

수치지도상의 공동주택위치 표현방법이 헤도닉 함수추정에 미치는 영향

손철, 강릉대학교 지역개발학과

K C I

---

Abstract

**Evaluating the Influences of Geocoding Methods on the Parameter Estimates of a Hedonic Apartment Price Function**

Sohn, Chul

Department of Regional Development  
Kangnung National University

This paper investigates the influences of two different types of geocoding methods on the size of a hedonic parameter estimate related with subway station accessibility by using Seoul's apartment housing market data. Results show that the size of the parameter estimate indeed varies around 35% to 53% between two different geocoding methods. Results suggest that readers of housing market studies using hedonic apartment price functions should pay attention to geocoding methods used in the studies of interest when she or he wants to compare the results of the studies mutually.

Keywords: Hedonic Function, GIS Apartment Market, Geocoding (헤도닉 함수, 지리정보시스템, 공동주택시장, 지오코딩)

---

## I. 서론

헤도닉 함수(Hedonic Function)는 1980년대 들어서부터 우리나라의 도시 및 지역 분야의 경험적 연구를 위해 사용되고 있다. 현재는 주택시장의 가격지표 작성, 환경 자원의 경제적 가치분석, 도시계획 및 교통계획에서의 프로젝트 영향분석, 대단위 부동산 감정평가 등 다양한 분야에서 이용되고 있다.

주택의 가격과 속성들 간에 존재하는 단순한 통계적 상관관계를 드러내 주는 헤도닉 함수가 다양한 분야에서 활발하게 사용되고 있는 것은 일면 단순해 보이는 헤도닉 함수가 일반인들이 축적한 부의 대부분을 차지하는 재화인 주택이 거래되는 주택시장에 대한 유용한 정보를 제공해주기 때문이다.

헤도닉 함수의 추정을 통해 얻을 수 있는 가장 중요한 주택시장 정보는 주택을 구성하는 속성들에 대해 개인들이 지불하는 비용에 대한 추정액, 즉 한계잠재가격(Marginal Implicit Price)이다. 헤도닉 함수의 추정을 통해 시장에서 거래되는 주택의 방 1개, 지하철로부터의 거리 1미터 당 개인들이 지불하고 있는 비용(한계잠재가격)을 통계적으로 추측할 수 있으며, 이 정보는 주택시장의 가격지표 작성, 환경 자원의 가치추정, 프로젝트 영향분석, 대단위 감정평가 등에서 유용하게 이용된다.

헤도닉 함수의 추정을 위해서는 주택가격 및 속성에 대한 정보와 주택이 위치한 근린에 대한 다양한 정보가 필요하다. 우리나라의 경우는 헤도닉 함수의 추정을 위해 부동산뱅크, 부동산 114 등 부동산정보업체가 제공하는 공동주택 가격 및 속성정보가 가장 일반적으로 이용되고 있다. 또한 다양한 근린속성에 대한 정보를 획득하기 위해서는 최근 국가지리정보시스템(NGIS) 사업을 통해 구축된 다양한 축척의 수치지도를 적극 이용하는 추세이다.

## II. 문제제기

헤도닉 함수 추정과정에서 부동산정보업체가 제공하는 공동주택 속성자료와 공동주택 및 기타 지리적 객체의 위치를 드러내 주는 수치지도를 지리정보시스템 환경에서 결합하여 이용함으로써 연구자는 공동주택의 가격에 영향을 미치

는 다양한 근린속성과 공동주택간의 공간적 상호관계를 손쉽게 파악할 수 있다. 또한 파악된 공간적 상호관계를 양적으로 측정하기 위해 다양한 공간분석 기능(버퍼링, 네트워크분석 등)을 이용할 수 있다. 이상과 같은 부동산정보업체 데이터, 수치지도, 지리정보시스템의 결합이 주는 시너지효과는 실제로 이들을 이용하여 추정된 헤도닉 함수의 추정결과를 통계적 측면에서 개선시키는 효과가 있는 것으로 보고되고 있다(Rodriguez, Sirmans and Marks, 1995).

<표 II-1> 부동산정보업체 데이터, 수치지도, GIS를 결합하여 이용한 연구

연구	연구대상	부동산 데이터	수치지도
Bae, Jun, and Park(2003)	공동주택	지하철 5호선 주변지역의 공동주택(부동산뱅크 자료)	불분명(서울시의 GIS 데이터라고만 언급됨)
박헌수, 정수연(2002)	토지	강남구의 471개 주거용 필지(공시지가자료)	강남구 필지 지적도
박헌수(2001)	공동주택	한강이남 11개구에서 2개 이상의 건물을 가지고 있는 공동주택단지(부동산뱅크 자료)	1:5000 수치지도
서경천, 이성호(2001)	토지	부산시 서면, 남포동 지역 1547개 필지(공시지가자료)	1:3000 지번도
고원용, 김홍규, 유완(2001)	공동주택	서울전역의 200세대 이상 및 동수가 2개 이상인 단지 중 서울에 위치한 358개 단지(부동산뱅크 자료)	1:3000 수치지도

우리나라의 경우도 이상의 이점 때문에 헤도닉 함수추정을 위해 실제로 부동산정보업체 데이터, 수치지도, 지리정보시스템(GIS)을 결합하여 이용한 연구들이 활발하게 보고되고 있다. <표 II-1>은 1995년 이후 도시 및 지역연구분야의 주요 학술지에 실린 글 가운데 헤도닉 함수추정을 위해 수치지도, 지리정보시스템을 사용한 연구들을 열거한 것이다. <표 II-1>에 열거된 연구 중 공동주택을 대상으로 한 연구는 모두 부동산정보업체(부동산뱅크)가 제공한 정보를 이용하고 있으며, 공동주택 위치정보의 파악을 위해서는 1:5000 혹은 1:3000 수치지도를 이용하고 있다.<sup>1)</sup> 이들 연구는 공통적으로 수치지도상에서 공동주택의

---

위치와 주요 근린속성(지하철역, 녹지 등)의 위치를 파악한 후 이들 간의 직선 혹은 네트워크 거리를 측정하는 데 지리정보시스템을 이용하고 있다.

그런데 이상의 연구들은 공동주택의 위치를 수치지도상에 표현하면서 개별 공동주택(공동주택단지의 개별 동)의 위치를 파악하는 대신 별도의 작업을 통해 1개 이상의 동으로 구성된 각 공동주택단지를 한 개의 포인트 객체로 수치지도(GIS 레이어)상에 표현하는 방법을 이용하고 있다. 따라서 헤도닉 함수의 추정에 이용되는 공동주택으로부터 여러 근린특성까지의 거리도 앞서 선정된 하나의 대표적인 점으로부터 측정되고 있다.

거리측정을 위한 이러한 접근은 부동산정보업체에서 제공하는 공동주택의 가격 및 속성정보가 개별 공동주택단위로 식별가능하지 않기 때문에 부득이한 것이다. 즉, 개별 공동주택(1개 동)의 위치는 각각의 동이 1:5000 수치지도나 1:5000보다 큰 대축척 수치지도상(예: 1:1000 수치지도)에서는 폴리곤으로 표시되기 때문에 식별가능하다. 그러나 부동산업체가 제공하는 공동주택의 가격과 속성정보는 단지단위로 제공되기 때문에 공동주택에 관한 위치와 가격과 속성정보를 지리정보시스템상에서 결합하여 사용하기 위해서는 개별 동을 나타내는 폴리곤을 이용하는 대신 공동주택 단지를 대표하는 하나의 포인트를 생성하여 이용할 수밖에 없는 것이다.

이러한 접근방법의 부득이함은 박현수(2001)의 다음과 같은 언급에도 잘 드러나 있다.

*“...국가수치지형도(1:5000)를 활용하여 각 아파트 단지를 공간상의 (x,y) 좌표로 표시...GIS를 이용하여 접근성에 대한 자료를 구하였다.(p. 159)”*

연구과정에서 각각의 공동주택단지가 하나의 포인트로 표시되는 것의 불가피성은 인정한다고 해도 이런 방법을 통해 추정된 헤도닉 함수를 여러 연구간

---

1) 현재 헤도닉 함수의 추정을 위해 1:1000 수치지도를 사용한 연구는 찾아 볼 수 없다. 그 이유는 헤도닉 함수의 추정을 위해서는 1:5000 수치지도의 상세함으로도 충분하기 때문이다. 그러나 본 연구에서의 수치지도 사용에 대한 논의는 1:1000 수치지도에도 그대로 적용된다.

---

에 비교할 경우 발생될 문제에 대해서는 주의할 필요가 있다. 일반적으로 우리나라에서의 헤도닉 함수 추정시 기본도 역할을 하는 1:5000 수치지도의 경우 공동주택단지의 단지별 경계를 나타내는 폴리곤은 존재하지 않으며, 단지내 존재하는 각각의 동이 각각 하나의 폴리곤으로 표현되고 있다. 이 밖에도 아파트 단지의 이름이 주기로 표현되어 있다. 이러한 상황에서 각 단지를 대표적인 하나의 포인트 객체로 표기하기 위해서는 여러 가지 방법을 취할 수 있다.

첫번째로, 각 단지에 포함되는 각 동 폴리곤의 중심점을 포인트로 표기한 후 단지내에 존재하는 포인트 중 임의의 하나를 단지를 대표하는 포인트로 선정하는 방법이다. 두번째로, 수도권 및 지방의 주요도시에 대해 존재하는 지번도나 도로교통지도를 이용하여 각 공동주택단지를 나타내는 폴리곤을 형성한 후 각 단지 폴리곤의 중심점을 단지를 대표하는 포인트로 선택하는 방법이다. 세번째로, 수치지도상에 존재하는 주기(공동주택 이름)의 위치를 포인트로 전환하여 이용하는 방법이다. <표 II-1>에서 열거한 연구 중 공동주택을 대상으로 한 연구들이 앞서의 방법 중 어떤 방법을 이용하였는지 명확하지 않다. 각 연구에 이 부분에 대한 서술이 없기 때문이다.

이점은 각 연구의 결과를 서로 명확히 비교할 수 없게 만드는 주요원인이 된다. 그것은 상이한 방법을 이용하는 연구들 간에는 측정된 위치변수(지하철로부터의 거리 등)의 변량이 비록 동일한 공동주택단지에 대한 데이터를 이용한다고 해도 다를 것이며, 그 결과 다중회귀식인 헤도닉 함수의 속성상 상이한 공동주택 위치표현 방법을 이용하는 연구들 간에 위치변수의 모수추정치가 달라질 것이기 때문이다.<sup>2)</sup>

일반적으로 헤도닉 함수를 이용한 연구에서 동일한 부동산 데이터를 이용하였음에도 불구하고 추정결과 얻어진 여러 헤도닉 함수의 모수추정치의 크기에 차이가 존재할 경우, 그 원인으로는 사용된 함수식(Function Form), 변수 수

---

2) 이 경우 위치변수의 모수추정치뿐만 아니라 비위치 변수의 모수추정치 역시 상이한 위치표현방법에 따라 달라질 가능성 역시 존재한다. 그러나 그 가능성은 위치표현 방법의 상이성 여부 이외에도 설명변수들 간에 존재하는 상관관계 등 다양한 요인에 의존하므로 본 연구에서는 논하지 않기로 한다.

---

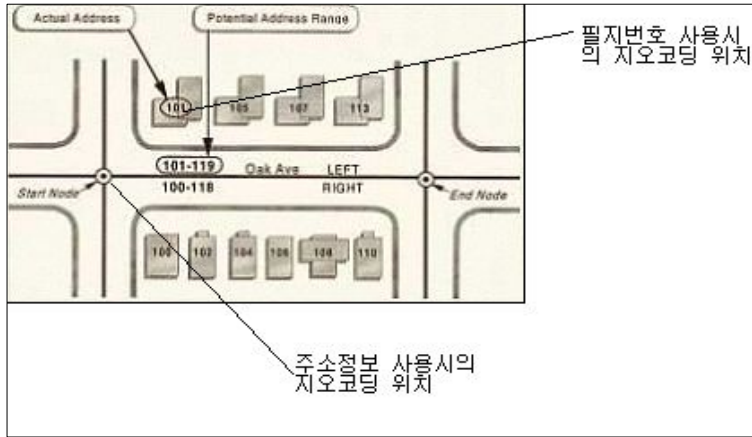
및 측정방법에서의 차이 등이 열거되곤 한다. 그러나 이처럼 우리나라의 상황에서 공동주택시장에 대한 헤도닉 함수추정 결과가 수치지도상의 공동주택 위치표시 방법에 따라 차이가 나면서도 연구자들 사이에 그 차이의 심각도가 적절하게 인식되지 못한다면 동일한 데이터로부터 추정된 헤도닉함수들의 여러 차이점들과 그것의 발생원인이 올바르게 파악되지 못할 가능성이 매우 크다. 따라서 헤도닉 함수의 추정결과가 바르게 해석되지 못할 가능성 또한 매우 크다. 이러한 문제인식에서 본 연구에서는 수치지도상의 공동주택 위치표현 방법 차이에 따른 헤도닉 함수 위치변수 모수추정결과 변화의 정도를 경험적으로 평가하여 제공함으로써 공동주택 위치표현 방법의 중요성에 대한 주의를 환기시키고자 한다.

### III. 관련연구의 검토

앞서 지적한 공동주택단지의 수치지도상의 위치표현 문제는 결국 현실세계에 존재하는 객체들의 위치를 어떠한 기준을 가지고 지리정보시스템내의 디지털 공간안에 표현시킬 것인가 하는 문제와 관련되어 있다. 현실에 존재하는 객체의 속성정보를 이용하여 이를 지리정보시스템 내의 디지털 공간의 특정위치에 표현하는 과정을 일반적으로 지오코딩(Geocoding)이라 한다(<http://www.ncjrs.org/html/nij/mapping>).

대부분의 상업용 GIS 소프트웨어가 지오코딩을 지원하고 있다. 지오코딩을 위해서는 일반적으로 2개의 정보가 필요한데 하나는 현실 객체가 가지고 있는 위치에 대한 속성정보(주소, 고유번호, 우편번호, 센서스 구역번호 등)이며, 나머지 하나는 현실 객체가 가진 위치속성을 포함하고 있는 GIS 레이어 혹은 수치지도이다. 후자에 속하는 것으로는 TIGER(Topologically Integrated Geographic Encoding and Referencing) 파일 또는 지적도(Parcel Map)등이 있다. TIGER 파일에는 주소범위에 대한 정보가 입력되어 있으며, 지적도에는 필지의 고유번호 또는 필지의 주소 등이 입력되어 있다.

<그림 III-1> 지오코딩의 예



출처 : [http://www.ncjrs.org/html/nij/mapping/ch4\\_3.html](http://www.ncjrs.org/html/nij/mapping/ch4_3.html)

상용 GIS 소프트웨어가 지오코딩을 수행하는 방식은 <그림 1>을 통해서 설명된다. <그림 III-1>은 “101 Oak Ave”라는 주소를 가진 집을 주소정보를 이용하여 지오코딩하는 경우를 나타낸다. 상용 GIS 소프트웨어가 주소정보와 TIGER 파일을 비교하여 위 주소를 가진 집을 지오코딩할 경우, 해당 소프트웨어는 집의 주소를 이용하여 TIGER 파일내에서 해당 주소가 포함되어 있을 도로구간을 검색한다. 물론 이러한 검색은 TIGER 파일내에 주소의 범위에 대한 정보가 입력되어 있기 때문에 가능하다. GIS 소프트웨어가 일단 해당 주소가 포함되어 있을 도로구간을 찾아내면, 주택의 위치는 해당 도로구간이 가지고 있는 주소범위와 집의 주소를 고려하여 적당한 위치에 내삽된다. <그림 III-1>의 경우는 해당 도로구간의 주소범위가 101-119 이므로 “101 Oak Ave”는 도로구간의 시작점에 그 위치가 표시된다. 그러나 상용 GIS 소프트웨어가 주소정보와 주소정보를 포함한 필지 혹은 건물지도를 이용하여 지오코딩을 수행한다면 “101 Oak Ave”는 도로상의 어떤 지점이 아닌 필지나 건물 폴리곤의 중앙에 그 위치가 표현된다.

<그림 III-1>에서 볼 수 있는 것처럼 지오코딩과 관련되어 주목할 점은 지오코딩 시 어떤 정보를 사용하느냐에 따라 동일한 객체라도 지리정보시스템의 디지털 공간 안에 표현되는 위치가 다를 수 있다는 점이다. 이러한 특징이 주는 시사점은 연구자가 사용하는 지오코딩 방법에 따라 지오코딩된 객체를 이용한 지리정보시스템의

---

분석결과가 다를 수 있다는 점이다.

지리정보시스템을 이용한 공간분석이나 경험적 연구에서의 분석결과가 연구에 사용되는 공간적 객체를 GIS 레이어나 수치지도상에 나타내는 방법에 따라 달라지는 문제점은 Ratcliffe(2001), 손철(2001, 2002)의 연구에 의해 지적되었다. Ratcliffe(2001)는 Sydney, Australia의 20,000개의 필지에 대한 주소를 TIGER 타입의 도로망도와 도로망도에 입력된 주소범위(Address Range)를 이용하여 지오코딩한 결과 약 5%-7.5%의 주소가 실제 주소가 소속되지 않은 센서스 구역(Census Tract)에 소속되는 것처럼 지오코딩되었으며, 50% 이상의 주소가 실제 주소가 부여된 필지 폴리곤 경계밖의 X,Y좌표를 부여받은 결과를 보고하고 있다.

손철(2001, 2002)은 Ratcliffe(2001)가 지적한 바와 같은 결과가 헤도닉 함수에 미치는 결과를 분석하였다. 손철(2001)의 연구는 미국 Farmers Branch, Texas의 주택 시장 데이터에 대해 서로 다른 지오코딩 방법을 적용한 결과가 추정된 헤도닉 함수로부터 구한 주택가격 상승률 계산에 미치는 영향을 분석하였다. 연구결과는 각기 다른 지오코딩 방법을 이용하여 추정된 95년도에서 97년도 사이의 가격상승률이 각각 9.88%과 6.95%로, 두 방법 사이에 2.93% 차이가 있음을 보이고 있다.

손철(2002)은 미국 Portland, Oregon의 토지감정평가 자료를 이용하여 각 필지를 주소와 필지고유번호를 이용하여 지오코딩한 후 각각의 위치를 이용하여 토지 감정가격과 CBD로부터의 거리, UGB(Urban Growth Boundary)로부터의 거리 등의 위치변수를 이용하여 헤도닉 함수를 추정하였다. 그 결과 위치변수의 추정모수의 크기가 두 방법간에 0%-27.54% 정도 차이가 남을 보고하였다. 또한 동 연구는 서로 다른 지오코딩방법을 적용하였을 경우, 두 방법간에 존재하는 모수추정치의 차이가 동일한 데이터에 대해 서로 다른 함수 형태(Function Form)를 적용하였을 경우보다 크다는 것을 보이고 있다.

이상의 연구결과가 시사하는 바는 공간분석을 위해 현실의 객체를 지리정보시스템 내 디지털 공간의 특정위치에 표현하기 위해 사용하는 여러 지오코딩방법에 따라 공간분석의 결과가 달라질 수 있으며, 그 달라짐의 크기는 우리가 일반적으로 중요시하는 요인, 예를 들어 분석에 있어 다른 함수식을 사용하는 것보다 클 수 있다는 점이다.

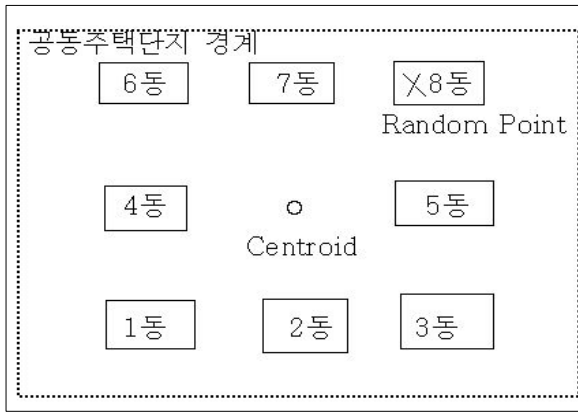
---

#### IV. 분석모형

앞서 살펴본 바와 같이 지금까지 수치지도를 이용하여 헤도닉 함수를 추정하는 연구들에서는 하나의 공동주택단지의 위치를 표시하기 위해 X,Y 좌표를 가진 하나의 포인트 객체를 수치지도상에 생성하여 왔다. 그러나 어떤 기준으로 공동주택단지를 대표하는 포인트 객체의 X,Y 좌표를 선택하였는지 기술하지 않고 있다. 공동주택단지의 위치표현에 대한 이러한 불명확성은 1개 동에서 수십 개 동으로 구성된 우리나라 공동주택단지의 규모를 고려할 때 이들 연구로부터 얻어진 연구결과의 수평적 비교를 어렵게 하는 요인이 되고 있다.

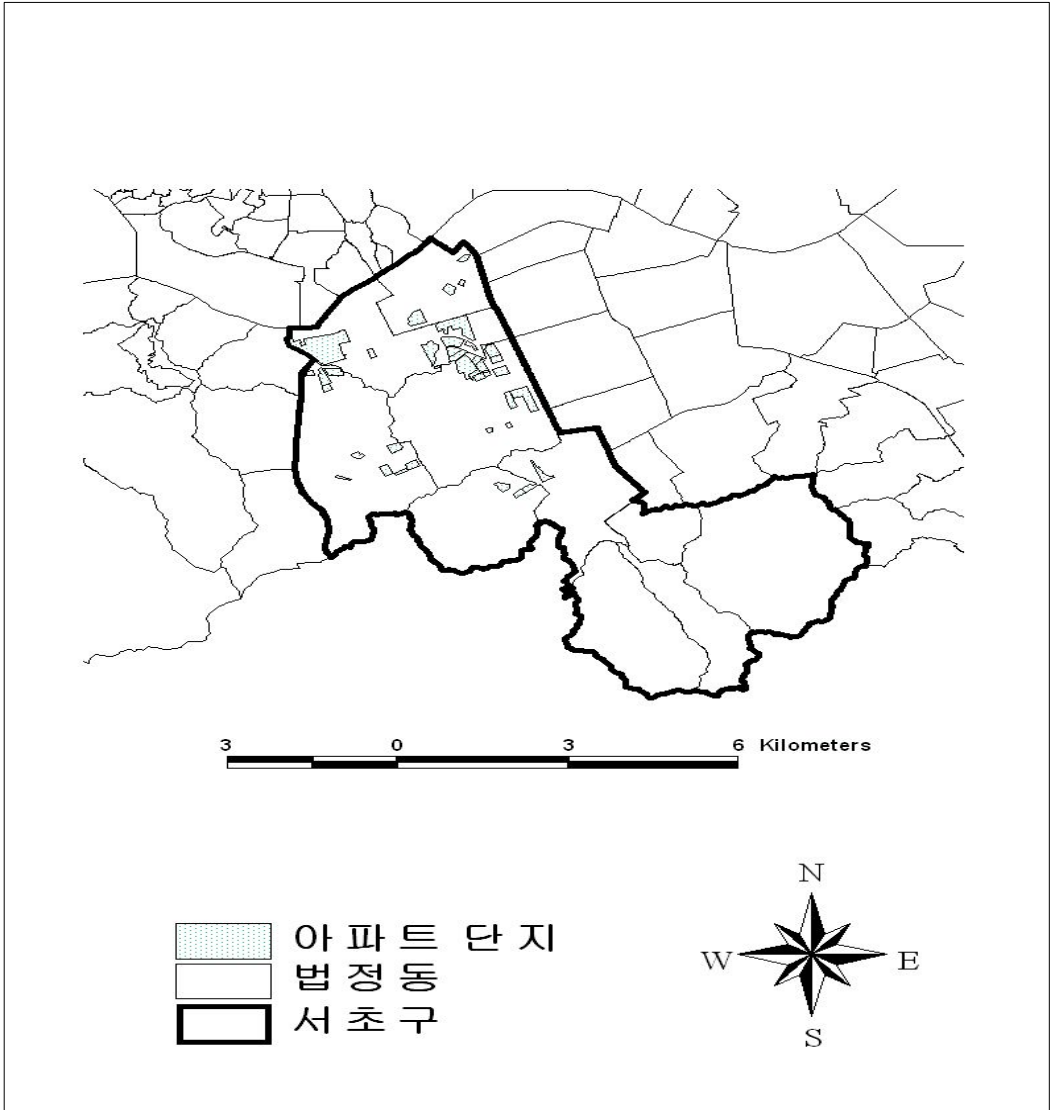
이상의 배경에서 본 연구에서는 연구자들이 서로 다른 기준에 의해 공동주택단지의 위치를 대표하는 포인트 객체의 좌표를 선택할 경우 헤도닉 함수의 위치변수 모수추정결과가 얼마나 달라질 수 있는가를 평가하고자 한다. 이와 같은 평가로부터의 결과는 지리정보시스템과 수치지도를 이용한 기존의 공동주택시장에 대한 헤도닉 함수의 추정결과들을 수평적으로 비교하는 데 유용한 정보를 제공할 것이다.

<그림 IV-1> 본 연구에서 사용된 공동주택단지 위치표현의 두 가지 방법



K C I

<그림 IV-2> 연구에 사용된 39개 서초구 공동주택단지



연구의 목적을 달성하기 위해 본 연구에서는 기존의 연구자들이 사용했으리라 추측되는 2개의 서로 다른 공동주택 위치표현 방법을 이용하여 서울시 서초구에서 선택한 임의의 공동주택 단지의 위치를 1:5000 수치지도에 표시한다. 그리고 이를 이용하여 서로 다른 공동주택 위치를 기반으로 하는 2개의 헤도닉 함수를 추정 한 후 두 모델 간에 존재하는 차이를 비교하고자 한다. 본 연구에서 이용하고자 하는 공동주택단지 위치표현 방식은 단지의 중심점(Centroid)을 이용하는 방법과 단지에 속하는 임의의 동 폴리곤의 중심점(Random Point)을 이용하는 방법이다. 두 방법의 차이점은 <그림 IV-1>에 나타나 있다. 두 방법은 <그림 IV-2>에 표시된 바와 같이 서울 서초구에 존재하는 39개 공동주택단지에 적용되었다.<sup>3)</sup>

### 중심점 모델 (Centroid Model)

$$MP = \alpha^0 + \alpha^1 PYOUNG + \alpha^2 THH + \alpha^3 DY1 + \alpha^4 DY2 + \alpha^5 DY3 + \alpha^6 DY5 + \alpha^7 DY6 + \alpha^8 DSUBC + e \quad \text{<식 IV-1>}$$

### 임의점 모델 (Random Point Model)

$$MP = \beta^0 + \beta^1 PYOUNG + \beta^2 THH + \beta^3 DY1 + \beta^4 DY2 + \beta^5 DY3 + \beta^6 DY5 + \beta^7 DY6 + \beta^8 DSUBR + e \quad \text{<식 IV-2>}$$

3) 중심점을 얻기 위해서는 단지들의 경계를 서울시 교통지도(중앙지도사 간)를 이용하여 파악한 후 CADLAND에서 제공한 1:5000 GIS 레이어에 폴리곤 객체로 그린 다음, 각 폴리곤의 중심점을 ArcView 3.2를 통해 취득하였다.

<표 IV-1> 변수의 정의

변수	정의
MP	최고가격과 최저가격의 평균(만원)
PYOUNG	평형
THH	단지내 총가구수
DY1	1975>=Year(입주시점) 이면 1, 아니면 0
DY2	1975<Year<=1980 이면 1, 아니면 0
DY3	1980<Year<=1985 이면 1, 아니면 0
DY4	1985<Year<=1990 이면 1, 아니면 0
DY5	1990<Year<=1995 이면 1, 아니면 0
DY6	1995>Year 이면 1, 아니면 0
DSUBR	임의점에서 가장 가까운 지하철역까지의 직선 거리(미터)
DSUBC	중심점에서 가장 가까운 지하철역까지의 직선 거리(미터)

본 연구에서 사용된 두가지 헤도닉 함수의 설정은 <식 IV-1>과 <식 IV-2>와 같다. 본 연구에서는 <식 IV-1>을 중심점 모델, <식 IV-2>를 임의점 모델이라 부르기로 한다. 두 모델의 차이점은 <식 IV-1>에서는 DSUBC라는 변수가 사용된 데 비해 <식 IV-2>에서는 DSUBR이 사용되었다는 점이다. <식 IV-1>과 <식 IV-2>에 사용된 변수의 정의는 <표 IV-1>에 나타난 바와 같다. <표 IV-1>에 설명된 변수들의 실제값들은 2003년 8월 1일자 부동산뱅크의 웹사이트([www.neonet.co.kr](http://www.neonet.co.kr))에서 취득되어 분석에 이용되었다. 또한 두 모델간의 위치변수 모수추정결과의 차이를 계산하기 위해 <식 IV-3>과 <식 IV-4>가 사용되었다.

### 모델간의 차이

$$\text{차이} = \left| \frac{\widehat{\alpha}^8 - \widehat{\beta}^8}{\widehat{\beta}^8} \right| \quad \text{<식 IV-3>}$$

$$\text{차이} = \left| \frac{\widehat{\alpha}^8 - \widehat{\beta}^8}{\widehat{\alpha}^8} \right| \quad \text{<식 IV-4>}$$

## V. 분석결과

<표 V-1>은 <식 VI-1>과<식 VI-2>의 추정에 사용된 변수들의 통계량을 보여주고 있다. <표 V-1>을 보면 총 100개의 관측치가 추정에 사용되었음을 알 수 있다. 이 관측치는 최종적으로 추정에 사용된 28개 서초구내 공동주택단지의 평형별 정보를 포함한다. 당초 고려되었던 39개 단지 가운데 최종적으로 28개 단지만이 추정에 이용된 것은 추정에 사용된 관측치를 DSUBR과 DSUBC가 1000미터 이내인 것으로 한정하였고 일부 (재)건축중인 단지가 제외되었기 때문이다.<sup>4)</sup>

<표 V-1>을 살펴보면 각 단지별 임의점과 중심점간의 직선거리를 나타내는 DIST 변수에 대한 통계량을 볼 수 있다. DIST 변수의 최소값과 최대값은 2미터와 553미터이며 그 평균값은 128미터이다. 따라서 이상의 특성을 가진 관측치를 이용하여 나타나는 모델간의 차이점은 공동주택의 위치표현 방법 사이에 평균 130미터 내외의 차이가 존재하는 경우 발생할 수 있는 결과로 해석되어야 할 것이다. 참고로 DIST가 553인 공동주택단지는 <그림 V-1>에 나타난 바와 같이 구반포주공이며 DIST가 2미터인 단지는 <그림 V-2>에 나타난 한신 12차이다.

<표 V-2>는 <식 IV-1>과 <식 IV-2>의 추정결과를 보여준다. <표 V-2>의 결과는 이분산에 강건한 White Estimator(White, 1980)를 이용하여 추정되었다. <표 V-3>은 <식 IV-1>과 <식 IV-2>의 추정에 사용된 설명변수의 Variance Inflation Factor를 보여주고 있다. <식 IV-1>과 <식 IV-2>에서의 최대 VIF는 모두 3을 넘지 않아 두 모델 모두에서 다중공선성의 문제는 심각하지 않음을 보여준다. 임의점 모델과 중심점 모델의 Adjusted R<sup>2</sup>는 각각 0.8952와 0.9019로, 추정에 사용된 24개 공동주택단지의 가격을 설명하는 데 중심점 모델이 약간 우월함을 보이고 있다. 이러한 결과는 공동주택의 위치표현시 단순하게 임의의 단지내 동을 선택하는 것보다

---

4) 다만 DSUBC의 경우 정확히 1000미터로 한정할 경우 선택된 공동주택단지의 수가 DSUBR과 다른 경우가 발생하여 <표 V-1>에서 볼 수 있는 바처럼 약 1002미터로 그 범위를 확대하였다. 이와 같은 DSUBC, DSUBR 값에 의해 관측치를 제한한 것은 지하철 접근도의 영향이 주택이 지하철역으로부터 1000미터 이상되는 지점에 위치할 경우 존재하지 않는다고 가정하였기 때문이다.

---

단지의 중심을 선택하는 방법이 이를 바탕으로 추정된 헤도닉 함수의 통계적 질을 개선할 수 있을 가능성을 시사해준다.<sup>5)</sup>

구체적으로 추정 결과를 살펴보면 임의점 모델과 중심점 모델에서 1평의 증가가 공동주택가격에 미치는 영향은 1,423만원에서 1,427만원 정도로 나타났다. 단지의 전체 세대수가 공동주택가격에 미치는 영향은 1세대 증가당 두 모델에서 모두 11만원 정도로 거의 차이가 없다. 주택의 건축연도와 가격의 관계를 나타내는 더미변수의 계수는 모두 5%와 1% 신뢰수준에서 유의하며 건축연도와 가격사이에 U자형 관계가 있음을 두 모델 모두에서 보이고 있다. 마지막으로 지하철역으로부터 직선 거리가 1미터 증가할 경우 주택가격의 하락폭은 임의점 모델에서는 8만원이며 중심점 모델에서는 12만원 정도로 추정되었다. 이 결과는 고용원의 2인(2001)에서 보고된 1km 증가당 836만원 감소(1미터 증가당 0.836만원 감소)<sup>6)</sup>와 서울시정개발연구원(2000)에서 보고된 1미터 증가당 3.28만원 감소<sup>7)</sup>보다는 상당히 높은 수치이다. 그러나 고용원, 김홍규, 유완(2001)과 서울시정개발연구원(2000)의 결과가 모두 1997년의 서울시 공동주택가격을 이용한 것이기 때문에 2003년과 1997년 사이의 물가인상폭을 감안하면 그 차이는 상당히 줄어들 것이다.<sup>8)</sup>

<표 V-2>에서 볼 수 있는 바처럼 <식 IV-3>과 <식 IV-4>를 이용하여 계산된 위치변수(DSUBR/DSUBC)의 모수추정결과의 차이는 각각 53%와 34% 정도이다. 이상과 같은 차이는 손철(2002)이 미국의 상황에서 발견한 서로 다른 지오코딩 방법 사이에 존재하는 위치변수 모수추정치 크기의 차이(0%-27.54%)보다 큰 것이다. 과연 이 같은 차이의 정도가 얼마나 심각한 것인가를 상대적으로 검토하기 위해 <표 V-4>에서는 중심점 모델을 선형함수와 세미로그함수 형태를 이용하여 추

---

5) 공동주택 구매자들이 공동주택단지의 전체적 입지적 특성을 고려할 때 단지의 임의의 지점보다는 단지의 중심지를 기준으로 할 가능성이 다분하다. <표 V-2>의 추정결과는 그러한 가능성을 반영하는 것으로 해석될 수 있다.

6) 고용원, 김홍규, 유완(2001)의 p.18의 <표 5>에서 인용

7) 서울시정개발연구원(2000)의 p.106의 <표 4-11>에서 인용

8) 고용원, 김홍규, 유완(2001)의 연구와 서울시정개발연구원(2000)의 연구는 동일한 연도(1997)의 서울 공동주택시장을 대상으로 했음에도 지하철 접근도의 한계잠재가격 측면에서 매우 다른 결과를 보이고 있어 흥미롭다. 두 연구에서의 1평의 한계잠재가격은 각각, 1129만원과 939.89만원으로 지하철 접근도에 비하면 상당히 유사하다.

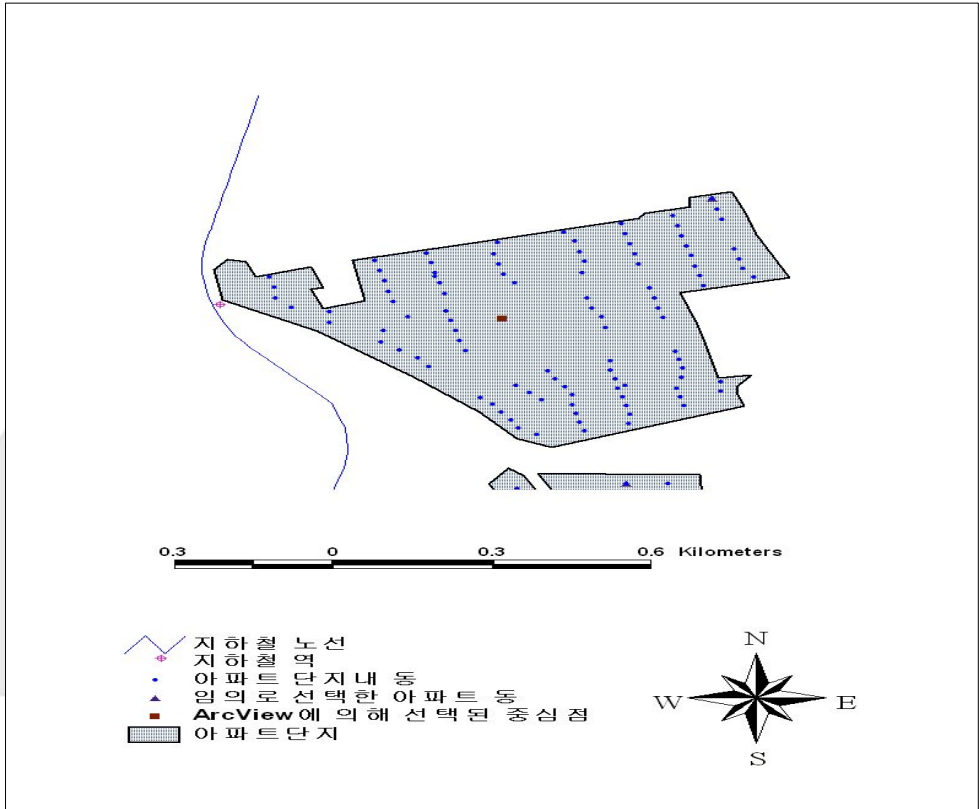
정한 두 가지 결과를 보고하고, 각각의 경우 한계잠재가격을 계산하여 보고하고 있다. <표 V-4>의 6번째 열과 7번째 열에서 볼 수 있는 것처럼 두 함수형태에서 계산된 위치변수의 한계잠재가격의 차이는 0.45% 정도이다. 이 정도의 차이는 서로 다른 위치표현방법을 사용하였을 때 나타나는 한계잠재가격의 차이보다 훨씬 작은 것이다(0.45% 대 53% 또는 35%).

<표 V-1> 추정에 사용된 관측치의 통계량

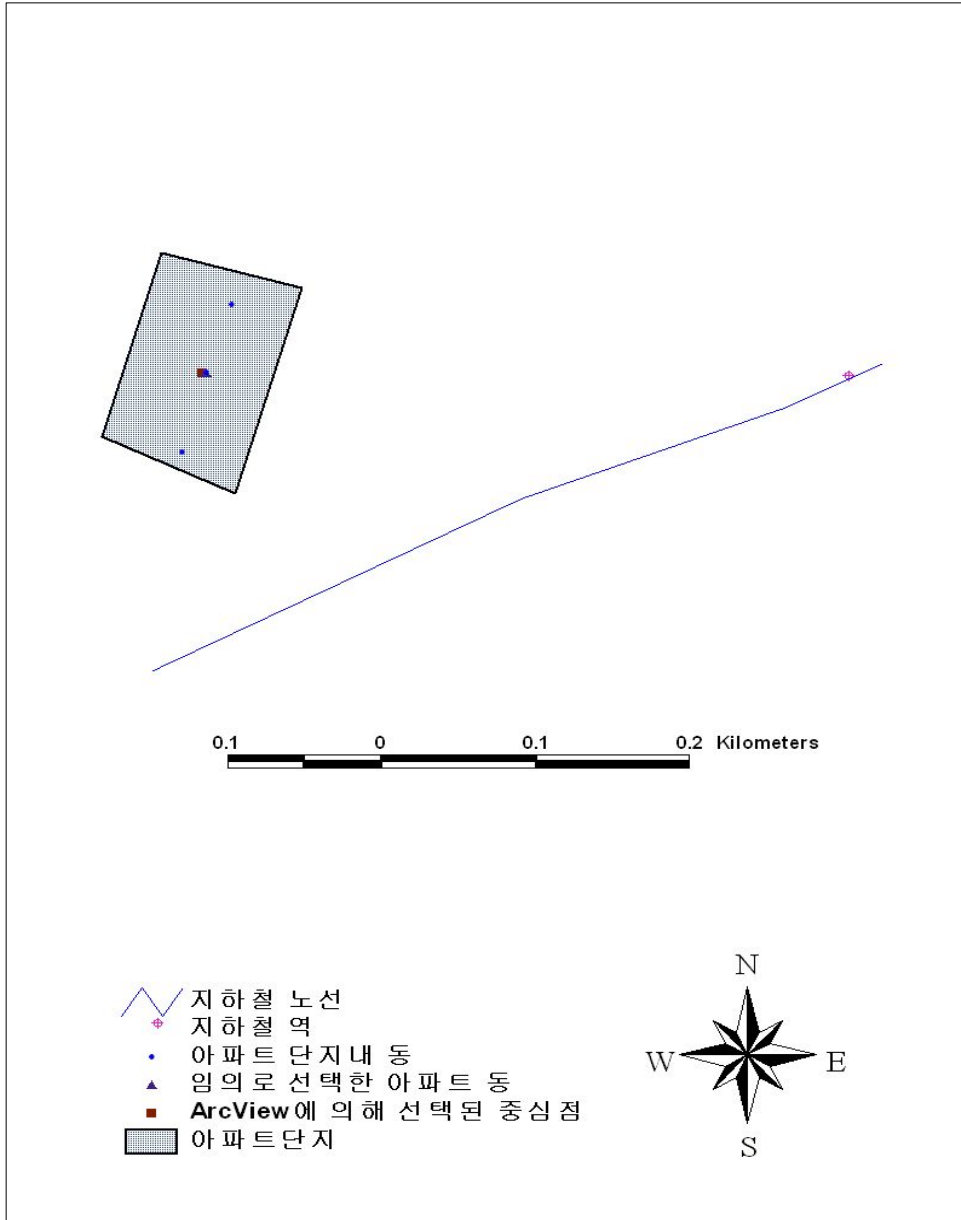
변수명	관측수	평균값	최소값	최대값
MP(만원)	100	59497.5	23000	147000
PYOUNG(평)	100	38.44	16	65
THH(가구)	100	808.73	140	3590
YEAR(연도)	100	1982.92	1973	2000
DSUBR(미터)	100	656.8586	231.148	1002.246
DSUBC(미터)	100	653.0759	243.289	995.347
DIST(미터) <sup>1)</sup>	100	128.1671	2.173752	553.083

1) 임의점과 중심점간의 직선거리

<그림 V-1> 구반포 주공아파트



<그림 V-2> 한신 12차 아파트



<표 V-2> 중심점 모델과 임의점 모델의 추정결과

변수	임의점 모델		중심점 모델		차이1	차이2
	계수(1)	유의성	계수(2)	유의성	$\frac{ (1)-(2) }{(1)}$	$\frac{ (1)-(2) }{(2)}$
PYOUNG	1422.997	***	1427.033	***	N/A	N/A
THH	11.437	***	10.766	***	N/A	N/A
DY1	19823.800	**	17023.600	**	N/A	N/A
DY2	10904.930	***	10588.540	***	N/A	N/A
DY3	10815.870	***	9964.538	***	N/A	N/A
DY5	7169.737	***	6260.045	***	N/A	N/A
DY6	17373.020	***	16307.020	***	N/A	N/A
DSUBR/ DSUBC	-7.894	*	-12.071	***	52.91%	34.60%
CONS	-9302.982	**	-5629.926	*		
R2	0.9037		0.9098			
Adj R2	0.8952		0.9019			
F-value	F(7,91)=106.85		F(7,91)=121.11			

1)\*\*\*: 1% 수준에서 유의함, \*\*: 5% 수준에서 유의함, \*:10% 수준에서 유의함

<표 V-3> Variance Inflation Factor

	임의점 모델	중심점 모델
THH	2.99	2.92
DY3	2.49	2.49
DY2	2.35	2.45
DY1	2.29	2.69
DY6	1.74	1.73
DSUBR/ DSUBC	1.22	1.52
DY5	1.09	1.11
PYOUNG	1.05	1.05
Mean VIF	1.9	1.99

<표 V-4> 중심점 모델의 한계잠재가격(Marginal Implicit Price) 비교

변수	중심점 모델 - 세미로그 형태 <sup>1)</sup>			중심점 모델 - 선형모델	차이1	차이2
	계수	유의 성	한계잠재 가격(1)	한계잠재가 격(2)	$ (1)-(2) $ (1)	$ (1)-(2) $ (2)
PYOUNG	0.0239	***	1421.235 <sup>2)</sup>	1427.033	N/A	N/A
TTHH	0.0002	***	10.52511 <sup>2)</sup>	10.766	N/A	N/A
DY1	0.1966	*	12510.61 <sup>3)</sup>	17023.600	N/A	N/A
DY2	0.2601	***	17613.54 <sup>3)</sup>	10588.540	N/A	N/A
DY3	0.2338	***	15614 <sup>3)</sup>	9964.538	N/A	N/A
DY5	0.2063	***	13592.2 <sup>3)</sup>	6260.045	N/A	N/A
DY6	0.3000	***	20661.22 <sup>3)</sup>	16307.020	N/A	N/A
DSUBC	-0.0002	***	-12.1256 <sup>2)</sup>	-12.071	0.45%	0.45%
_cons	9.7941	***				
R2	0.8900					
Adj R2	0.8800					
F-value	F(7,91)=107.52					

1)이분산에 강건한 White Estimator를 이용하여 추정됨

2)연속변수의 경우: 계수 x 평균 공동주택가격 (59,497.5 만원 )

3)터미변수의 경우:  $[\exp(\text{계수}-0.5 \times \text{계수의 분산})-1] \times \text{평균 공동주택가격}(59,497.5 \text{ 만원})$ <sup>9)</sup>

4)\*\*\*: 1% 수준에서 유의함, \*\*: 5% 수준에서 유의함, \*:10% 수준에서 유의함

9) 세미로그 형태의 회귀식에서 터미변수 계수의 올바른 추정치(Estimator)에 대해서는 Kennedy (1981) 참조. 본 연구에서의 한계잠재가격은 기존의 터미변수 계수를 Kennedy(1981)의 변환방법에 의해 변환한 후 계산됨

---

지금까지의 분석은 상이한 공동주택 위치표현 방법을 기반으로 한 헤도닉 함수들의 위치변수의 모수추정치 간에 35% 또는 53%의 크기차이가 존재한다는 것을 보여준다. 또한 그 차이의 정도가 동일한 데이터에 다른 함수식을 적용하였을 경우보다 심각함을 보이고 있다.

본 연구의 결과가 주는 주요교훈은, 헤도닉 함수를 추정하거나 추정결과를 이용하고자 하는 연구자는 위치변수의 측정결과에 직접적 영향을 미치는 공동주택 위치표현방법에 주의를 기울여야 한다는 것이다. 그것은 위치표현방법간의 평균 130미터 정도의 차이에도 35%에서 53%정도의 위치변수 모수추정치의 크기 차이가 발생할 수 있기 때문이다. 실제로 이 수치들이 특정 프로젝트의 효과를 평가하는 데 사용될 경우, 어떤 수치를 사용하는가가 프로젝트의 전체 효과를 평가하는 데 커다란 영향을 미칠 것이 분명하다.

## VI. 결론

본 연구에서 다루어진 주제는 우리나라 공동주택 시장의 데이터를 이용한 헤도닉 모델의 추정시 연구자들이 임의로 선택한 공동주택 위치표현방법에 따라 헤도닉 함수의 위치변수 모수추정결과가 얼마만큼 달라지는가 하는 점이다. 연구결과는 본 연구에서 설정한 서로 다른 위치표현 방법에 따라 추정된 헤도닉 함수들 사이에 53%(또는 35%)정도의 위치변수 모수추정치 크기의 차이가 존재한다는 것이다. 본 연구의 결과는 지리정보시스템을 이용한 헤도닉 함수추정시 우리나라의 경우, TIGER 파일이 널리 이용되고 있는 미국과는 환경이 다르지만 미국과 유사한 지오코딩형태에 따라 헤도닉 함수의 추정결과가 달라지는 문제가 존재한다는 것을 보여주는 첫번째 경험적 결과이다.

일반적으로 1:5000 수치지도와 부동산정보업체가 수집한 단지별 공동주택정보를 지리정보시스템을 이용한 헤도닉 함수 추정에 사용하는 우리의 상황에서 연구자들이 공동주택 위치표현 방식에서 통일된 기준을 사용하거나 공동주택 개별 동에 대한 위치정보를 기타의 속성정보와 같이 이용하지 않는 한 다양한 공동주택 위치표현방식에 따른 추정결과의 차이는 해소되지 않을 것이다. 따라서 본 연구의 결과가

---

주는 주요 시사점은 연구자나 연구결과의 사용자가 이상과 같은 수치지도상의 위치 표현방법 때문에 존재할 차이를 항상 인식하고 연구결과를 이용하거나 연구결과들을 비교해야 하며, 연구자의 경우 자신이 사용한 위치표현 방법을 명확히 서술해야 한다는 점이다. 즉, 연구자나 연구결과의 이용자는 드러난 연구결과나 연구결과의 차이가 연구에 사용된 데이터, 추정방법, 시점, 함수형태, 포함된 변수의 유형뿐만 아니라 공동주택의 위치표현의 방식에 의해서도 다를 수 있음을 염두에 두어야 하며, 따라서 연구결과가 다른 위치표현 방식을 적용할 경우 달라질 가능성이나 여러 연구 간에 존재하는 차이가 동일한 위치표현방법을 사용하면 해소될 가능성을 항상 고려해야 한다.



---

## 참고문헌

- 고원용, 김홍규, 유완. 2001. "서울시 주택 하위시장별 주거환경이 공동주택가격에 미치는 영향," 지역연구 제17권: pp9-27.
- 박헌수. 2001. "모수적방법과 준모수적방법에 의한 주택가격 함수 추정에 관한 연구." 국토계획 제 36권: pp153-165.
- 박헌수, 정수현. 2002. "A Study on the Estimation of the Hedonic Price Function of Land with Spatial Variation." 국토계획 제 37권: pp255-265.
- Bae, June and Park. 2003. "The impact of Seoul's subway Line 5 on residential property values." Transport Policy 10: pp85-94.
- 서경천, 이성호. 2001. "지가의 공간적 변동에 따른 입지지의 분석에 관한 연구." 국토계획 제 36권: pp55-71.
- 서울시정개발연구원. 2000. 지하철 5호선건설 사후평가 연구.
- 손철. 2001. "Comparing the Impacts of Different Geocoding Methods on the Estimated Parameters of the Non-Location Variables Insensitive to Geocoding in a Hedonic Model Specification." 한국 GIS학회 제9권: pp643-658.
- 손철. 2002. "Do Different GIS-based Geocoding Methods Result in Different Parameter Estimates for Location Variables from a Hedonic Housing Price Model?" 국토계획 제37권: pp153-168.
- Bae, Jun, and Park. 2003. "The Impact of Seoul's Subway Line 5 on Residential Property Values." Transport Policy 10: pp85-94.
- Kennedy, P. 1981. "Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations." American Economic Review 71: pp801
- Ratcliffe, Jerry H. 2001 "On the accuracy of TIGER-type geocoded address data in relation to cadastral and census areal units." International Journal of Geographic Information Sciences 15(5): pp473-486.
- Rodriguez, M., Sirmans, C. and Marks, A. 1995. "Using Geographic Information Systems to Improve Real Estate Analysis." *Journal of Real Estate Research* 10( 2): pp163-173.

---

White, Hal. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica*, 48, 4: 817-838.

[http://www.ncjrs.org/html/nij/mapping/ch4\\_3.html](http://www.ncjrs.org/html/nij/mapping/ch4_3.html)

K C I