

수도권의 공간적 주택하위시장 식별에 대한 연구

Identification of Spatial Housing Submarkets in the Capital Region
Using Self Organizing Map and Chow Test

손 철

Sohn Chul

강릉원주대학교 도시계획부동산학과 부교수

Associate Professor, Dept. of Urban Planning and
Real Estate, Gangneung-Wonju National Univ.
(csohn@gwnu.ac.kr)

목 차

- I. 연구의 배경
- II. 선행연구에 대한 검토
 - 1. 공간적 하위시장 관련 이론에 대한 검토
 - 2. 공간적 하위시장 관련 경험적 연구에 대한 검토
- III. 연구목적, 설계, 방법, 데이터
 - 1. 연구목적 및 설계
 - 2. 연구방법
 - 3. 연구데이터
- IV. 분석결과
 - 1. SOM 분석결과
 - 2. Chow Test 결과
- V. 결론

※ 본 연구는 2009년 강릉원주대학교 해외장기과건연구비 지원에 의해 이루어진 “Neural Network과 GIS를 이용한 도시주택시장의 분석” 연구의 일부임.

I. 연구의 배경

주택시장에 대한 헤도닉 분석은 일반적으로 주택이라는 상품을 구성하는 다양한 속성들의 한계잠재가격(Marginal Implicit Price)을 추정하는 것을 목적으로 한다. 그런데 주택하위시장이 존재할 경우 주택하위시장에 따라 동일한 속성에도 다른 크기의 한계잠재가격이 존재할 가능성이 있다. 이 때문에 헤도닉 분석의 수행 시 연구자는 자신이 분석하고자 하는 주택 데이터가 여러 개의 주택하위시장으로부터 얻어진 것인지를 사전에 파악할 필요가 있다. 그 이유는 실제로 존재하는 여러 주택하위시장으로부터 수집된 주택 데이터를 통합하여 하나의 헤도닉 함수를 추정할 경우 어떤 주택하위시장의 특성도 제대로 반영하지 못하는 한계잠재가격에 대한 추정치를 얻을 수 있기 때문이다.

서울특별시, 인천광역시, 경기도로 구성되는 수도권은 그간 이 지역이 가진 다양한 주택 관련 문제, 민간 부동산정보업체가 구축해온 공동주택에 대한 가격 및 속성 데이터베이스, 해당 자치단체가 구비한 지리정보 데이터베이스로 다양한 헤도닉 분석의 대상이 되어왔다. 또한 앞서 언급한 주택하위시장의 중요성 때문에 강창덕(2010), 정건섭·이상엽(2007), 김주영·우경(2004), 고원용·김홍규·유완(2001) 등의 선행연구들은 수도권 전체 혹은 서울시를 대상으로 공간적 주택하위시장을 분석했다.

그러나 이들 연구의 대부분이 수도권 가운데 서울의 주택시장만을 분석하고 있으며 비서울권의 주택하위시장의 공간적 영역에 대해서는 그것이 개별 시·군·구 혹은 개별동을 단위로 한 개별화된 형태인지 아니면 몇 개의 시·군·구가 통합된 좀 더 광역화된 형태인지에 대해 명확한 정보를 제공해주지 못하고 있다. 따라서 선행연구 결과는 서울을 포함한 전체 수도권을 대상으로 한 헤도닉 분석 시 과연 어떤

형태의 공간적 하위시장을 고려해야 할 것인가에 대한 질문에 적절한 대답을 제공해주지 못하고 있다. 이에 본 연구에서는 수도권을 대상으로 선행연구에서 구체적으로 다루지 않은 연구질문, 즉 수도권 내 비서울 지역 공동주택시장에 존재하는 주택하위시장을 공간적으로 보다 세분화하는 것이 타당한지 아니면 보다 광역적으로 파악하는 것이 타당한지를 검토해보고자 한다.

II. 선행연구에 대한 검토

1. 공간적 하위시장 관련 이론에 대한 검토

주택하위시장은 상호 대체성이 크며 다른 주택하위시장에 속한 주택과는 대체성이 작은 주택들의 집합으로 정의된다(Bourassa et al. 2003). 주택이 다른 재화와 구분되는 특징은 위치가 고정되어 있고 내구성을 가지며, 다양한 속성으로 구성되어 있다는 점이다(Schnare and Struyk. 1976; Bitter et al. 2007). 이러한 특성 때문에 Schnare and Struyk(1976)는 특정 유형의 주택공급이 매우 비탄력적이고 해당 유형의 주택에 대해 비탄력적 수요가 많은 소비자에 의해 공유될 때 주택하위시장이 존재할 수 있다고 지적했다. 또한 이 경우 형성되는 주택하위시장에는 고유의 공급조건과 수요조건이 존재하기 때문에 이 시장에서 결정되는 주택특성들의 균형잠재가격(Implicit Price) 역시 이러한 고유의 특성을 반영하여 결정된다고 지적했다.

주택하위시장별로 특유의 주택특성들에 대한 균형잠재가격이 존재할 경우 헤도닉 함수의 추정 시 주택하위시장에 대한 바른 인식과 고려는 추정된 함수의 예측력을 향상시키고 주택가격의 공간적·시간적 변화를 보다 잘 모델링할 수 있도록 도와준다(Goodman and Thibodeau. 2007). 헤도닉 함수의 추

정 시 주택하위시장은 주택의 구조적 특성(단독주택, 공동주택 등), 물리적 속성(건축연령 등), 근린의 질(교육의 질, 안전, 소득 등 근린의 매력도) 등 다양한 속성에 의해 정의되고 분석에 이용된다(Islam and Asami, 2009; Goodman and Thibodeau, 1998). 이렇듯 다양한 기준이 주택하위시장의 정의에 이용되고 있지만 아직까지 최선의 기준이 무엇인가에 대해서는 연구자 간에 의견이 일치되지 않고 있다(Islam and Asami, 2009).

공간적 하위시장(Spatial Submarket)은 상호 대체성이 큰 주택들의 집합이 공간적으로 근접하여 존재하는 주택하위시장을 의미한다. 이 하위시장은 주로 근린의 질에 대한 비탄력적 수요(Schnare and Struyk, 1976), 주택시장 참여자의 정보제약(Michaels and Smith, 1990) 등에 의해 생성된다. 공간적 하위시장이 존재할 경우 수요자는 해당지역에 위치한 주택만을 선택하려 하고, 해당지역에는 특수한 주택공급 조건이 존재하기 때문에 주택특성에 대해 하위시장 고유의 균형잠재가격이 존재할 가능성이 크다(Schnare and Struyk, 1976). 이런 맥락에서 Bitter et al.(2007)은 주택특성의 균형잠재가격이 공간적 하위시장을 따라 변함에도 이를 무시하고 연구지역에 대해 단 하나의 헤도닉 함수를 추정할 경우 여기서 얻어진 추정된 계수는 “여러 지역의 계수값이 평균화된 값(average value of the parameter over all location)”을 나타낼 뿐이라고 지적했다.

헤도닉 분석 시 공간적 하위시장의 정의를 이용하는 연구자들은 이러한 하위시장별 고유의 균형잠재가격의 존재 가능성을 분석에서 고려하기 위해 일반적으로 다음과 같은 3개의 함수형태를 사용한다.

$$P_i^m = \alpha^m + \sum_k \beta_k^m X_{ik}^m + \varepsilon_i \quad <식 1>$$

$$P_i = \alpha + \sum_k (\beta_{k0} + \beta_{k1} NQ) X_{ik} + \varepsilon_i \quad <식 2>$$

$$P_i = \alpha + \sum_k \beta_k (\mu_i, \nu_i) X_{ik} + \varepsilon_i \quad <식 3>$$

단, m은 공간적 하위시장, P 는 주택가격, X_{ik} 는 i 샘플의 k번째 주택특성 변수, NQ는 근린의 질 지표, α, β 는 추정계수, μ_i, ν_i 는 i가 위치한 지점의 좌표, ε 는 오차항

<식 1>은 공간분할에 의해 주택시장 데이터를 분할하고 각 하위시장별로 개별적인 헤도닉 함수를 추정하는 방법이다. 공간분할을 위해서는 전문가의 판단에 의하거나 일반적으로 근린의 질(Neighborhood Attraction, Neighborhood Quality) 관련 변수를 이용하여 개별 행정구역 또는 센서스 구역을 몇 개의 커다란 지역으로 묶는 방법이 이용된다(Watkins, 2001). <식 2>는 Casetti(1972)의 공간확장방법(Spatial Expansion Method)을 Can(1992)이 헤도닉 함수 추정에 적용한 것이다. <식 2>에서 주택속성의 한계잠재가격은 근린의 질과의 상호작용에 의해 결정된다. 그리고 <식 3>은 헤도닉 함수 추정에 지리가중회귀분석(Geographically Weighted Regression: GWR)(Fortheringham et al. 2002)을 적용하여 추정계수의 크기가 공간상의 위치에 따라 변할 수 있도록 허용한 것이다. <식 3>은 Bitter et al.(2007) 등에 의해 헤도닉 함수 추정을 위해 사용되었다.

<식 1~3>은 공간적 하위시장을 보는 근본적으로 다른 시각을 나타내고 있다. <식 1>은 대도시에 존재하는 공간적 하위시장을 이산적인(discrete) 형태로 보고 존재하는 하위시장별로 독립적인 함수를 추정하기 위한 식이다. 그리고 <식 2>와 <식 3>은 이와는 달리 하위시장을 연속적(Continuous)인 형태로 보는 시각을 담고 있다. 즉, 헤도닉 함수 추정계수의 연속적 변화를 추정하려는 것이 <식 2>와 <식 3>이다. 두 식의 차이점을 살펴보면 <식 2>

에서는 추정계수의 변화를 초래하는 요인으로 근린의 질을 명시적으로 고려했다는 점이고, <식 3>은 이러한 고려 없이 위치에 따라 추정계수가 변화하도록 자유롭게 허용했다는 점이다.

2. 공간적 하위시장 관련 경험적 연구에 대한 검토

현재까지 강창덕(2010), Song and Quercia(2008), Alkay(2008), 정건섭·이상엽(2007), 김주영·우경(2004), Bourassa et al.(1999) 등 국내외 다수의 연구들이 소득 등 근린의 질을 나타내는 사회·경제적 지표와 전문가의 판단을 이용하여 공간적 하위시장을 정의하고 대도시 내 이산적 혹은 연속적 공간적 하위시장이 과연 존재하는가를 경험적으로 검토했다. 주요 연구들의 결과를 살펴보면 다음과 같다.

강창덕(2010)은 헤도닉 함수 추정 시 주택위치에 따른 한계잠재가격의 변화를 허용하는 지리가중회귀분석과 2008년 3월 서울시 공동주택의 실거래가 자료를 이용하여 서울시 주택시장에 존재하는 공간적 하위시장에 대해 분석했다. 분석결과는 공동주택의 위치에 따라 한계잠재가격의 크기가 상당히 크게 변화하는 모습을 보여주었다. 이는 서울시가 공간적으로 상당히 분화된 다수의 주택하위시장으로 구성되어 있음을 시사하는 것이다.

Song and Quercia(2008)는 미국 오리건주 워싱턴 카운티의 186개 센서스 구역(Census Block Group)으로부터 근린디자인에 대한 다양한 지표를 구한 후 군집분석에 의해 186개 구역을 6개 유형의 군집으로 묶었다. 그리고 6개의 군집에 대해 독립된 헤도닉 함수를 추정하고 Chow Test를 수행한 결과 군집간에 추정계수의 차이가 존재함을 발견했다.

Alkay(2008)는 터키 이스탄불 주택시장을 근린의 평균소득수준을 이용하여 3개 주택하위시장으로 구분했다. 그리고 3개 주택하위시장별로 헤도닉 함수

를 추정한 결과 3개 모델에서의 추정계수가 전체 시장을 대상으로 한 추정계수와 통계적으로 유의한 차이가 있음을 보고했다.

정건섭·이상엽(2007)은 서울 강북지역의 지리적으로 인접한 노원, 중랑, 도봉, 강북, 성북 등 5개 구를 대상으로 2002년 공동주택 자료를 이용하여 헤도닉 함수를 추정했다. 추정 후 Wald Test를 통해 추정된 함수계수들이 구별로 다른지를 검증했다. 분석결과 노원·중랑, 도봉, 강북, 성북 등 4개의 주택하위시장이 드러났다.

Bitter et al.(2007)은 미국 애리조나주 투산시의 주택시장에 대한 헤도닉 함수 추정에 공간확장방법(Spatial Expansion Method)과 지리가중회귀분석(GWR)을 적용했다. 분석결과는 두 방법 모두 설명력과 예측력 측면에서 통상적인 회귀모델에 비해 우월한 결과를 보였으며, 지리가중회귀분석의 결과가 가장 우월한 것으로 나타났다.

Kuoko(2005)는 네덜란드 전체의 통계 집계구를 20개의 인구·사회 및 경제적 지표와 16x12 뉴론으로 구성된 자기조직화 지도(Self Organizing Map)를 이용하여 군집화했다. 분석결과는 네덜란드 전체의 국가적 주택시장이 5개 하위시장(Truely Urban, Urban Periphery, Pseudo-rural, Agricultural Rural, Areas Between)으로 나뉘질 수 있음을 나타냈다.

김주영·우경(2004)은 1993년 실시된 수도권 도시거주 가구를 대상으로 한 주거만족도 및 주택수요 조사자료를 이용하여 수도권 주택하위시장의 존재를 분석했다. 분석결과는 서울시청으로부터의 물리적 거리를 기준으로 주택시장을 3개 권역으로 나누어 3개의 헤도닉 함수를 추정한 결과가 수도권 전체를 대상으로 하나의 헤도닉 함수를 추정한 결과에 비해 우월함을 보여주었다. 이상의 결과는 수도권에서 몇 개의 시·군·구를 통합한 광역적 주택하위시장이 존재할 가능성을 보여주는 것이다.

Watkins(2001)는 연구자의 사전지식에 근거해 영국 글래스고시의 주택시장을 6개 하위시장(Central, Northwest, East, South, Southwest 등)으로 구분했다. 그리고 6개 시장별로 헤도닉 함수를 추정하고 Chow Test를 이용하여 하위시장 간 추정된 헤도닉 함수의 안정성을 검증했다. 분석결과는 대부분의 Chow Test에서 계수의 안정성이 기각되어 글래스고시의 주택 시장에 공간적 하위시장이 존재함을 보였다.

고원용·김홍규·유완(2001)은 1997년 서울시의 공동주택의 실제 매매가격을 이용하여 하위시장을 전체 공동주택 평당가격의 평균치보다 높은 시장과 낮은 시장으로 구분했다. 그리고 두 개의 시장에 대해 추정된 헤도닉 함수에 대해 Chow Test를 수행한 결과 서울시에 상위가격 시장과 하위가격 시장이라는 주택하위시장이 존재하는 것을 발견했다. 이 연구는 공동주택의 평당 가격을 기준으로 공동주택시장을 구분했다. 그러나 이는 평당가격이 유사한 공동주택 단지가 공간적으로 가까운 위치에 입지할 가능성이 있어 서울 지역에 공간적 하위시장이 존재함을 간접적으로 암시하는 결과라 할 수 있다.

Bourassa et al.(1999)은 호주의 시드니와 멜버른 대도시권에 속하는 행정구역(Local Government Area: LGA)에 대해 대도시 중심으로부터의 거리를 이용한 5개 구분, 행정구역의 사회·경제적 속성을 이용한 군집분석으로 얻은 5개 구분, 행정구역의 사회·경제적 속성과 주택의 속성을 이용해 얻은 5개 구분으로 각각 주택하위시장을 정의했다. 그리고 전체 대도시권을 대상으로 추정된 헤도닉 함수의 Mean Square Error와 3개 주택하위시장 정의에 5개씩 추정된 헤도닉 함수의 Weighted Mean Square Error를 비교했다. 분석결과는 3개의 주택하위시장의 정의를 이용한 헤도닉 함수의 추정결과가 단일시장을 가정한 헤도닉 함수의 추정결과보다 우월함을 나타냈다.

Can(1992)은 미국 오하이오주 콜럼버스시의 센서

스 구역에서 얻어진 사회·경제적 특성 자료(비백인 인구 비율, 가계소득, 실업률 등)에 주성분 분석을 적용하여 근린의 질 지수를 구하고 이를 동일지역의 헤도닉 함수 추정 시 설명변수들과 상호작용항으로 포함시켰다. 그리고 상호작용항이 통계적으로 유의미함을 발견하여 근린의 질에 따라 연속적으로 추정 계수가 변화함을 보였다.

Michaels and Smith(1990)는 부동산중개인의 전문 지식을 이용하여 미국 보스턴시 교외의 소도시를 5개 하위시장(Premier, Above Average, Average, Below average)으로 지리적으로 구분하고 각 하위시장에 대해 독립변수로 쓰레기 매립지에 대한 거리를 포함하는 헤도닉 함수를 추정했다. 그리고 Tiao-Goldberger Test를 이용하여 5개 하위시장에 대해 추정된 함수로부터의 추정계수가 동일한가를 검정했다. 검정결과, 추정된 모델에 포함된 20개의 독립변수 중 16개의 독립변수에서 동일성이 기각되어 이 지역에 공간적 하위시장이 존재함이 통계적으로 드러났다.

이상의 연구결과들은 공간적 하위시장 정의에 다양한 기준이 이용되고 있으며 국내외 대도시권 주택 시장에 공간적 하위시장이 존재할 수 있음을 보여준다. 또한 Chow Test가 공간분할 후 하위시장의 존재를 검증하기 위해 다수의 연구에서 이용되었음을 보여준다. 그리고 강창덕(2010), Bitter et al.(2007), Can(1992) 등의 연구는 연속적 성격의 공간적 하위시장이 존재할 수 있음을 나타낸다.

이상의 연구들 가운데 서울을 포함한 수도권 지역에 대한 강창덕(2010), 정건섭·이상엽(2007), 김주영·우경(2004), 고원용·김홍규·유완(2001)의 연구결과는 공통적으로 공간적 하위시장이 수도권에 존재함을 보여준다. 그러나 그 공간적 범위에서는 서로 다른 연구결과가 나타난다. 김주영·우경(2004)은 수도권에 몇 개의 시·군·구를 통합한 3개의 광역적 주택하위시장이 존재할 가능성을 나타냈다. 반면 강창

덕(2010), 정건섭·이상엽(2007) 등은 서울에서 주로 개별 구별로 혹은 그 이하의 공간적 수준에서 매우 세분화된 주택하위시장이 존재할 가능성을 나타내고 있다. 고원용·김홍규·유원(2001)의 연구는 공간적 하위시장의 존재를 암시할 뿐 그 구체적 형태에 대한 정보를 제공하지 않는다.

만약 김주영·우경(2004)의 연구결과와 같이 광역적인 주택하위시장이 수도권에 존재한다면 연구자들은 인접 몇 개 시·군·구의 주택시장 데이터를 보다 자유롭게 통합하여 이용할 수 있을 것이다. 그러나 강창덕(2010), 정건섭·이상엽(2007) 등의 연구에서와 같이 개별 구 단위 혹은 그 이하의 공간적 단위에서 독립된 주택하위시장이 존재할 가능성이 있다면 연구자는 한 개의 행정구역을 넘어서는 주택시장 데이터를 연구에 이용할 경우 주택하위시장의 존재를 사전에 엄밀하게 검증할 필요가 있다.

수도권의 공동주택시장을 대상으로 하는 헤도닉 분석이 활발하게 이루어지고 있는 시점에서 수도권에 존재하는 주택하위시장의 형태에 대한 연구는 매우 중요한 의미를 가진다. 이 경우 헤도닉 분석 시 주택하위시장의 존재를 명시적으로 고려해야 하기 때문이다. 선행연구들은 수도권에 주택하위시장이 존재함을 나타낸다. 그러나 그 시장의 형태에 대해서는 명확한 답을 주고 있지 않다. 특히 서울 지역에 대해서는 비교적 공간적으로 세분화된 하위시장이 존재하는 것으로 판단할 수 있는 여러 경험적 증거가 제공되고 있으나, 비서울 지역에서 주택하위시장이 개별 시·군·구 혹은 그 이하의 공간적 단위를 기준으로 하는 개별화된 형태를 보이고 있는지 아니면 몇 개 이상의 시·군·구를 포함하는 보다 광역화된 형태를 보이고 있는지 선행연구들은 확실한 답을 주고 있지 않다.

III. 연구목적, 설계, 방법, 데이터

1. 연구목적 및 설계

본 연구는 선행연구에서 구체적으로 다루지 않은 연구질문, 즉 비서울 지역의 수도권 주택시장에 존재하는 주택하위시장을 공간적으로 보다 세분화하는 것이 타당한지 아니면 보다 광역적으로 파악하는 것이 타당한지를 다룬다.

이러한 연구목적을 달성하기 위해 본 연구는 정건섭·이상엽(2007)과 유사하게 지리적으로 인접한 시·군·구에 대해 헤도닉 함수를 추정하고 Chow Test를 통해 헤도닉 함수 계수의 동일성을 검증하는 방법을 사용했다. 그러나 정건섭·이상엽(2007)과 달리 연구자 개인의 판단에 의해 단순한 지리적 인접성을 기준으로 비교대상 시·군·구를 선택하지 않고 Song and Quercia(2008), Kuoko(2005), Bourassa et al. (1999)과 같이 근린의 질과 관련 있는 사회·경제적 특성 자료를 이용하여 수도권 시·군·구를 몇 개의 동질적 군집으로 나누었다. 그 후 비서울 지역 시·군·구를 포함한 1개의 군집을 선택하고 해당 군집에 속한 지리적으로 인접한 시·군·구에 대해 연구를 수행했다. 즉, 본 연구에서는 지리적 근접성만이 아니라 근린의 질의 유사성을 모두 고려하여 Chow Test의 대상을 설정했다. 본 연구에서의 이러한 고려는 수도권 주민의 주거선택 시 근린의 질 측면에서 서로 긴밀한 대체관계에 있을 시·군·구를 연구자 주관이나 단순한 지리적 기준이 아닌 근린의 질을 고려한 보다 객관적 기준에 의해 선택하기 위한 것이다.

본 연구의 연구설계는 2단계로 구성되었다. 1단계에서는 수도권에 위치한 시·군·구를 시·군·구의 근린의 질을 대리하는 사회·경제적 지표를 이용하여 몇 개의 군집으로 분류했다. 하위시장의 정의를 위해 근린의 질을 대리하는 사회·경제적 지표를 사용

하는 것은 근린의 질(Neighborhood Quality)이 공간적 하위시장을 구분하는 중요한 요소라는 점을 이용한 것이다(Bourassa et al. 1999; Schnare and Struyk. 1976). 2단계에서는 1단계에서의 군집들에서 비서울 지역 시·군·구를 포함하는 1개의 군집을 선택하고 해당 군집에 속하고 행정경계를 접한 시·군·구에 대해 공동주택 헤도닉 함수를 추정하여 각 함수의 추정계수 간에 차이가 존재하는가를 Chow Test를 통해 검증했다.

2단계의 Chow Test가 많은 경우 헤도닉 함수들로부터의 추정계수가 동일하다는 결과를 보여준다면 이것은 수도권 비서울 지역에 광역적 형태의 주택하위시장이 존재할 가능성을 보여주며, 반대의 경우는 보다 세분화된 시·군·구 단위의 개별화된 주택하위시장이 존재할 가능성을 보여주는 것으로 해석할 수 있다.

1단계에서의 군집을 설정하기 위해서 본 연구에서는 2005년 센서스와 정부기관에서 나온 사회·경제적 지표와 Neural Network에 기초한 군집분석방법 중 하나인 자기조직화 지도(Self Organizing Map: SOM)를 이용했다. SOM은 공동주택 가격데이터가 없는 옹진군을 제외한 78개 시·군·구를 군집화하기 위해 이용되었다. SOM은 K-means 방법 등 통상적 군집방법과 동일한 군집화 방법이나 군집결과를 2차원의 지도에 시각적으로 알기 쉽게 표현해준다는 장점이 있다(Budayan et al. 2009).

2. 연구방법

1) 자기조직화 지도(SOM: Self Organizing Map)¹⁾

SOM은 다차원의 자료벡터를 통상 2차원의 격자망

(Lattice)에 맵핑시키는 비감독 기계학습(Unsupervised Machine Learning) 방법이다. 2차원의 격자망은 육각형 혹은 사각형의 노드(Node)로 구성된다. 격자망의 노드는 노드를 대표하는 가중치(Weight)를 가진다. 이 노드의 가중치는 SOM에 투입되는 자료벡터와 같은 차원을 갖는다.

SOM의 학습과정이 시작되면 2차원 노드 가중치 값이 임의로 초기화된다. 그리고 투입되는 자료벡터와 모든 노드 가중치 사이의 거리가 측정된다. 이때 통상 유클리디안 거리가 이용된다. 거리측정 결과 가장 가까운 노드에 자료벡터가 배정된다. 이때 투입된 자료벡터와 가장 가까운 노드를 Winning Node 혹은 Best Matching Unit이라 한다. 매번 새로운 자료벡터가 투입될 때마다 이 자료벡터와 가장 가까운 노드와 이웃 노드의 가중치 값이 투입된 자료벡터 값에 가깝도록 조정된다. 이때 사용되는 식이 <식 4>다.

$$m_i(t+1) = m_i(t) + h_{ci}(t) [x(t) - m_i(t)]$$

$$h_{ci}(t) = \alpha(t) \cdot e^{-d_{ci}^2 / 2\delta_i^2(t)} \quad \text{<식 4>}$$

x: n차원의 자료 벡터

m: n차원의 가중치 벡터

t: 시간

h_{ci} : 이웃을 정의하는 함수(Neighborhood Function). 통상 Gaussian Model

$\alpha(t)$: 학습률(Learning Rate)

d_{ci} : 이웃하는 노드와의 거리

δ : 이웃의 최대범위를 정의하는 Kernel Width(δ 는 SOM이 자료데이터를 훈련하는 초기에는 넓게 후반기는 좁게 정의됨)

1) SOM에 대한 설명은 Yao(2007)와 Agarwal and Skupin(2008)에 근거한 것임.

이런 과정을 거쳐 모든 자료벡터가 SOM 알고리즘에 이용되면 1회의 반복(Iteration)이 끝나게 된다. 이런 반복과정은 통산 일정한 범위에서 노드 가중치의 변화가 없을 때까지 반복된다. 최종적으로 모든 반복과정이 끝나면 격자망을 구성하는 노드의 가중치들은 격자망에서 가까이 위치해 있을수록 유사한 값을 가지게 된다. 최종적으로 2차원 격자망의 훈련(Training)이 완성되면 SOM에 투입된 각각의 자료벡터는 가장 유사한 값을 가진 Winning Node에 소속시킬 수 있다. 즉, Winning Node를 기준으로 몇 개의 군집이 형성된다.

일반적으로 SOM을 구성하는 노드의 수가 많기 때문에 Winning Node를 기준으로 생성되는 군집의 수도 많다. 따라서 군집의 수를 좀 더 축소하기 위해 최종적으로 얻어진 가중치 값을 이용하여 2차적으로 새로운 군집을 형성하는 방법이 많이 이용된다. 노드의 가중치값을 이용하여 2차적 군집을 형성하기 위해서는 일반적으로 Hierarchical Clustering 또는 K-means Clustering 등의 군집방법이 이용된다(Lu et al. 2006). 본 연구에서는 1차적으로 얻어진 최종적 노드 가중치 값을 K-means 군집방법을 이용하여 2차적으로 군집화하는 방법을 사용했다.

K-means 군집방법을 이용하기 위해서는 총군집의 수가 사전적으로 설정되어야 한다. 본 연구에서는 수도권 사회·경제적 군집의 수를 최대 10개로 설정하고 Vesanto and Alhoniemi(2000)에서와 같이 1에서 10까지의 군집 수 중 Davies-Bouldin index (v_{DB})를 최소화하는 군집 수를 최종적으로 선택했다.²⁾

$$v_{DB}(C) = \frac{1}{C} \sum_{k=1}^C \max_{l \neq k} \left\{ \frac{S_c(Q_k) + S_c(Q_l)}{d_{ce}(Q_k, Q_l)} \right\} \quad <식 5>$$

C: 군집의 수

$S_c = \sum_i \|X_i - c_k\| / N_k$: 군집중심과 군집에 포함된 데이터 간의 거리

c_k : 군집 k의 중심값

N_k : 군집 k에 존재하는 데이터의 수

$d_{ce} = \|c_k - c_l\|$: 군집 Q_k 와 Q_l 의 거리

2) Chow Test

2개의 주택시장(주택시장[1]과 주택시장[2])에 대해 추정된 헤도닉 함수의 추정계수가 서로 동일한가를 검정하기 위한 Chow Test는 다음과 같은 방법에 의해 수행될 수 있다. 첫째, 주택시장[1]과 주택시장[2]에 대한 데이터를 결합한 후 <식 6>과 같은 회귀식을 추정한다. 둘째, 추정된 계수를 이용하여 <식 7>과 같은 귀무가설을 검정한다.

$$Price = \alpha + \beta^1 x + \beta^2 z + \alpha' d + \beta^1 dx + \beta^2 dz + u \quad <식 6>$$

Price: 주택가격

x, z: 주택속성

d=1, 주택시장[2]일 경우, d=0, 주택시장[1]일 경우

$\alpha, \beta^1, \beta^2, \alpha', \beta^1', \beta^2'$: 추정계수

u: 오차항

$$H_0: \alpha' = \beta^1' = \beta^2' = 0 \quad <식 7>$$

2) 최적의 군집분류에서는 군집 내의 거리는 최소화되고 군집 간의 거리는 최대화되어야 함. 따라서 최적의 군집 수는 v_{DB} 를 최소화하는 점에서 얻어짐(Vesanto and Alhoniemi, 2000).

3. 연구데이터

표 1_ 주요 선행연구에서의 사용변수

1) SOM 분석을 위한 변수의 정의

본 연구에서는 SOM을 이용해 수도권 시·군·구를 몇 개의 사회·경제적 구성이 유사한 동질권역으로 구분하기 위해 2005년 센서스와 관련 정부기관의 통계로부터 13개 사회·경제적 지표를 구하여 이용했다. 지표의 설정을 위해 Kauko (2005), Bourassa et al.(1999) 등의 연구에서 호주, 네덜란드 등의 공간적 하위시장 식별을 위해 사용된 변수들을 참고했다. <표 1>은 두 연구에서 군집분석을 위해 사용된 변수를 보여주며, Kauko (2005)와 Bourassa et al.(1999)은 공통적으로 소득, 실업률 등 근린의 경제적 수준과 관련된 변수들을 사용한 것을 보여준다. 본 연구에서는 Kauko(2005)와 Bourassa et al. (1999)을 따라 근린의 경제적 수준과 관련 있고 우리나라의 통계 상황에서 구할 수 있는 변수들을 군집을 위해 고려했다. 본 연구에서 고려된 변수는 개인 소득과 밀접한 관계가 있는 센서스 구역별 교육수준, 단위면적당 공동주택가격, 재정자립도 등이다.³⁾ <표 2>는 본 연구에서 고려된

Kuoko(2005)	Bourassa et al.(1999)
Addresses per neighbourhood (density proxy)	Persons per square km
Extent of urbanisation	Dwellings per square km
Population density (inhabitants per km ²)	Percentage unemployed
Percentage of children 10-14 years old	Percentage in public housing
Percentage of people 15-24 years old	Percentage driving car to work
Percentage of people 25-44 years old	Average cars per household
Percentage of people 45-64 years old	Average bedrooms per house
Percentage of people 65 or over years old	Owner-occupation rate
Percentage of non-Westerners	Median household income(\$A)
Percentage of one-person households	Distance to CBD (km)
Number of families	Distance to subcenter (km)
Percentage of families with children	Distance to coast(km)
Average family size	
Percentage of people on a low income	
Percentage of people on a high income	
Percentage of people 15-65 years old with unemployment benefit as the primary source of income	
Assessed market value of dwelling	
Percentage of industrial enterprises	
Percentage of commercial enterprises	
Percentage of noncommercial enterprises	

표 2_ 군집분석을 위한 변수

내용	출처
2005년 대학원 이상 학력 비율(%)	2005년 인구 및 주택센서스
2005년 가구의 최종학력이 초등학교인 가구비율(%)	"
2005년 가구의 최종학력이 중학교인 가구비율(%)	"
2005년 가구의 최종학력이 고등학교인 가구비율(%)	"
2005년 가구의 최종학력이 대학·4년제 미만인 가구비율(%)	"
2005년 가구의 최종학력이 대학·4년제인 가구비율(%)	"
2005년 가구의 최종학력이 석사과정인 가구비율(%)	"
2005년 가구의 최종학력이 박사과정인 가구비율(%)	"
2005년 가구의 최종학력이 미취학인 가구비율(%)	"
2008년 재정자립도	재정고 (http://lofin.mopas.go.kr/)
2008년 11월 공동주택 평당 평균가격(만 원)	부동산 114
시청역까지 직선거리(m)	GIS 데이터
삼성역까지 직선거리(m)	"

3) 김강호(2009), 신경수·최창렬(2007), 김유선(2009) 등의 연구는 우리 사회에서 학력과 소득 사이에 강한 상관관계가 있음을 보여줌.

표 3_ 주요 선행연구에서의 사용변수

구분	고원용·김홍규·유완(2001)	정건섭·이상엽(2007)	김주영·우경(2004)	강창덕(2010)
종속변수	공동주택 매매가격	아파트가격	주택가격	아파트 실거래가격
주거특성	공동주택평수, 건축연도, 층수, 주거의향	평형, 방수, 층수, 동수	주택유형, 사용면적, 방수, 건축 경과년수, 부엌종류, 입식 여부, 화장실종류, 목욕탕, 온수, 화장 실사용 여부, 난방방식, 출입구	층, 면적
근린환경특성	세대수, 주차대수, 도심까지의 거리, 대중교통의 접근성(지하철), 근린공원, 산, 강	총세대수, 지하철과의 거리, 버스노선	전철(기차역)거리, 시청(구청)거리, 시장(대형슈퍼)거리, 백화점(도소매점)거리, 초등학교 거리, 중학교 거리, 1차 진료기관 거리, 종합병원 거리, 근린공원 거리, 직장과의 거리	지하철 역, 직선거리, 개발제한구역, 직선거리, 시청 직선거리, 핵심 상권, 직선거리, 공원 직선거리, 고등학교, 직선거리, 중학교 직선거리
지역환경특성	동별 인구밀도, 동별 1인당 공원면적, 동별 의료기관(병상) 수, 동별 시장/백화점 수	-	인구밀도, 오염물질 배출, 시설 수, 사업체 수	인구밀도, 고용밀도

13개 변수의 정의와 출처를 보여준다.

2) 헤도닉 함수 Chow Test를 위한 변수의 정의

본 연구의 분석을 위해 2008년 11월 수도권 공동주택가격 조사자료를 이용했다. 수도권 공동주택에 대한 가격조사 정보는 ‘부동산114(www.r114.co.kr)’에서 획득했다. 헤도닉 함수의 추정을 위해서는 설정의 오류를 피하기 위해 적절한 설명변수를 포함하는 것이 필요하다. 강창덕(2010), 정건섭·이상엽(2007), 김주영·우경(2004), 고원용·김홍규·유완(2001)의 선행연구에서 사용된 설명변수를 정리하면 <표 3>과 같다. 수도권 전체를 연구대상으로 하는 본 연구의 성격상 강창덕(2010), 김주영·우경(2004), 고원용·김홍규·유완(2001)과 같이 설문조사나 현지방문을 통해 얻을 수 있는 상세한 설명변수를 포함하기에는 어려움이 있어 상대적으로 간결한 모델을 추정한 정건섭·이상엽(2007)의 모델에 포함된 변수들을 참조하여 설명변수를 결정했다.

본 연구에서는 <표 4>와 같은 공동주택의 물리적 특성변수와 철도역 접근도, 서울 도심에 대한 접

표 4_ 분석을 위한 변수

변수명	정의
mp	2008년 11월 18일 조사된 단지별 평형별 최대값과 최소값의 평균(만 원)
size	크기(평)
age	연령(건축 이후 경과연도)
thh	단지 내 층가구 수
dcity	서울 도심 접근도(행정동 중심에서 서울 시청 역까지의 직선거리)(미터)
dstep	1. 계단형일 경우 0. 계단형이 아닌 경우
drail	1. 행정동에 철도역세권(역사 500m 이내)이 있을 경우 0. 행정동에 철도역세권(역사 500m 이내)이 없을 경우

근도를 나타내는 위치변수를 설명변수로 고려했다. 위치변수의 측정을 위해 공동주택이 위치한 행정동의 위치가 이용되었다. Chow Test는 <표 4>의 변수를 이용하여 <식 8>과 같은 세미로그 함수 형태의 헤도닉 함수를 추정한 후 <식 9>의 귀무가설을

검정하여 이루어졌다.

$$\begin{aligned} \ln(MP) = & \alpha + \beta^1 size + \beta^2 age + \beta^3 thh \\ & + \beta^4 dstep + \beta^5 dcity + \beta^6 drail \\ & + \alpha' d + \beta^1 d* size + \beta^2 d* age + \beta^3 d* thh \\ & + \beta^4 d* dstep + \beta^5 d* dcity + \beta^6 d* drail + u \end{aligned}$$

<식 8>

d: 2개의 비교대상 시·군·구 중 1개 시·군·구를 의미하는 더미변수

$$H_0 : \alpha' = \beta^1 = \beta^2 = \beta^3 = \beta^4 = \beta^5 = \beta^6 = 0$$

<식 9>

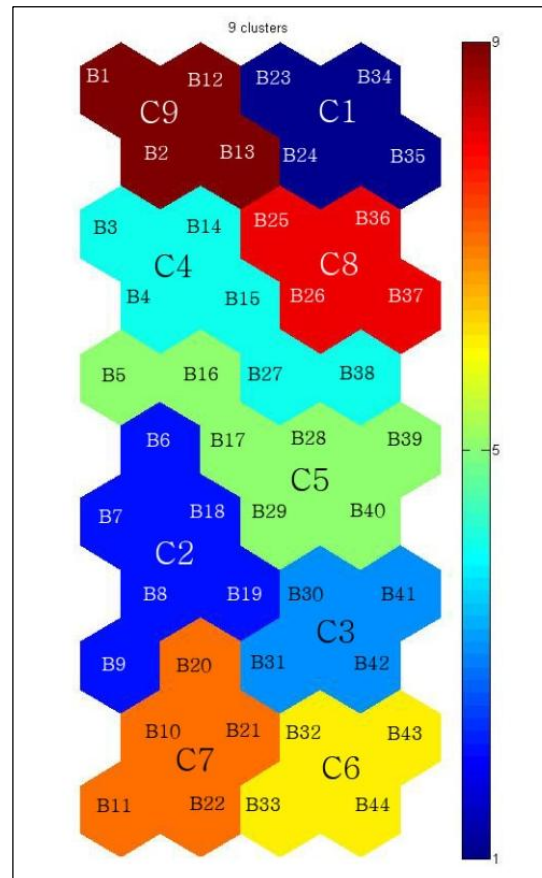
IV. 분석결과

1. SOM 분석결과

수도권 78개 시·군·구별 13개 사회·경제적 지표로 구성된 데이터를 이용하여 4x11개의 육각형 노드로 구성된 SOM에 대한 훈련(Training)이 이루어졌다.⁴⁾ 그리고 노드들에 대한 최종적 가중치 벡터(Weight Vector)가 얻어졌다. 그 다음 78개 시·군·구의 데이터 벡터를 44개 노드의 최종 가중치 벡터에 대입하여 가장 유사한 속성을 가진 노드(Winning Node)를 찾아낼 수 있었다. 그 결과 얻어진 것이 <그림 1>의 B1-B44 군집이다. 본 연구에서는 B1-B44를 중분류 군집으로 정의한다. 그리고 B1-B44를 다시 군집화하여 대분류 군집을 형성했다. 44개 노드를 이용

하여 대분류 군집을 형성하기 위해 수도권 대분류 군집의 수를 1에서 10까지 설정하고 노드의 가중치 벡터를 이용하여 각 군집 수별 500번의 K-means 군집을 반복 수행했다. 그리고 각각의 군집 수에 대해 분류오류를 최소화하는 군집에 대해 Davies-Bouldin index를 계산한 결과 군집 수가 9개일 때 Davies-Bouldin index가 가장 최소화되었다.⁵⁾ <그림 1>에서 최종적으로 얻어진 C1-C9 등 9개 대분류 군집의

그림 1_SOM 결과 I



4) SOM의 추정에는 Matlab SOM Toolbox(URL: <http://www.cis.hut.fi/projects/somtoolbox/>)를 이용하여 이루어졌음. 본 연구에서는 SOM의 훈련(training)을 위해 Matlab SOM Toolbox의 명령어인 som_make()를 이용했고 격자망의 차원에 대한 결정은 동 명령어의 기본환경에 의해 이루어졌음. 훈련에 이용된 13개 변수는 투입 전 0과 1 사이 값으로 변환되었음.

5) 분류오류는 다음과 같은 함수에 의해 계산됨(Vesanto and Alhoniemi, 2000).

$$E = \sum_{k=1}^C \sum_{x \in Q_k} \| X - c_k \|^2,$$

X: 데이터 벡터, C: 클러스터의 수, Q_k : k번째 클러스터, c_k : cluster k의 중심

그림 2_SOM 군집결과 II

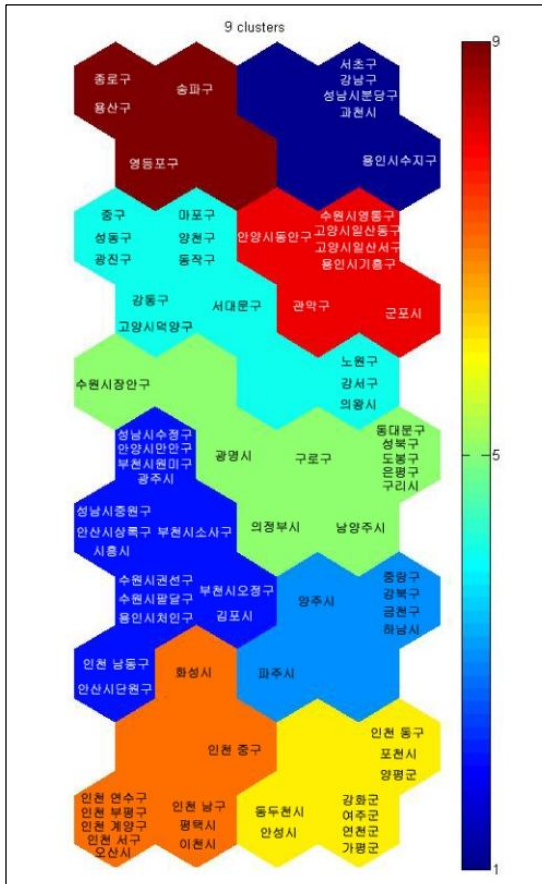
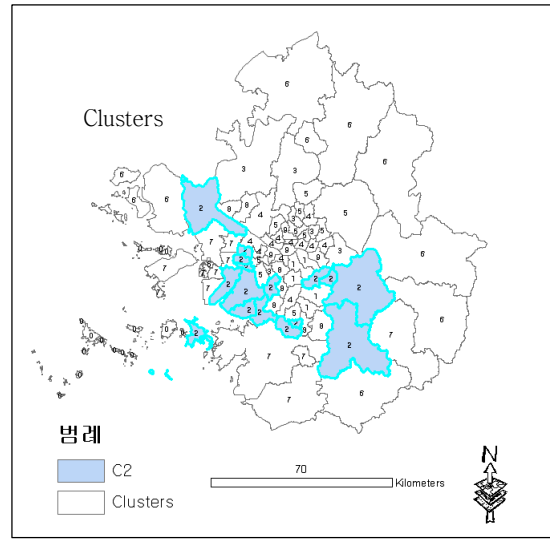


그림 3_대분류 군집의 분포



적으로도 인접한 것을 알 수 있다. 이 점은 유사한 사회·경제적 속성을 가진 시·군·구가 공간적으로도 군집해 있음을 의미한다.

2. Chow Test 결과⁶⁾

본 연구에서는 연구설계 2단계를 수행하기 위해 사회·경제적 지표를 통해 구분된 C1-C9의 대분류 중 경기 및 인천지역 시·군·구로 구성된 C2를 선택했다. <그림 3>은 지도상에서 C2의 위치를 보여준다.

대분류 C2에는 중분류 B6(성남시 수정구, 안양시 만안구, 부천시 원미구, 광주시), B7(성남시 중원구, 안산시 상록구, 시흥시), B8(수원시 권선구, 수원시 팔달구, 용인시 처인구), B9(인천시 남동구, 안산시 단원구), B18(부천시 소사구), B19(부천시 오정구, 김포시) 등 15개 시·군·구가 포함되어 있다. 이 가운데 성남시 수정구, 성남시 중원구, 광주시, 용인시 처인구가 수도권 남동부 지역에 군집해 있으며 나머지

모습을 볼 수 있다.

<그림 2>는 B1-B44, C1-C9 군집에 소속된 시·군·구의 모습을 보여준다. <그림 2>를 살펴보면 SOM의 위상단에 주로 서울 및 경기권에서 주거지역으로 가장 선호되는 시·군·구가 군집해 있는 것을 볼 수 있다. 그리고 좌상단에는 서울의 나머지 자치구가 입지해 있으며 우하단에는 수도권의 접경지역 및 외곽에 위치한 시·군·구가 위치했다.

<그림 3>은 SOM의 9개 대분류 군집을 지도상에 표시한 것이다. <그림 3>에서 직관적으로 파악할 수 있듯이 동일한 군집에 속한 시·군·구가 공간

6) Chow Test를 위해 <식 8>형태의 헤도닉 함수가 추정되었고 추정계수의 분산은 이분산에 강건한 Huber-White-Sandwich estimator를 이용하여 추정되었음(Baum, 2006).

표 5_14개 시·군·구의 분석 샘플 수

시·군·구	Freq.	Percent
인천시 남동구	490	12.44
수원시 권선구	278	7.06
수원시 팔달구	189	4.8
성남시 수정구	137	3.48
성남시 중원구	156	3.96
안양시 만안구	270	6.85
부천시 원미구	495	12.57
부천시 소사구	268	6.8
부천시 오정구	250	6.35
안산시 상록구	239	6.07
안산시 단원구	318	8.07
시흥시	569	14.45
용인시 처인구	103	2.61
광주시	177	4.49
Total	3,939	100

지역이 수도권 남서부와 북서부 지역에 군집해 있다.

본 연구에서는 C2에 속한 시·군·구가 유사한 사회·경제적 속성을 공유하는 시·군·구로 수도권 거주자의 주거선택 시 긴밀한 대체관계에 있다고 가정한다. 그리고 C2에 속하는 시·군·구 가운데 직접적으로 행정경계를 접하고 있는 시·군·구들에 대해 <식 8>의 헤도닉 함수를 추정하고 <식 9>의 가설을 검정했다.

그런데 C2에 속하는 시·군·구 가운데 김포시는 직접적으로 행정경계를 접하는 시·군·구가 없어 분석에서 제외했다. <표 5>와 <표 6>은 분석에 사용된 각 시·군·구별 샘플의 수와 변수별 통계량을 보여준다.

<표 7>은 C2에 속하는 시·군·구 가운데 행정경계를 접하는 시·군·구가 없는 김포시를 제외한 14개 시·군·구에 대해 직접적으로 인접한 시·군·구와 Chow Test를 실시한 결과를 보여준다. 분석결과는

표 6_사용변수의 통계량

Variable	Obs	Mean	Min	Max
mp	3939	24423	3750	260000
size	3939	28.996	10	82
age	3939	12.74	0	30
thh	3939	495.87	35	5076
dcity	3939	23814	13393.69	41547.25
dstep	3939	0.79	0	1
drail	3939	0.47	0	1

모든 경우에 회귀계수의 안정성이 기각되고 있음을 보여준다. 이 결과는 각각의 시·군·구가 모두 개별적인 주택하위시장을 형성할 가능성을 시사한다. 중분류 B7에 속하면서도 행정경계를 접하고 있는 안산시 상록구와 시흥시 간, 중분류 B8에 속하면서도 경계를 접하고 있는 수원시 팔달구와 수원시 권선구 간의 Chow Test에서도 회귀계수의 안정성이 기각된 것은 근린의 질이나 지리적 근접성 면에서 긴밀한 대체관계에 있어 보이는 시·군·구들에서조차 헤도닉 함수 추정 시 데이터를 통합하여 사용할 경우 어떤 지역의 특성도 반영하지 못하는 평균화된 주택특성의 잠재가격 추정치를 얻을 가능성이 있음을 말해준다.

<표 7>에 정리된 결과는 비서울 지역 1개의 군집에서만 얻어진 한정된 결과다. 그러나 비서울 지역에서 시·군·구별로 분화된 주택하위시장이 존재할 가능성이 있다는 것을 보여준다. 따라서 선행연구 가운데 광역적인 공간적 하위시장의 가능성을 보여주는 김주영·우경(2004)의 연구결과는 지나치게 단순화된 것이며 강창덕(2010), 정건섭·이상엽(2007) 등의 연구결과가 비록 서울에 한정된 것일지라도 수도권의 세분화된 주택하위시장의 모습을 보다 정확하게 보여주는 것이라 판단할 수 있다.

표 7 _ Chow Test 결과

비교대상	통계량
성남시 수정구: 성남시 중원구	F(7, 279) = 2.99 Prob > F = 0.0048
성남시 중원구: 광주시*	F(6, 321) = 5.55 Prob > F = 0.0000
광주시: 용인구 처인구*	F(6, 268) = 2.47 Prob > F = 0.0244
수원시 팔달구: 수원시 권선구	F(7, 453) = 29.41 Prob > F = 0.0000
수원시 권선구: 안산시 상록구	F(7, 503) = 33.65 Prob > F = 0.0000
안산시 상록구: 안산시 단원구	F(7, 543) = 6.49 Prob > F = 0.0000
안산시 상록구: 안양시 만안구	F(7, 495) = 17.85 Prob > F = 0.0000
안양시 만안구: 시흥시	F(7, 825) = 40.37 Prob > F = 0.0000
안산시 상록구: 시흥시	F(7, 794) = 15.64 Prob > F = 0.0000
안산시 단원구: 시흥시	F(7, 873) = 19.42 Prob > F = 0.0000
시흥시: 부천시 소사구	F(7, 823) = 4.97 Prob > F = 0.0000
부천시 소사구: 부천시 원미구	F(7, 749) = 20.28 Prob > F = 0.0000
부천시 원미구: 부천시 오정구*	F(6, 733) = 16.76 Prob > F = 0.0000
시흥시: 인천시 남동구	F(7, 1045) = 14.06 Prob > F = 0.0000
인천남동구: 부천시 소사구	F(7, 744) = 9.41 Prob > F = 0.0000

주: *는 샘플 모두에서 $d_{rail}=0$

V. 결론

본 연구에서는 수도권 지역의 비서울 지역에 존재하는 공동주택 하위시장의 존재와 형태에 대해 분석해보았다. 분석결과 서울 지역에서와 마찬가지로 비서울 지역에서도 개별 시·군·구를 단위로 한 공간적 하위시장이 존재할 가능성이 있음이 나타났다.

본 연구의 분석결과와 강창덕(2010), 정건섭·이상

엽(2007) 등의 연구결과는 수도권에서 시·군·구나 그 이하의 공간적 단위를 기본단위로 한 세분화된 형태의 공동주택 하위시장이 존재할 가능성이 크다는 점을 보여준다. 이러한 결과가 수도권에 대한 하위시장 분석에 주는 시사점은 수도권에 공간적으로 매우 세분화된 주택하위시장이 존재할 가능성이 크며, 수도권 전체를 대상으로 한 하위시장 분석을 실시할 경우 하위시장의 존재로 인한 추정계수의 공간적 변화를 보다 본질적으로 보는 인식의 전환이 필요하다는 점이다. 그리고 이러한 인식의 전환을 반영하여 추정계수의 공간적 변화를 명시적으로 허용하는 공간확장방법(Spatial Expansion Method)이나 지리가중회귀분석(Geographically Weighted Regression)의 사용을 수도권에 대한 하위시장 분석에 적극적으로 고려해볼 필요가 있다는 점이다.

참고문헌

- 강창덕. 2010. "GWR 접근법을 활용한 부동산 감정평가 모형 연구: 서울시 아파트를 사례로". 부동산연구 제20권 제2호. pp107-132.
- 고원용·김홍규·유환. 2001. "서울시 주택 하위시장별 주거환경이 공동주택가격에 미치는 영향". 지역연구 제17권 제2호. pp9-27.
- 김강호. 2009. "학력과 직업훈련 참여가 임금에 미치는 효과". 농업교육과 인적자원개발 제41권 제3호. pp123-151.
- 김유선. 2009. "한국 노동시장의 임금결정요인-OLS 회귀분석과 분위회귀분석". 산업관계연구 제19권 제2호. pp1-25.
- 김주영·우경. 2004. "수도권 주택하위시장 분석에 관한 연구". 국토연구 제41권. pp101-112.
- 신경수·최창렬. 2007. "학력별 임금격차와 성별 생산성격차 분석". 생산성논집 제21권 제3호. pp97-121.
- 정건섭·이상엽. 2007. "주택하위시장 구분방법과 정책적 시사점". 정책분석평가학회보 제17권 제1호. pp193-216.
- Agarwal, Pragma and Andre Skupin. 2008. *Self Organizing Maps: Applications in Geographic Information Science*. West Sussex,

- England : Wiley.
- Alkay, Elif. 2008. "Housing Submarkets in Istanbul". *International Real Estate Review* vol.11, no.1. pp113-127.
- Baum, Christopher F. 2006. *An Introduction to Modern Econometrics Using Stata*. College Station, Texas : Stata Press.
- Bitter, Christopher, Gordon F. Mulligan, and Sandy Dall'erba. 2007. "Incorporating Spatial Variation in Housing Attribute Prices: a Comparison of Geographically Weighted Regression and the Spatial Expansion Method". *Journal of Geographical Systems* vol.9. pp7-27.
- Bourassa, Steven C., Foort Hamelink, Martin Hoesli and Bryan D. MacGregor. 1999. "Defining Housing Submarkets". *Journal of Housing Economics* vol.8. pp160-183.
- Bourassa, Steven C., Martin Hoesli, and Vincent S. Peng. 2003. "Do Housing Submarkets Really Matter?". *Journal of Housing Economics* vol.12. pp12-28.
- Budayan, Cenk, Irem Dikmen, M. Talat Birgonul. 2009. "Comparing the Performance of Traditional Cluster Analysis, Self-Organizing Maps and Fuzzy C-means Method for Strategic Grouping". *Expert Systems with Applications* vol.36. pp11772-11781.
- Can, Ayse. 1992. "Specification and Estimation of Hedonic Housing Price Models". *Regional Science and Urban Economics* vol.22. pp453-474.
- Cassetti, E. 1972. "Generating Models by the Expansion Method: Applications to Geographical Research". *Geographical Analysis* vol.4. pp81-92.
- Fotheringham, A. Stewart, Chris Brunsdon, Martin Charlton. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. West Sussex, England : Wiley.
- Goodman, A. C. and T. G. Thibodeau. 1998. "Housing Market Segmentation". *Journal of Housing Economics* vol.7. pp121-143.
- _____. 2007. "The Spatial Proximity of Metropolitan Area Housing Submarkets". *Real Estate Economics* vol.35, no.2. pp209-232.
- Islam, Kazi Saiful and Yasushi Asami. 2009. "Housing Market Segmentation: A Review". *Review of Urban and Regional Development Studies* vol.21, no.2/3. pp93-109.
- Kauko, Tom. 2005. "Using the Self-organising Map to Identify Regularities across Country-specific Housing-market Contexts". *Environment and Planning B: Planning and Design* vol.32. pp89-110.
- Lu, Hsin-Chung, Chung-Liang Chang and Jen-Chieh Hsieh. 2006. "Classification of PM10 Distribution in Taiwan". *Atmospheric Environment* vol.40. pp1452-1463.
- Michaels, R. Gregory and V. Kerry Smith. 1990. "Market Segmentation and Valuing Amenities with Hedonic Models: The Case of Hazardous Waste Sites". *Journal of Urban Economics* vol.28. pp223-242.
- Schnare, A. B. and R. J. Struyk. 1976. "Segmentation in Urban Housing Markets". *Journal of Urban Economics* vol.3. pp146-166.
- Song, Yan and Roberto G. Quercia. 2008. "How are Neighbourhood Design Features Valued across Different Neighbourhood Types?". *Journal of Housing and the Built Environment* vol.23. pp297-316.
- Vesanto, Juha and Esa Alhoniemi. 2000. "Clustering of the Self Organizing Map". *IEEE Transactions on Neural Networks* vol.11. no.3. pp586-600.
- Watkins, Craig A. 2001. "The Definition and Identification of Housing Submarkets". *Environment and Planning A* vol.33. pp2235-2253.
- Yao, Xiaobai. 2007. "Where are Public Transit Needed - Examining Potential Demand for Public Transit for Commuting Trips". *Computers, Environment and Urban Systems* vol.31. pp535-550.

-
- 논문 접수일: 2011. 7. 7
 - 심사 시작일: 2011. 7.19
 - 심사 완료일: 2011. 8.31

Identification of Spatial Housing Submarkets in the Capital Region Using Self Organizing Map and Chow Test

Keywords: Spatial Housing Submarket, Self Organizing Map, Chow Test

In this study, the existence of spatial housing submarkets in the Capital Region was tested using Self Organizing Map and Chow Test. The Self Organizing Map was used to form 9 large sized area-wide socio-economically homogeneous regional clusters and 44 middle sized area-wide socio-economically homogeneous regional clusters. To form the clusters, 13 socio-economic indicators of the 78 spatial administrative units in the Capital Region were used. One of the 9 large sized regional clusters, which includes non-Seoul spatial administrative units, was selected and, then, the stability of the coefficients of the hedonic housing price functions across the adjoining spatial administrative units in the cluster was tested using Chow Tests. The results of these tests showed that there exist spatially very disaggregated housing submarkets in this area. The results imply that we need to see the housing market of the Capital Region as intrinsically spatially disaggregated when we estimate hedonic housing price functions for this area and one should consider using Spatial Expansion Method or Geographically Weighted Regression which allows the spatial changes of the parameter values to conduct hedonic analyses about the Capital Region.

수도권의 공간적 주택하위시장 식별에 대한 연구

주제어: 공간적 주택하위시장, 자기조직화 지도, 회귀계수 안정성 검증

본 연구에서는 수도권 내 공간적 주택하위시장의 존재와 형태를 Neural Network의 한 유형인 자기조직화 지도(Self Organizing Map)와 Chow Test를 이용하여 분석하였다. SOM은 수도권 내 78개 시·군·구 지역을 사회·경제적 특성이 유사한 9개 대분류, 44개 중분류로 군집화하는 데 사용되었다. 군집화를 위해 78개 시·군·구에 대한 13개 사회·경제적 지표가 사용되었다. 9개 대분류 지역 중 비수도권에 위치한 1개 대분류 지역을 택하여 소속된 인접 시·군·구에 대해 공동주택 헤도닉 모델을 추정하고 추정된 함수 간의 계수의 동일함에 대한 가설이 Chow Test를 통해 검증되었다.

분석결과는 수도권 비서울 지역에 시·군·구를 단위로 하는 매우 개별화된 공간적 주택하위시장이 존재할 가능성이 높음을 보여주고 있다. 이러한 결과가 수도권에 대상으로 하는 헤도닉 분석에 주는 시사점은 수도권 전체를 대상으로 한 헤도닉 분석을 실시할 경우 하위시장의 존재로 인한 추정계수의 공간적 변화를 보다 본질적으로 보는 인식의 전환이 필요하다는 점이다. 그리고 이러한 인식의 전환을 반영하여 추정계수의 공간적 변화를 명시적으로 허용하는 공간확장방법(Spatial Expansion Method)이나 지리가중회귀분석(Geographically Weighted Regression)의 사용을 수도권에 대한 헤도닉 분석에 적극적으로 고려해볼 필요가 있다는 점이다.