연결성을 제약식으로 부과할 수 있는 동시에 실무상 적용 가능한 효율적 방법이라고 할 수 있으며, 계산 논리는 다음과 같다(Assunção *et al.*, 2006).

첫째, 개체들 간의 근린관계를 반영한 연결 그래프

(connectivity graph)를 작성한다. 둘째, 연결 그래프상의 정점(edge)이 연결하고 있는 두 개 관찰치의 유사성(similarity)을 측정한다. 본 연구에서는 각 관찰치가 가지는 가변 회귀계수 값을 기준으로 유사성을 측정하는 셈이 된다. 셋째, 측정된 유사성 값에 반비례하도록 정점 간 비용을 계산한다. 본 연구에서는 유클리디안 거리(Euclidean distance)로 비용값을 계산하였다. 넷째, 그래프의 모든 정점을 최소 비용으로 연결하는 최소신장트리(MST)를 구성하여 근린구조를 표현한다. 마지막으로 가장 유사성이 떨어지는 정점부터 순차적으로 끊어 최소신장트리를 분할한다. 정점은 다음의 목적함수가 최대가 되도록 끊는다.

$$f(S_l^T) = SSD_T - (SSD_{Ta} + SSD_{Tb}) \quad (5)$$

위 식에서 S_l^T 는 트리 T 에서 정점 l 을 끊었을 때 형성되는 트리 구조를 나타낸다. Ta 및 Tb 는 정점 l 을 끊음으로써 트리 T 에서 새로이 형성된 보다 작은 개별 트리를 나타낸다. 따라서 정점 l 을 끊음으로써 SSD (Sum of the intracluster Square Deviations) 차이가 가장 크게 발생하도록 최소신장트리를 분할한다. SSD 는 다음과 같은 수식으로 계산한다.

$$SSD_k = \sum_{j=1}^m \sum_{i=1}^{n_k} (x_{ij} - \bar{x}_j)^2 \quad (6)$$

위 식에서 n_k 는 트리 k 에 속한 관찰치들의 수를 의미한다. x_{ij} 는 관찰치 i 의 j 번째 속성값을 나타내며, \bar{x}_j 는 트리 k 에 속한 모든 관찰치들의 속성 j 에 대한 평균값을 의미한다. m 은 분석에서 고려한 속성의 개수를 나타낸다(강남구 5개, 중랑구 3개). 이와 같은 절차를 걸쳐 분할된 결과는 가장 동질적인 개체들끼리 그룹화된 연속된 공간이 된다.

그림 6은 이와 같은 과정을 거쳐 최종적으로 구획된 강남구 및 중랑구 하부시장의 위치 및 형태를 보여준다. 강남구는 북쪽의 신사동에서부터 시작해서 주로 동-서 방향의 간선도로를 따라 지역이 구분된 것으로 보인다. 반면, 중랑구는 간선도로와는 상관없이 분할된 것으로 보이며, 일부는 기존 읍면동 경계와 유사하게 분할되었다.

그러나 이러한 분할은 단지 인위적 결과물의 하나일 수 있으며, 하부시장 구획으로 실제 헤도닉 모형의 성능이 향상되었는지 확인한 연후에야 하부시장별 해석에 의미를 부여할 수 있을 것이다.

4. 하부시장에 대한 해석

본 연구에서는 구획된 하부시장의 유용성을 확인하는 방법으로 헤도닉 모형 성능의 개선 정도를 살펴 보았다. 개선의 정도를 측정하는 지표로는 일반적인 모형 적합 지표인 AIC(Akaike Information Criterion)와 부동산 대량평가모형에서 주로 사용하는 지표인 COD(Coefficient Of Dispersion; 분산계수)를 사용하였다. COD는 아래와 같은 산식을 통해 계산된다⁹⁾.

$$COD = \frac{\left[\frac{\sum |\text{개별비율} - \text{비율들의 중위수}|}{\text{비율들의 개수}} \right]}{\text{비율들의 중위수}} \times 100 \quad (7)$$

식(7)에서 비율은 실제 가격(표준지 공시지가) 대비 모형을 통해 산출된 추정가격(estimated price)의 비율을 말한다. COD가 작을수록 실제 가격과 추정가격의 격차가 작은 것으로 해석할 수 있으며, 통상 20.0 ~ 25.0 정도를 상한선으로 하여 이보다 작은 값을 갖는 경우 과세표준 산정 등 정부 행정업무에서 받아들

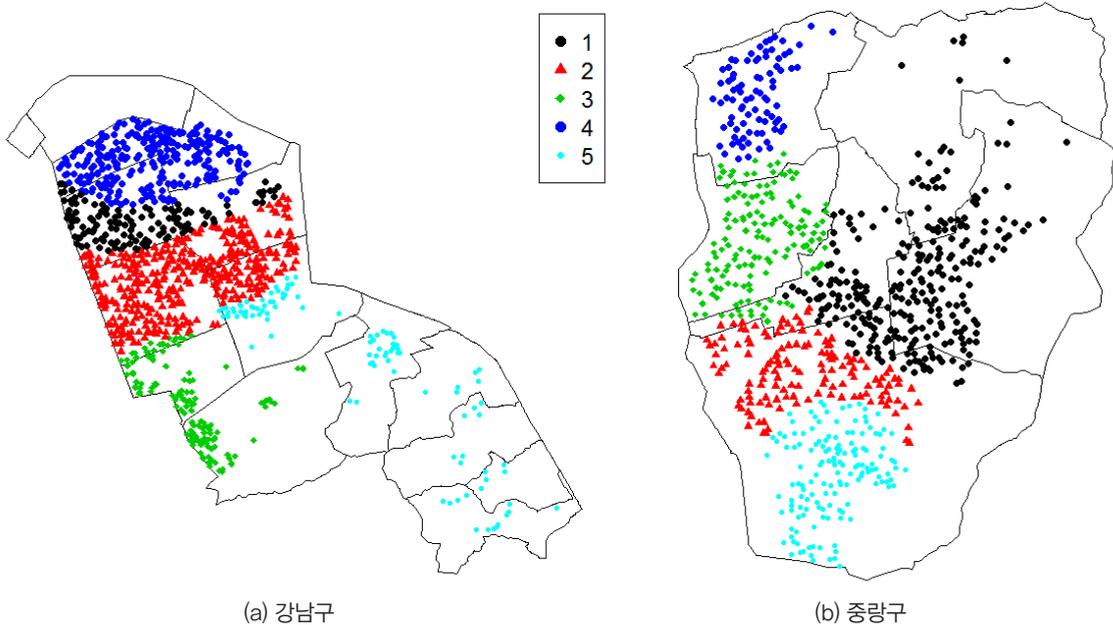


그림 6. 하부시장 구획 결과

일 수 있는 수준으로 간주한다(IAAO, 2010)¹⁰⁾.

모형 성능의 개선 여부는 ① 별도의 하부시장을 구획하지 않고 전체 지역에 대해 하나의 헤도닉 모형을 적용한 경우, ② 읍면동 행정구역을 하부시장으로 보고 모형을 적용한 경우, ③ 본 연구에서 구획한 하부시장 결과를 기준으로 모형을 적용한 경우를 비교하였다. 별도의 하부시장을 구획하지 않은 경우에는 OLS 모형을 적용하였으나, 하부시장을 구획한 경우에는 표준지와 이러한 표준지가 포섭되는 하부시장 간의 관계를 고려하여 다수준 모형을 적용하였다. 구획된 하부시장을 더미변수로 처리하기보다는 하위 수준의 개별 필지와 상위 수준의 하부시장이라는 위계구조를 감안한 다수준 모형을 적용하는 것이 하부시장 간 정확한 비교가 가능하기 때문이다. 비단 하부시장 뿐 아니라 위계구조를 가지는 데이터에서 단일 오차항을 가정하는 것은 부적절할 수 있다. 예를 들어 데이터가 2개 이상의 상이한 공간 수준에서 측정되었다면, 각 공간 수준에 따라 상이한 오차항 분산(error variance)을 가정하는 것이 보다 합리적이기 때문이다(Jones & Bullen, 1993).

다수준 모형의 적용 형태는 해석의 편의를 위해 임의 상수항(random intercept) 모형을 선택하였고, 표 4는 이에 따른 강남구 및 중랑구의 모형 적합 결과를 보여준다.

표 4를 보면 강남구의 경우 AIC 및 COD 값이 모두 하락하였는데, 하부시장 구획을 통해 헤도닉 모형의 성능이 개선되었음을 확인할 수 있다. 특히 읍면동보다는 본 연구에서 구획한 하부시장이 보다 우수한 것을 알 수 있으며, 읍면동의 개수(13개)보다 본 연구에서 구획한 하부시장 개수(5개)가 적다는 사실을 고려할 때, 공간적 연결성을 고려한 SKATER 구획방법이 보다 효율적임을 알 수 있다.

반면, 중랑구의 경우 하부시장 구획에 관계없이 모형의 성능이 비교적 일정함을 알 수 있는데, 이는 중랑구가 비교적 동질적인 부동산 집단으로 구성되어 하부시장 자체가 존재하지 않는 것으로 볼 수 있다. 특히 AIC 값의 경우 하부시장 구획으로 오히려 소폭 증가한 것을 알 수 있다. 또한 중랑구는 전체를 하나로 보아 OLS 모형을 적용한 경우에도 COD가 이미

표 4. 모형 적합 결과

구분		AIC	COD
강남구	전체	128.9	18.87
	읍면동 구획(13개)	58.1	16.98
	SKATER 구획(5개)	-11.9	16.55
중랑구	전체	-994.1	9.50
	읍면동 구획(6개)	-959.7	9.39
	SKATER 구획(5개)	-935.7	9.44

10.0 미만으로 나타나는 등 별도의 하부시장 구분이 필요 없을 정도로 처음부터 모형의 적합도가 매우 우수하게 나타난 지역이다. 따라서 하부시장을 구획하더라도 더 이상 모형 성능의 개선 여지가 거의 없는 지역에 해당한다. 결론적으로 그림 6(b)에서 5개 하부시장 구분은 인위적인 결과물에 지나지 않는 것으로 보인다.

헤도닉 모형 성능의 개선 관점에서 비교적 유의미한 결과가 도출된 강남구의 다수준 모형 적합 결과(SKATER를 활용한 5개 하부시장 구획)는 표 5와 같다.

표 5를 보면 고정효과 부분의 회귀계수는 표 3의 혼합 GWR 모형 적합 결과와 대부분 유사하다(가변 회귀계수의 경우 중위수 기준). 또한 개별 필지 수준(수준 1)의 표준편차 0.23, 하부시장 수준(수준 2)의 표준편차 0.15의 크기를 고려할 때 하부시장은 오차항 변동성의 약 43%(0.15/0.38)를 설명하는 등, 랜덤효과를 모형의 구성요소로 포함시킨 것은 합리적인 조치였음을 알 수 있다.

다수준 모형의 상수항 값 14.88은 기준범주(2종 일반주거지역, 주거용, 세불(폭 8m 미만의 자동차 통행이불가능한 도로))에 해당하는 표준지의 지가수준을 나타낸다. 즉, 강남구 이면골목에 소재하는 전형적인 주택 표준지의 지가수준으로 해석할 수 있다($\exp(14.88)=2,900,000\text{원}/\text{m}^2$). 이러한 주택 표준지의 지가수준은 어떠한 하부시장에 위치하는지에 따라 최대 $3,370,000\text{원}/\text{m}^2(\exp(14.88+0.149))$, 최소 $2,360,000\text{원}/\text{m}^2(\exp(14.88-0.204))$ 의 격차를 보이고 있다. 여타 회귀계수도 이와 유사한 해석이 가능하다.

표 5. 강남구 다수준 모형 적합 결과

설명변수	고정효과(fixed effect)			랜덤효과(random effect)			
	회귀계수	표준오차	t-value	상수항		회귀계수	
상수항	14.88	0.14	108.53	가격수준 중	하부시장 1	-0.001	
용도 지역	1종전주	-0.27	0.04	-6.40	가격수준 상	하부시장 2	0.124
	1종일주	-0.06	0.04	-1.72	가격수준 하	하부시장 3	-0.204
	3종일주	0.09	0.02	4.47	가격수준 상	하부시장 4	0.149
	준주거	0.09	0.05	1.66	가격수준 하	하부시장 5	-0.068
	일반상업	0.43	0.03	16.05	구분		표준편차
	자연녹지	-1.15	0.14	-8.48	수준 1(잔차항)		0.23
	개발제한	-0.34	0.10	-3.51	수준 2(상수항)		0.15
이용상황(상업용)	0.45	0.02	18.94				
형상(자루형 외)	0.27	0.06	4.29				
도로 접면	광로	0.56	0.11				5.22
	중로	0.27	0.11				2.49
	소로	0.14	0.11				1.30
	세로	0.03	0.10				0.29

하부시장을 나타내는 랜덤효과 상수항을 보면 가격수준이 강남구 평균보다 높은 지역은 하부시장 2 및 4에 해당한다. 하부시장 2는 중심 부분을 동-서 방향으로 지하철 2호선(강남역-선릉역-삼성역; 테헤란로)이 통과하는 우리나라의 대표적인 업무중심 지대이다. 반면, 하부시장 4는 강남구 북쪽의 신사동, 압구정동, 청담동을 아우르는 강남의 ‘관문’으로 상가, 갤러리 등이 밀집분포하고 유동인구가 특히 많은 지역에 해당하며, 중심부에 지하철 3호선(신사역; 도산대로)이 통과하고 있다.

강남구의 평균 가격수준에 해당하는 하부시장 1역시 중심부를 동-서 방향으로 지하철 7호선(논현역-강남구청역-청담역; 학동로)이 통과하고 있다. 이 지역은 유동인구, 상권 활성화 정도, 지가수준 등에서 하부시장 2 및 4보다 열세한 지역이다.

마지막으로 강남구 평균 가격수준에 미치지 못하는 하부시장 3 및 5를 보면 하부시장 3은 강남구 남서쪽 도곡동 및 개포동 일대의 아파트 위주로 이루어진 주거지대에 해당한다. 하부시장 5는 강남구 남동쪽의

수서동-세곡동 외곽에 해당하며, 보금자리주택 개발이 진행되고 있는 지역이기도 하다. 하부시장 3과 5는 주로 강남구의 배후 주거지 역할을 하는 지역에 속한다.

이와 같이 본 연구에서 구획한 강남구 하부시장의 특징 중 하나는 강남구를 동-서로 관통하는 주요 간선도로(북쪽에서부터 도산대로, 학동로, 테헤란로)를 중심으로 지역이 구분된 점이다. 따라서 이러한 간선도로의 위치나 형태를 하부시장 구획에 있어 중요한 고려요인 중 하나로 고려할 필요가 있다.

간선도로, 지하철 또는 경전철과 같은 도시의 인프라라는 부동산 가격형성에 미치는 중요성에도 불구하고 이를 하부시장 구획에 적극적으로 활용한 연구는 많지 않다. 즉, 하부시장 구획이 아닌 이후의 모형 적합과정에서 간선도로나 지하철과의 거리를 설명변수의 하나로 채택하여 모형의 구성요소로 사용하는 데 그친 경우가 대부분이다(Gatzlaff & Smith, 1993 등). 현행 공시지가 제도의 경우에도 마찬가지로여서 표준지와 간선도로와의 거리를 비준표 모형 구성요소의

하나로 취급하고 있을 뿐, 간선도로의 위치 등을 비준표 모형의 작성단위 결정에 사용하지는 않고 있다.

공간현상을 설명하는 데 있어 기존의 단핵도시(monocentric city)이론 대신 다핵도시(polycentric city)이론이 보다 폭넓게 수용되는 이유의 하나는 도시의 인프라 구축에서 찾을 수 있으며(Haider and Miller, 2000), 부동산의 가격형성과정 역시 주요한 인프라를 중심으로 가격이 차별적으로 형성될 가능성이 높다.

따라서 향후 하부시장과 관련된 다양한 업무(상권의 구분, 공시지가 산정을 위한 유사가격권의 구분, 거래사례 수집을 위한 인근지역 범위의 설정 등) 수행에 있어 간선도로의 위치 및 간선도로를 따라 형성된 상가지대의 특징(고밀도 상가지대, 역주변 상가지대, 후면 상가지대 등)을 중요한 요인으로 고려할 필요가 있다.

5. 결론

본 연구는 부동산 가격추정에 흔히 사용되는 헤도닉 모형의 성능 개선에 주된 목적을 두고, 가장 효율적인 하부시장 구획 방법에 대해 검토하였다. 서울시 25개 자치구 중 공시지가의 평균, 표준편차, 변동계수 등을 고려하여 상대적으로 이질적인 부동산 집단으로 구성된 강남구와 비교적 균일한 부동산으로 구성된 중랑구를 사례지역으로 정하였다. 헤도닉 모형에서 산출된 회귀계수는 해당 부동산을 구성하는 속성의 잠재가격으로 볼 수 있으므로, 혼합 GWR 모형에서 산출된 가변 회귀계수를 하부시장 구획을 위한 투입변수로 활용하였다. 이후 공간적 연결성을 명시적으로 고려하기 위해 최소신장트리(MST)에 기반한 SKATER 알고리즘을 사용, 하부시장을 구획하였다.

구획된 하부시장의 적정성은 개별 필지 수준과 하부시장 수준 간의 위계구조를 감안한 다수준 모형을 적용하여 살펴 보았고, 그 결과 중랑구는 실제로 하부시장이 존재하지 않음을, 강남구는 5개의 하부시장으로 구분하는 것이 (적어도 부동산 가격추정에 있어

서) 합리적임을 확인하였다.

GWR 모형에서 산출된 회귀계수를 사용하여 하부시장을 구획하려는 시도는 있었으나(강창덕, 2010 등) 대부분 산출된 회귀계수를 도면에 연속표면으로 표시하는 등 계수값을 시각화하는 수준에 그쳤다. 또한 혼합 GWR 모형을 적용한 연구(김현중·이성우, 2011 등)의 경우 고정 회귀계수와 가변 회귀계수를 구분해 낸 후, 가변 회귀계수의 공간적 변동성을 해석하는 정도의 시도가 있었다. 본 연구에서는 이러한 시도를 좀더 발전시켜 혼합 GWR 모형에서 산출된 가변 회귀계수를 기초로, 공간적 연결성을 고려한 군집 분석 과정에 투입하여 하부시장을 구획하였는데 차별성을 찾을 수 있다.

분석 결과, 강남구의 경우 간선도로(도산대로, 학동로, 테헤란로)를 중심으로 한, 비교적 의미 있는 하부시장 구획 결과가 도출되었다. 하부시장 구획을 시도한 기존의 연구들이 지역의 지가수준, 인구 수, 소득수준, 주민 연령 등을 주요 변수로 인식하였다면 본 연구에서는 간선도로와 같은 도시의 인프라가 하부시장을 구분 짓는 중요한 요인임을 확인하였다. 간선도로, 고속도로, 교량 등으로 인해 지가분포의 이질성이 증폭되고 더 나아가 지역의 단절 현상이 발생할 수 있다는 것은 일반적인 수준에서도 이해할 수 있는 사안이다. 따라서 향후 부동산 가격 추정을 위해 지역을 세분화하는 노력에 있어 인프라의 구축 여부, 구축되었다면 그 위치와 특징에 보다 관심을 기울일 필요가 있다.

마지막으로 본 연구에서는 부동산 전문가인 감정평가사가 평가한 가격인 표준지 공시지가를 기초 자료로 활용하였다. 그러나 표준지 공시지가 자체에 '평가 주체에 따른 편이'가 포함되었을 가능성이 높으며, 공시지가가 시장가치(market value)를 적정하게 대리할 수 있는지 여부도 논란의 여지가 있다. 따라서 향후에는 공시지가가 아닌 부동산 실거래가 신고가격에 기초한 하부시장 구획을 시도할 필요가 있다.

다만, 부동산 실거래가 자료는 표준지와 달리 공간적 편중이 심한 편이며, 거래가격 자체에 잡음(noise)이 상당 부분 포함되어 있다. 즉 급매에 의한 거래, 매수자 및 매도자의 정보 비대칭, 부동산 중개인의 개

입 정도, 탈세 등 저가신고의 유인 등 수많은 요인으로 인해 인접하여 위치한 유사 부동산이 동일 시점에 거래된다 하더라도 그 가격은 상이할 수 있다. 따라서 실거래가 신고가격에 기초한 하부시장 구획은 이러한 거래잡음에 강건(robust)하거나 적극적으로 처리할 수 있는 모형(예를 들어 거래잡음을 유형별로 구분하여 모형의 구성요소로 반영하는 것 등)을 구성하여 접근할 필요가 있다.

하부시장의 구획은 부동산 가격 추정에 있어 모형의 구성만큼 중요한 양대 이슈임에도 불구하고, 상대적으로 주목받지 못한 분야였다. 그러나 부동산 가격 형성 과정의 공간적 범위 파악이 선행되어야 정확한 가격 추정이 가능해질 것이고, 본 연구도 이러한 측면을 부각시켰다는 점에서 의미를 찾을 수 있다.

주

- 1) 2013년 표준지 공시지가 기준, 강남구는 평균지가 7,290,000원/㎡, 표준편차 5,320,000원, 변동계수(Coefficient of Variation, 표준편차 ÷ 평균 × 100) 73으로, 중랑구는 평균지가 2,220,000원/㎡, 표준편차 1,010,000원, 변동계수 45로 산출되어, 상호 대비되는 지역으로 상기 지역을 선정하였다.
- 2) 공간적 자기상관을 측정하는 Moran's I 값의 경우 강남구 0.10(p-value < 0.01), 중랑구 0.04(p-value < 0.01)로 산출되어(역거리 공간가중행렬 기준) 강남구가 중랑구보다 자기상관성도 강한 것을 확인할 수 있었다.
- 3) 모든 분석은 통계 패키지 R(ver. 3.0.2)을 이용하여 수행하였다.
- 4) 자료의 정규성 등에 부합되도록 공시지가 단가(원/㎡)에 자연로그를 취한 값을 종속변수로 사용하였다.
- 5) k-근접이웃(k-nearest neighbor) 방법을 사용하였고, cross-validation을 통해 AIC 값이 최소가 되는 근접 관찰치 수(강남구 170개, 중랑구 83개)를 가변대역폭으로 정하였다(표 3 하단 참조).
- 6) 종속변수가 자연로그를 취한 값이므로 회귀계수 0.35는 $\exp(0.35)=1.4$ 배의 가격 격차를, 0.59는 $\exp(0.59)=1.8$ 배의 가격 격차를 의미한다.
- 7) 정규 크리깅(ordinary kriging)을 활용하여 연속표면을 구축하였다.
- 8) 하부시장 수를 5개가 아닌 2~8개로 정한 후 후술하는 구획

방법과 다수준 모형(본 연구에서 최종적으로 적용한 모형)을 적용한 결과, 모형성능을 측정하는 지표 COD(Coefficient Of Dispersion, 작을수록 모형성능 우수)의 변화 추이는 다음과 같았다.

하부시장 수	2	3	4	5	6	7	8
강남구	18.76	17.96	17.04	16.55	16.54	16.51	16.55
중랑구	9.50	9.49	9.48	9.44	9.40	9.40	9.44

즉, 강남구는 하부시장 수 5개 이상부터 COD 값이 일정하게 유지되었고, 중랑구는 전반적으로 큰 변화가 없었다. 따라서 강남구 하부시장을 5개를 초과하여 구획하는 것은 큰 의미가 없는 것으로 판단되며, 중랑구에 대한 해석은 본 논문의 「4. 하부시장에 대한 해석」 참조.

- 9) 공시지가 등 국내 가격공시제도에서 가격 적정성은 COD를 기준으로 검토하고 있으며 이는 해외의 경우도 마찬가지이다.
- 10) IAAO(International Association of Assessing Officers) 기준의 경우 아파트 등 비교적 표준화된 부동산은 COD 10.0 이하, 단독주택은 15.0 이하, 상업용은 25.0 이하 등으로 상한선을 정하고 있다. 주로 미국, 캐나다 등 북미의 지역 사정에 맞는 기준으로 비교적 좁은 공간에 다수의 이질적 부동산이 분포하는 우리나라의 경우에는 상한선을 이보다 높여 적용하여야 할 것으로 판단된다.

참고문헌

강창덕, 2010, “GWR 접근법을 활용한 부동산 감정평가 모형 연구: 서울시 아파트를 사례로”, 부동산 연구, 20(2), 107-132.

김현중 · 이성우, 2011, “수도권 5대 범주의 결정요인: Mixed GWR 모형의 적용”, 서울도시연구, 12(4), 137-155.

원제무 · 정광섭 · 김상원 · 백진호 · 백기형, 2009, “시장세분화를 통한 주상복합주택 가격 결정요인 특성에 관한 연구”, 국토계획, 2009, 44(3), 137-147.

Assunção, R. M., M. C. Neves, G. Camara and C. Da Costa Freitas, 2006, Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees, *International Journal of Geographical Information Science*, 20(7), 797-811.

Bayer, P., McMillan, R. and Rueben, K., 2003, *An Equilib-*

- rium Model of Sorting an Urban Housing Market: The Causes and Consequences of Residential Segregation*, Vol.860, Economic Growth Center Discussion Paper, Yale University, New Haven.
- Bourassa, S. C., Cantoni, E. and Hoesli, M., 2007, Spatial Dependence, Housing Submarkets, and House Price Prediction, *Journal of Real Estate Finance and Economics*, 35(2), 143-160.
- Bourassa, S. C., Hamelink, F., Hoesli, M. and MacGregor, B. D., 1999, Defining Housing Submarkets, *Journal of Housing Economics*, 8(2), 160-183.
- Bourassa, S. C., Hoesli, M and Peng, V. S., 2003, Do housing markets really matter?, *Journal of Housing Economics*, 12(1), 12-28.
- Brunsdon, C., Fotheringham, S. and Charlton, M., 1998, Geographically weighted regression - modeling spatial non-stationarity, *The Statistician*, 47, Part 3, 431-443.
- Fotheringham, S. A., Charlton, M. and Brunsdon, C., 2002, *Geographically weighted regression: the analysis of spatially varying relationships*, Wiley, Chichester.
- Fotheringham, S. A. and Rogerson, P. A., 2009, *The SAGE Handbook of Spatial Analysis*, SAGE Publications, London.
- Gatzlaff, D. H. and M. T. Smith, 1993, The Impact of Metrorail on the Value of Residences Near Station Locations, *Land Economics*, 69, 54-66.
- Goodman, A. C., 1978, Hedonic Price, Price Indices and Housing Markets, *Journal of Urban Economics*, 5(4), 471-484.
- Goodman, A. C. and Thibodeau, T. G., 2007, The Spatial Proximity of Metropolitan Area Housing Submarkets, *Real Estate Economics*, 35(2), 209-232.
- Guo, D., 2008, Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP), *International Journal of Geographical Information Science*, 22(7), 801-823.
- Haider, M. and Miller, E. J., 2000, Effects of transportation infrastructure and location on residential real estate values: application of spatial autoregressive techniques, *Journal of the Transportation Research Board*, 1722(1), 1-8.
- Haining, R., Wise, S. and Ma, J., 2000, Designing and implementing software for spatial statistical analysis in a GIS environment, *Journal of Geographical System*, 2, 257-286.
- Helbich, M., Brunauer, W., Hagenauer, J. and Leitner M., 2013, Data-Driven Regionalization of Housing Markets, *Annals of the Association of American Geographers*, 103(4), 871-889.
- International Association of Assessing Officers, 2010, *Standard On Ratio Study*, Kansas City.
- Islam, K. S. and Asami, Y., 2009, Housing Market Segmentation: A Review, *Review of Urban & Regional Development Studies*, 21(2), 93-109.
- Jones, K. and Bullen, N., 1993, A Multi-level Analysis of the Variations in Domestic Property Prices: Southern England, *Urban Studies*, 30(8), 1409-1426.
- Legendre, P. and Legendre, L., 2012, *Numerical Ecology*, 3rd english edition, Elsevier, London.
- Martin, D., 1998, Optimizing census geography: the separation of collection and output geographies, *International Journal of Geographical Information Science*, 17, 181-196.
- Michaels, R. and Smith, V., 1990, Market Segmentation and Valuing Amenities with Hedonic Models: The Case of Hazardous Waste Sites, *Journal of Urban Economics*, 28, 223-242.
- Openshaw, S., 1977, A geographical solution to scale and aggregation problems in region - building, partitioning and spatial modeling, *Transactions of the Institute of British Geographers*, 2, 459-472.
- Paez, A., Fei, L. and Farber, S., 2008, Moving window approaches for hedonic price estimation: An empirical comparison of modeling techniques, *Urban Studies* 45(8), 1565-1581.
- Rothenberg, J. G., Galster, R. and Pitkin, J., 1991, *The Maze of Urban Housing Market: Theory Evidence and Policy*, University of Chicago, Chicago.
- Tan, P.N., Steinbach, M., Kumar, V., 2006, *Introduction to Data Mining*, WP Co, New Delhi.
- Tu, Y., Sun, H. and Yu, S. M., 2007, Spatial Autocorrela-

- tions and Urban Housing Market Segmentation, *Journal of Real Estate Finance and Economics*, 34(3), 385-406.
- Watkins, C. A., 2002, The Definition and Identification of Housing Submarket, *Environment and Planning*, A 33(12), 2235-2253.
- Wheeler, D. C., Paez, A., Spinney, J. and Waller, L. A., 2013, A Bayesian approach to hedonic price analysis, *Papers in Regional Science*, online article.
- Wheeler, D., Tiefelsdorf, M., 2005, Multi-collinearity and Correlation among Regression Coefficients in Geographically weighted regression, *Journal of Geographical Systems*, 7(2), 161-187.
- Wu, C. and Sharma, R., 2012, Housing submarket classification: The role of spatial contiguity, *Applied Geography*, 32, 746-756.
- Yates, J. and Mackay, D. F., 2006, Discrete Choice Modeling of Urban Housing Markets: A Critical Review and an Application, *Urban Studies*, 43(3), 559-581.
- 교신: 박기호, 151-742, 서울시 관악구 관악로 599, 서울대학교 지리학과(이메일: khp@snu.ac.kr, 전화: 02-880-6453, 팩스: 02-876-9498)
- Correspondence: Key Ho Park, Department of Geography, Seoul National University, 599, Gwanangno, Gwanak-gu, Seoul, 151-742, Korea (e-mail: khp@snu.ac.kr, phone: +82-2-880-6453, fax: +82-2-876-9498)

최초투고일 2014. 5. 19

수정일 2014. 6. 16

최종접수일 2014. 6. 24