The Journal of GIS Association of Korea, Vol. 16, No. 4, pp. 489-497, December 2008

Evaluating the Performance of a Polygon based Approach to Represent Apartment Complexes in a GIS based Hedonic Housing Price Analysis

Sohn, Chul*

ABSTRACT

Currently, GIS has been widely used in the hedonic analyses of urban apartment housing markets in Korea. In those analyses, the apartment complexes are typically represented as the points or the polygons on the GIS maps and the location variables of the analyses are measured based on the points or the polygons. In this study, the relative performance of the point based approach and the polygon based approach in a GIS based hedonic analysis was compared using the apartment housing market data from the north eastern part of the city of Seoul and Davidson and MacKinnon Test. The results from this study indicate two things. First, two approaches can produce substantially different results in a hedonic price model estimation. Second, the polygon based approach produces a hedonic price model which explains the price variations better than the point based approach. These findings suggest that Korean researchers who are interested in improving quality of hedonic price model estimations and use GIS to measure the location variables for hedonic price models should consider using the polygon based approach with the point based approach. This is because the polygon based approach can produce the location variables with the shortest straight line distances and can explain the housing price variations well.

Keywords : Hedonic Analysis, GIS, Polygon based Approach

요 약

주택시장에 대한 헤도닉분석을 위해 GIS가 활발하게 사용되고 있다. 그 것은 GIS를 이용할 경우 헤도닉 분석에서 필수적인 위치 변수를 손쉽게 측정할 수 있기 때문이다. 현재까지 국내의 주택시장을 대상으로 하는 헤도닉 분석은 전자지도상에서 포인트들로 표현된 공동주택 단지와 여러 중요 도시내 위치간의 직선최단거리를 측정하는 데 GIS를 이용하고 있다. 그런데

* Department of Regional Development, Kangnung National University

이상의 연구들에서 간과한 점은 실제의 직선최단거리는 포인트보다는 폴리곤으로 보다 정확하게 측정될 수 있다는 점이다. 물론 GIS에 기반한 헤도닉분석에서 공동주택을 나타내는 데 포인트를 사용하느냐와 폴리곤을 사용하느냐는 연구의 목적에 따라 관심의 대상이 아닐 수 있다. 그러나 추정되는 모델의 질을 향상시키는 것이 목적이라면 선택에 신중을 기해야 할 문제이다. 이런 배경에서 본 연구에서는 서울 강북지역 공동주택의 위치를 포인트와 폴리곤으로 전자지도상에 각각 나타낸 후 각각의 경우 추정된 헤도닉 모델에 대해 Davidson and MacKinnon Test를 적용하여 어떤 경우 보다 우월한 모델이 추정되었는지 검증해 보았다. 검증결과 공동주택 단지를 폴리곤으로 나타낼 경우 이에 기반한 최단 직선거리 변수를 포함한 모델추정 결과가 우월한 것으로 나타났다. 이 결과가 의미하는 것은 GIS를 이용한 공동주택시장에 대한 헤도닉 모델추정에서 공동주택단지를 GIS 지도상에 폴리곤으로 나타내고 이를 기반으로 다양한 위치관련 변수를 추정하는 것이 보다 나은 모델 추정결과를 가져올 수 있다는 점이다.

주요어 : 헤도닉 분석, 지리정보시스템, 폴리곤 기반 분석

I. Background and Objective

Currently, GIS has been widely used in the hedonic analyses of urban housing markets in Korea and other countries. This is because the location variables, which are essential in those hedonic analyses, can be easily measured by spatial analysis functions of GIS. In the case of Korea, the hedonic analyses have been mainly conducted for the urban apartment housing market because the data regarding the apartment housing market are easily obtained from the real estate market data providers. When the hedonic analyses for the Korean apartment housing market are conducted, GIS is used to measure the location variables of apartment complexes which are typically represented as points in the GIS maps.

Instead of representing an apartment complex as a point, we can also represent an apartment complex as a polygon on a GIS map and measure the location variables based on the polygons. However, this approach to represent an apartment complex as a polygon is not popular in the hedonic studies of Korean urban housing market. This is because there are no empirical studies on the usefulness of the polygon based approach.

Because of this, most of GIS based hedonic analyses on the Korean urban apartment housing market simply use the point based approach while not knowing the possibility of the polygon based approach. In this vein, this study is designed to compare and report the relative explanatory performances of two approaches to inform about the possibility of the polygon based approach.

II. Problem Statement

To conduct the hedonic analyses on urban housing market, we need to measure the various types of the location variables such as distances to major employment centers or transportation nodes. To measure the location variables in a GIS based hedonic analysis for the Korean apartment housing market, the location of each apartment complex should be represented on a GIS map as a point or a polygon. As mentioned before, the location of each apartment complex is typically represented as a point on a GIS map in Korea.[1)]

When we use the points to represent the complexes, we can choose the centroid of the apartment complexes or the coordinates for the main and side gates of the apartment complexes to locate the complexes. After we represent the complexes on the GIS map, the distances can be measured as either straight line distance or network distance.

Table 1. Combinations for Distance Measurement: the Point based Approach

Centroid + Straight Line Distance
Centroid + Network Distance
Point for Gate(Main or Side Gates) + Straight Line Distance
Point for Gate(Main or Side Gates) + Network Distance

1) Recent GIS based hedonic analyses such as Seo (2006) and Choi and Yoon (2004) used the point based approach while Sohn and Shin (2007) used the mixture of the point based approach and the polygon based approach.

Table 1 shows the various combinations of the distance types and the types of the points representing the apartment complexes. Among the combinations, the combination of 'Centroid and Network Distance' or 'Point for Gate and Network Distance' can give us the most realistic distance information. However, to use this combination, we need the information for the coordinates of all the main & side gates of the apartment complexes and road network inside the complexes, which are very difficult to obtain.[2)] Thus the most realistic option to choose is to use the combination of 'Centroid and Straight Line Distance'. To use this combinations, we only need the centroid points and simple GIS functions such as 'spatial join' of ArcView 3.2. This is the reason why most researchers use this combination in their hedonic studies.

However, because the size of each apartment complex is relatively large in Korea, the problem of using 'Centroid and Straight Line Distance' combination is that it can not produce the shortest distance between the apartment complexes and the urban destinations in most of cases.

The real example indicated in Figure 1 shows the problem of using a point to represent an apartment complex. Figure 1 shows the several apartment complexes close to the subway stations of Seoul's northern region. The complexes are represented as both centroid points and polygons

2) The coordinates information for the side gates are hard to get because many side gates are created by the residents after the completion of the apartment building project. When we deal with many apartment complexes to conduct the areawide hedonic studies, it is almost impossible to survey all the locations of the side gates.

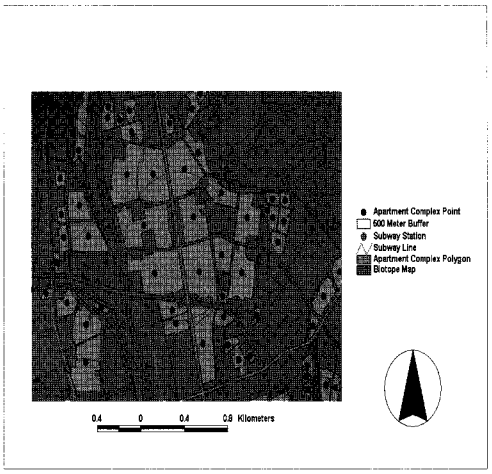


Figure 1. Two Approaches for Representing Apartment Complexes

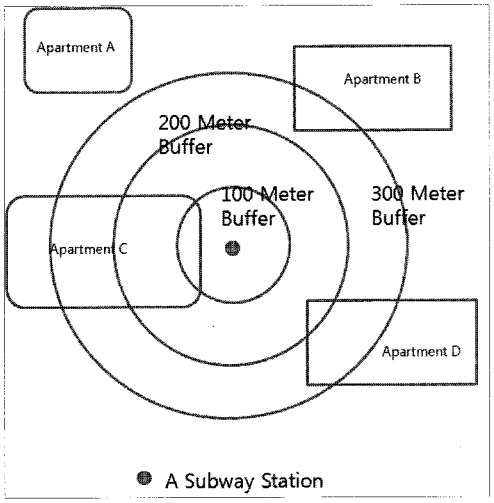


Figure 2. A Simple Method for Polygon based Distance Measurement

for the purpose of comparison. As we can see easily, a complex can be identified as within the subway station catchment area, typically defined as a 500 meter buffer area from the stations, or as outside the subway station catchment area, according to the approaches used to represent the apartment complexes.

This problem occurs because the centroid points are within the relatively large apartment complexes. Therefore the shortest straight line distances from the complexes cannot be produced from these points. An alternative to the point based approach is the approach to use polygons to represent the apartment complexes. When we use the polygons to represent the apartment complexes, we can measure the shortest straight line distances between the apartment complexes represented by the polygons and the various urban destinations using the spatial analysis functions of GIS.

For example, as indicated in Figure 2, if we want to measure the straight line distance between the a subway station and the apartment complexes, we can do this by calculating the apartment complexes' area within the distance bands or the multiple buffer rings created from the subway station.[3] In addtion to this method, one may create a code which measures the shortest straight line distance between the polygons and the various urban destinations.

In sum, if we do not have the good GIS data about the all types of gates and inside road networks of the apartment complexes, the realistic option to conduct the GIS based hedonic studies on Korean apartment housing market is to measure the simple straight line distances based on the points or the polygons. However, as mentioned before, so far, the point based approach is dominant among the researcher although the polygon based

3) In case of ArcView 3.2, we can do this by using the 'Tabulate Area' function in the 'Spatial Analyst' Extension.

approach is useful in producing the shortest straight line distances. Thus, if the goal of the researchers is to obtain the shortest straight line distances, then the polygon based approach should be considered with the point based approach.

III. Research Method, Study Area, Data, and Model Specification

1. Research Method

In this study, to compare the explanatory performance of the point based approach and the polygon based approach, the Davidson and MacKinnon test is used. The logic of the Davidson and MacKinnon test, also called the J-test, is well explained in Maddala (2005).

$$H_0: \quad p = \beta x + e_0 \quad e_0 \sim N(0, \sigma_0^2) \tag{1}$$

$$H_1: \quad p = \gamma z + e_1 \quad e_1 \sim N(0, \sigma_1^2) \tag{2}$$

where p: dependent variable, x, z: competing independent variables, β, γ: coefficients. e: error term

According to the Davidson and MacKinnon test, when we have two competing models like (1) and (2), to check which one is preferred, we need to first estimate (1) and (2) by OLS, then we need to calculate the predicted values from the models such as $\widehat{p_0} = \hat{\beta} x$ and $\widehat{p_1} = \hat{\gamma} z$.

$$p = \beta x + \eta \widehat{p_1} + u \tag{3}$$

$$p = \gamma z + \delta \widehat{p_0} + \nu \tag{4}$$

Table 2. Suggested Interpretations

		Hypothesis: $\eta = 0$	
		Not Rejected	Rejected
Hypothesis $\delta = 0$	Not rejected	①	②
	Rejected	③	④

① Both H_0 and H_1 are acceptable
② H_1 is acceptable, H_0 is not
③ H_0 is acceptable, H_1 is not
④ Neither H_0 nor H_1 is acceptable
Source: Maddala (2005)

Then we need to estimate (3) and (4) by OLS and evaluate the statistical significance of η and μ. Table 2 shows the possible outcomes from the statistical significance tests and suggested interpretations.

For the detailed discussion about the theoretical logic behind the Davidson and MacKinnon test, please refer to Maddala (2005).

2. Study Area

To compare the explanatory performance of the point based approach and the polygon based approach in the hedonic analyses under the Korean housing market context, the housing market of northern parts of the city of Seoul is selected. This housing submarket consists of Dongdaemun-Gu, Jungnang-Gu, Seongbuk-Gu, Gangbuk-Gu, Dobong-Gu, and Nowon-Gu. Figure 3 shows the study area.

3. Data

The apartment housing prices and attributes data used in this study come from 「Real Estate 114」. The total number of 242 apartment com-

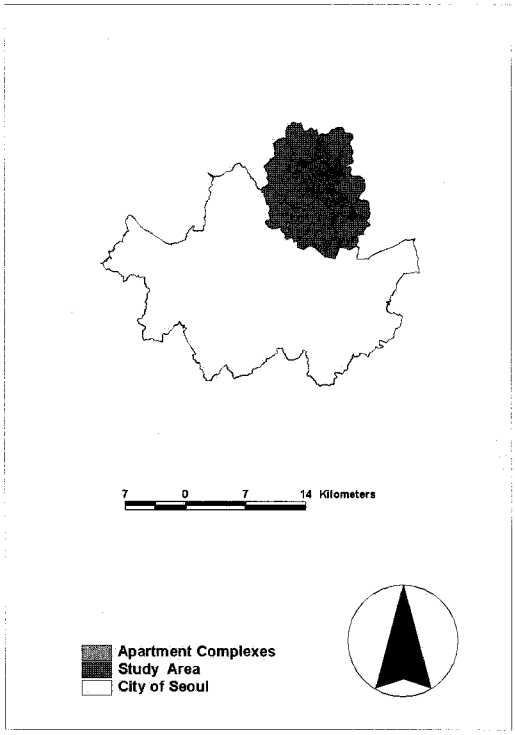


Figure 3. Study Area

plexes in the study area are considered in this study. The apartment housing prices used were surveyed in November, 2005. The GIS maps used to measure the location attributes come from the Seoul Development Institute.

4. Model Specification

Two competing apartment housing hedonic models for the study area are specified as (5) and (6) to achieve the research objective. The linear function form is used and the independent variables come from the previous hedonic studies on this submarket. Table 3 and Table 4 show the definitions and summary statistics of the variables.

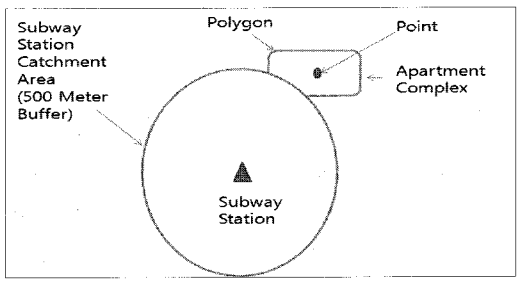
The Specification (5) can be called a "Point based Model" because this includes the independent variable for access to subway stations measured from the point representing apartment complexes, as indicated in Figure 4, while the specification (6) can be called a "Polygon based Model" because it includes the independent variable for access to subway stations measured from the polygon representation of apartment complexes.[4)]

H_0: Point based Model

$$mprice = \alpha + \beta^1 size + \beta^2 age + \beta^3 dthh500 + \beta^4 dheat + \beta^5 dstep + \beta^6 dsamsung + \beta^7 pointsub500 + e \quad (5)$$

H_1: Polygon based Model

$$mprice = \alpha + \beta^1 size + \beta^2 age + \beta^3 dthh500 + \beta^4 dheat + \beta^5 dstep + \beta^6 dsamsung + \beta^7 polysub500 + e \quad (6)$$



[Figure 4] Point based Approach Vs. Polygon based Approach

4) The variables representing the access to the subway stations are measured by "select by theme" function of ArcView 3.2 after creating the 500 meter buffer area from the subway stations.

Table 3. Definition of Variables[5)]

Variable	Definition
mprice	Mean price of a certain apartment size in the apartment complex in ten thousand won
size	Apartment size in pyoung
age	Apartment age in year
dthh500	dthh500=1 if the number of total households >=500
dheat	dheat=1 if the apartment complex is in a regional heating system
dstep	dstep=1 if the apartment is corridor type
dsamsung	Straight line distance to Samsumg subway station
pointsub500	pointsub500=1 if the apartment exists within 500 buffer from the subway station based on a point representation
polysub500	polysub500=1 if the apartment exists within 500 buffer from the subway station based on a polygon representation

Table 4. Summary Statistics of Variables

Variable	Obs	Mean	Std. Dev.	Min	Max
mprice	1023	20580.43	9812.1	5500	66500
size	1023	29.89736	8.83875	10	74
age	1023	10.33236	5.70671	1	34
thh	1023	739.5503	700.7392	40	3003
dthh500	1023	0.487781	0.500095	0	1
dstep	1023	0.396872	0.489488	0	1
dheat	1023	0.200391	0.400489	0	1
dsamsung	1023	13352.96	3342.316	6398.187	19808.23
point sub500	1023	0.423265	0.494318	0	1
poly sub500	1023	0.512219	0.500095	0	1

After (5) and (6) are estimated and the predicted values from the estimated (5) and (6) are used as indicated in (7) and (8) to conduct the Davidson and MacKinnon test.

$$mprice = \alpha + \beta^1 size + \beta^2 age + \beta^3 dthh500 + \beta^4 dheat + \beta^5 dstep + \beta^6 dinc1000 + \beta^7 dsamsung + \beta^8 pointsub500 + \eta \widehat{PLM} + e$$

where $\widehat{PLM}$ is the predicted value form Polygon based Model (7)

$$mprice = \alpha + \beta^1 size + \beta^2 age + \beta^3 dthh500 + \beta^4 dheat + \beta^5 dstep + \beta^6 dinc1000 + \beta^6 dsamsung + \beta^8 polysub500 + \delta \widehat{PTM} + e$$

where $\widehat{PTM}$ is the predicted value form Point based Model (8)

IV. Results

Table 5 and Table 6 show the estimates from two models specified in (5) and (6).[6] Table 5 and Table 6 indicate that estimation results from regressions are acceptable because R^2s are around 0.80 and F-Values are between 392 and 405.

The regression results for (5) and (6) reported in Table 5 and Table 6 suggest that the access to subway stations is positively associated with the apartment housing price. In the case of (5), the reported coefficient for pointsub500 is 484 whereas, in case of (6), the reported coefficient for polysub500 is 687. These results mean that if an apartment complex exits within the 500 meter buffer area from the neighboring subway stations, the average price of the apartment units in the complex goes up to around 4,840,000 or 6,870,000 Won when compared with the apartment units located outside the buffer area. As we can see, the difference between the two figures is 2,030,000 Won. We cannot say that the difference between the two figures is small.

Table 7 reports the results from the Davidson and MacKinnon test conducted as formulated in (7) and (8). As shown in Table 7, the hypothesis: $\eta = 0$ is rejected while the hypothesis: $\delta = 0$ is not. According to the interpretation rule suggested in Table 2, these results mean that a

Table 5. Estimation Results for Point based Model

mprice	Coef.	T	P>|t|
size	892.2769	32.36	0
age	-162.196	-5.53	0
dthh500	3051.672	9.92	0
dstep	-2600.38	-9.7	0
dheat	1896.79	5.17	0
dsamsung	-0.25156	-5.55	0
pointsub500	484.086	1.82	0.069
_cons	-2102.93	-1.86	0.063
Number of obs = 1023, F(7, 1015) =392.39, Prob > F = 0.0000, R-squared = 0.8084, Root MSE = 4310.2			

Table 6. Estimation Results for Polygon based Model

mprice	Coef.	T	P>|t|
size	889.755	32.08	0
age	-163.93	-5.58	0
dthh500	3016.732	9.89	0
dstep	-2605.49	-9.75	0
dheat	1795.346	4.79	0
dsamsung	-0.2542	-5.65	0
polysub500	686.7138	2.46	0.014
_cons	-2081.71	-1.85	0.064
Number of obs = 1023, F(7, 1015) = 404.68, Prob > F = 0.0000, R-squared = 0.8089, Root MSE = 4303.7			

Table 7. Estimation Results for the Davidson and MacKinnon Test

		Hypothesis: $\eta = 0$
		Rejected (t=1.96)
Hypothesis: $\delta = 0$	Not rejected (t=-0.66)	H_1 is acceptable H_0 is not

5) The independent variable, "dsamsung" is measured by the point based approach in the two competing models. I assume that the difference between two approaches in this variable is very small.

6) White Estimator (1980) is used to estimate (5) and (6) due to heteroskedasticity.

polygon based approach is superior to a point based approach in explaining variations in the apartment prices.

V. Conclusion

In this study, the relative performance of the point based approach and the polygon based approach was compared using the apartment housing market data from the north eastern part of the city of Seoul.

The results from this study indicate two things. First, two approaches can produce substantially different results in a hedonic price model estimation. Second, the polygon based approach produces a hedonic price model which explains price variations better than the point based approach.

These findings suggest that Korean researchers who are interested in improving quality of hedonic price model estimations and use GIS to measure the location variables for hedonic price models should consider using the polygon based approach with the point based approach. This is because the polygon based approach can produce the location variables with the shortest straight line distances and can explain the housing price variations well.

However, using the polygon based approach is not a simple task because currently all the functions for measuring distances in the GIS environment are designed for the point based approach. Thus, to use the polygon based approach actively, the GIS functions which allow the researchers to measure various distances from the polygon representing an apartment complex to many destinations should be developed.

References

[1] Choi, Changsik and Hyukyool Yoon, "Spatial Impact Analysis of the Seoul Subway Line 7 on Apartment Properties", Seoul Urban Studies, 2004, 5(4), pp.1-12

[2] Maddala, G. S., "Introduction to Econometrics", John Wiley & Sons, 2005, pp. 502-505.

[3] Seo, Kyung Chon, "Exploratory Analysis of Real Estate Price using Tight Coupling with GIS and Statistics", Journal of Korea Geographic Information Society, 2006. 9(3), pp.67-81.

[4] Sohn, Chul and Sang-Young Shin, "A Study on the Spatial Extent of Arae Under the Negative Price Impact from an Urban Resource Recovery Facility in Seoul, Korea", Journal of the Korea Real Estate Analyst Associations, 2007, 13(3), pp.117-127.

[5] White, Hal. "A Heteroskedasticity -Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", Econometrica, 1980, 48, pp.817-838.

Received (September 19, 2008)
Revised (December 22, 2008)
Accepted (December 23, 2008)