

능형회귀분석을 활용한 부동산 헤도닉 가격모형의 정확성 및 해석력 향상에 관한 연구 - 서울시 구로구 아파트를 대상으로 -

구본상* · 신병진¹

¹서울과학기술대학교 건설시스템공학과

Using Ridge Regression to Improve the Accuracy and Interpretation of the Hedonic Pricing Model : Focusing on apartments in Guro-gu, Seoul

Koo, Bonsang*, Shin, Byungjin¹

¹Department of Civil Engineering, Seoul National University of Science and Technology

Abstract : The Hedonic Pricing model is the predominant approach used today to model the effect of relevant factors on real estate prices. These factors include intrinsic elements of a property such as floor areas, number of rooms, and parking spaces. Also, The model also accounts for the impact of amenities or undesirable facilities of a property's value. In the latter case, euclidean distances are typically used as the parameter to represent the proximity and its impact on prices. However, in situations where multiple facilities exist, multi-colinearity may exist between these parameters, which can result in multi-regression models with erroneous coefficients. This research uses Variance Inflation Factors(VIF) and Ridge Regression to identify these errors and thus create more accurate and stable models. The techniques were applied to apartments in Guro-gu of Seoul, whose prices are impacted by subway stations as well as a public prison, a railway terminal and a digital complex. The VIF identified colinearity between variables representing the terminal and the digital complex as well as the latitudinal coordinates. The ridge regression showed the need to remove two of these variables. The case study demonstrated that the application of these techniques were critical in developing accurate and robust Hedonic Pricing models.

Keywords : Hedonic price model, Multi-colinearity, Variance Inflation Factor, Ridge Regression

1. 서론

1.1 연구의 배경 및 목적

국내 부동산 중 아파트는 국민 60%이상의 주거지역으로 자리 잡았으며, 국민 개인의 자산 중 평균 67.8%를 차지하고 있어(Statistics Korea 2014), 경제적 안정과 삶의 주거 수준에 지대한 영향을 미치고 있다.

이에 따라 지자체 및 지역 주민들은 아파트 가격 변동 및 영향 요인들에 예의주시한다. 이 중에서도 인근 지역에 이미

존재하거나 계획 예정인 입지시설이 가격에 주는 영향에 대해 매우 민감하고, 해당 입지시설의 특성에 따라 대비된 선호도를 보이고 있다.

공원, 지하철역 혹은 산업단지과 같이 부동산 가격의 긍정적인 선호시설에 대해서는 유치가 치열하다. 이른바 PIMPY(Please In My Front Yard)현상으로 최근 호남고속철도 노선을 놓고 대전시와 충남도가 대립한 것이나, 삼성의 승용차 공장을 놓고 대구 및 부산이 갈등을 빚은 경우가 대표적인 사례이다.

이에 비해 원자력 발전소, 쓰레기 소각장 등 비선호 또는 혐오시설에 대해서는 NIMBY(Not In My Back Yard)현상으로 지역 주민들이 강력하게 반대를 한다. 밀양 송전탑이나 양재동 분향소의 입지계획에 해당 지역 주민들이 강하게 반발한 사례를 들 수 있다.

두 현상 모두 지역 이기주의라는 비판을 받지만 그 만큼 국

* Corresponding author: Koo, Bonsang, Department of Civil Engineering, Seoul National University of Science and Technology, Seoul 139-743, Korea
E-mail: bonsang@seoultech.ac.kr
Received August 3, 2015; revised August 10, 2015
accepted August 20, 2015

민이 부동산 가격, 특히 아파트 가격에 영향을 미치는 사업에 민감하다는 방증이다.

이에 따라 국내에서는 입지시설이 아파트 가격에 영향을 미치는 인자에 대한 연구가 꾸준히 이어져 왔으며(Table 1), 선호 혹은 비선호 시설 종류와 가격의 상관관계를 정립하는 연구가 대부분이다. 이때 활용되는 대표적 기법이 헤도닉 가격 모형(Hedonic Pricing Model)인데, 이들 연구에서는 모형 적용 시 영향 인자(변수)들 간 발생할 수 있는 ‘다중공선성’(Multi-colinearity)에 대한 추가 분석을 명시적으로 하지 않고 있다.

입지시설과 부동산 가격의 상관관계를 파악할 경우, 해당 부동산과 입지시설간의 거리를 주요 변수로 사용하게 되는데, 해당 부동산에 영향을 주는 입지시설이 다수 존재할 경우에는 다중공선성이 더욱 두드러지게 발생할 수 있다.

더불어, 선형 회귀에 근간을 둔 헤도닉 가격 모형은 해당 시설의 영향 범위 및 경계를 파악하기 힘들어, 해석력이 떨어지는 한계를 지니고 있다. 이에 따라 입지시설의 영향을 파악하고자 할 경우, 헤도닉 가격 모형만으로는 분석의 신뢰도가 떨어지고, 결과에 대한 해석이 어려워진다.

본 연구에서는 이런 점을 개선하기 위한 방안으로 헤도닉 가격 모형과 함께 분산팽창지수(Variance Inflation Factor) 및 능형회귀(Ridge Regression)분석을 실시하였다. 분산팽창지수는 선형 회귀 분석에서 나온 결과에 대해 다중공선성 여부를 파악하는데 유용한 통계적 산정 수단이다. 이와 더불어 능형회귀 기법은 다중공선성이 존재하는 변수의 제거 여부를 판단하는데 활용될 수 있다.

본 연구에서는 실제 사례에 헤도닉 가격 모형과 더불어 상기 두 가지 기법을 적용함으로써 입지분석 수행 시 결과에 대한 정확성과 해석력을 높이는 수단으로 활용될 수 있음을 입증하는 것을 목적으로 한다.

1.2 연구의 범위 및 방법

상기 목적을 수행하기 위해서 다음과 같은 연구를 진행하였다.

첫째, 국내에서 헤도닉 가격 모형을 활용한 주요 연구들을 고찰하였고 이들 연구의 한계를 파악하였다.

둘째, 헤도닉 가격 모형의 한계와 이를 보완하기 위해 분산팽창지수와 능형회귀 기법의 이론적 고찰을 수행하였다.

셋째, 실제 사례에 적용하기 위해 데이터를 수집하였다. 사례는 서울시의 구로구의 아파트 가격을 대상으로 하였다.

구로구에는 대표적인 비선호시설인 서울 남부 구치소가 있으며, 동시에 선호시설인 전자단지가 입주해 있고, 가까운 지하철 역까지의 거리 등 입지시설의 영향을 파악하는데 좋은 대상이 되었다.

넷째, 본 데이터를 대상으로 일반적인 헤도닉 가격 모형을

구축한 뒤, 다중공선성의 존재 여부 및 해결을 위해 능분산팽창지수를 산정하고 능형회귀 기법의 실시를 통해 각 시설이 미치는 영향 정도 뿐 아니라 실제 영향 거리를 파악할 수 있었다.

본 연구는 주변 입지가 아파트 가격에 미치는 영향을 파악하는 것이 목적이므로 부동산 정책 등에 대한 영향 등을 배제하기 위해 단기간의 실거래가 가격으로만 분석하였다.

2. 이론적 고찰

2.1 부동산 가격 영향 인자에 관한 선행연구

2.1.1 선행 연구

부동산 가격 예측에 관한 연구는 과거부터 다양하게 진행되었으며, 이 중에서 헤도닉 가격모형을 이용한 가격 예측에 관한 최근의 국내 연구는 Table 1에 제시하였다.

Table 1. Selected research in real estate pricing methods

Title	Author (year)
The Hedonic Impacts Accrued from the Han-River Civic Park on the Surrounding Residential Areas	Yang S. D. and Choei N. Y. (2003)
Urban Non-preference Facilities Effect on Residential Real Estate Price : Focusing on Incineration Plant in Capital Area	Shin J. T. (2008)
A Study on the Effect of Environment-related Facilities : Focusing on Gangnam-Gu in Seoul	Kim M. H. (2010)
A study on the Effect of Improvement in Accessibility to Capital Area on the Sale Price of Apartment	Moon J. H. (2012)
The Analysis of Importance Ratio Factors of Sales Price for Apartment Building Construction Projects	Yang O. H., Kim M. S., Hwang U. S and Kim Y. S. (2011)
PSM based Price Estimating for Local Mixed-Use Apartment Development	Park J. K., Cho Y. K. and Lee S. Y. (2014)
A Study on the Analysis of Apartment Price affected by Urban Infrastructure System	Hwang S. D., Jeong M. O and Lee S. Y. (2015)

Yang and Choei(2003)는 한강시민공원이 주변 아파트에 대한 영향력을 헤도닉 가격모형을 통해 추정하여 한강시민공원과의 거리가 가까울수록 가격이 상승하지만 강변북로가 가까이 위치하면 교통 소음 등에 의해 평당 가격이 하락하는 현상이 파악되어 이에 대한 후속 연구가 필요하다는 것으로 결론을 맺었다.

Shin(2008)는 비선호시설의 입지선정시의 분쟁 해결을 위한 합리적이고 객관적인 자료를 얻기 위하여 비선호시설 입지가 부동산 가격에 미치는 영향범위에 대하여 헤도닉 가격 모형을 사용하여 정태적 분석을 실시하였다. 그 결과 비선호 시설인 소각장으로부터 3km 이상 떨어지면 소각장의 영향을

받지 않으며, 3km 범위 내에서도 거리가 멀어질수록 영향력이 약해지는 것으로 나타났다.

Kim(2009)은 강남구의 환경과 관련된 비선호시설과 선호시설이 주택가격에 미치는 영향에 대한 영향력을 정량적으로 밝히고자 하였고 자원회수시설은 거리가 멀어질수록 주택가격이 상승하고, 한강시민공원의 경우 멀어질수록 가격이 하락한다는 것을 검증하였다.

Moon(2012)는 춘천시의 아파트 가격모형을 통해 수도권 접근성이 개선(서울춘천고속도로, 경춘복선선철) 이후 주택 매매가격, 토지가격, 자산가치 전망지수 등 부동산 관련 지표가 상승함을 보였다.

이 연구들은 공통적으로 부동산 가격의 영향요인을 분석하는데 대표적으로 활용되는 헤도닉 가격 모형을 활용하고 있다.

최근의 연구들은 헤도닉 가격 모형 이외에 다양한 방법을 이용하고 있다. Yang et al.(2011)은 AHP기법을 활용하여 가격 결정에 대한 영향 요인을 도출한 후 중요도를 분석하였으며, 서울지역의 주택은 하나의 투자 방식으로 인식하는 지역적 특성으로 인해 투자가치요인, 정책요인, 원가요인이 주요 인자로 파악되었고, 충청지역(비수도권 지역)은 지역요인, 교통요인, 주택요인이 주요 영향인자로 파악되어 이는 삶의 질적 수준에 영향을 미치는 인프라 구축시 미진한 지역특성을 반영하고 있는 것으로 나타났다.

Park et al.(2014)은 PSM(Price Sensitivity Method)을 기반으로 주상복합 아파트의 분양가를 추정하였으며 주상복합 아파트의 경우 일반 아파트 보다 가격이 10.8% 높게 추정되었으며, 수요 탄력성을 추정하여 주상복합 아파트의 사업성 분석을 위한 기초자료로서 활용할 수 있다는 가능성을 제시하였다.

Hwang et al.(2015)은 도시기반시설(변전소)이 공동주택가격에 미치는 영향분석을 위해 다중회귀분석을 수행하고, 인공신경망을 통해 다중회귀모형을 검증하였다. 그 결과 지역적으로 변전소가 주변 아파트 가격에 미치지 않는 경우도 있으며, 변전소와 최단거리에 있음에도 가장 최근 입주한 아파트인 경우 변전소의 영향이 없는 것으로 나타났다.

2.1.2 선행 연구 고찰

상기 국내에서 진행된 대다수의 연구는 헤도닉 가격 모형을 기초로 하고 있으며, 특정 시설의 영향을 파악하기 위해서 해당 부동산과 입지시설의 인접거리를 변수로 활용하고 있다. 또한 대부분 선호 또는 비선호 시설 한 개를 대상으로 분석을 실시하였다.

그러나 다수의 입지시설이 존재한다면, 해당 부동산은 이들 모두에게 영향을 받게 되며, 이 경우 거리를 나타내는 변수들간의 다중공선성의 문제가 발생할 수 있는데, 선행 연구에서는 이에 대한 문제 제기나 해결방안을 명시적으로 제시

하지 않고 있다.

2.2 헤도닉 가격 모형

2.2.1 개요

헤도닉 가격모형(Hedonic Pricing Model, '쾌락모형')은 “재화의 가치는 해당 재화에 내포되어 있는 특성에 의해 결정된다.”는 가정 하에 Rosen(1974)에 의해 제시된 모델이다. 주택의 가격은 그것이 가지는 단순한 물리적 특성뿐 아니라, 환경적 특성들의 복합적 작용에 의해 결정된다는 가정 하에 아파트의 내재가치를 산정해 보고, 그에 근거한 잠재적 분양가를 결정하는 방식이다.

헤도닉 가격모형은 주택이라는 재화가 가지는 속성을 잘 반영하는 방법으로 분양가격의 산정은 물론 가격결정에 영향을 미치는 요소 파악 등 다양하게 사용될 수 있는 유용한 통계적 기법이라고 할 수 있다. 물리적 특성(평수, 향, 층, 방 수 등), 아파트가 속한 단지(용적률, 주차시설, 녹지 상태 등) 및 지역 특성(교육시설, 문화시설, 통학 편리성, 도심까지의 거리, 대기오염도 등) 다양한 특성을 지니는 복합재화인 아파트의 가치는 이러한 특성들 가치의 합이며, 궁극적으로 개별 소비자들이 주관적인 판단에 기초하여 개별 특성에 대하여 소비자들이 지불하고 있는 가치의 합이라고 볼 수 있다. 그러나 아파트의 개별 가치는 시장에서 관찰될 수 없으므로 이를 파악하기 위하여, 아파트들의 특성변수와 아파트의 분양가격과의 관계를 표시하는 것이며, 아래 식 (1)과 같은 다중회귀식의 형태로 표현한다.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n + \epsilon \quad (1)$$

여기서 Y는 주택 가격을 의미하는 종속변수이며 $x_1 \sim x_n$ 은 주택 가격에 영향을 미치는 인자들인 독립변수가 된다. 주택 가격은 각각의 독립변수들에 대한 계수 ($\beta_0, \beta_1, \beta_2 \dots$) 들의 크기를 통해 측정할 수 있다.

2.2.2 헤도닉 가격 모형의 한계

헤도닉 가격 모형은 주택의 가격이 단순한 물리적 특성뿐 아니라, 환경적 특성들의 복합적 작용에 의해 결정된다는 가정 하에 그 주택의 내재가치를 산정한다. 그러나 헤도닉 가격 모형은 선형 회귀분석에 기초하고 있어, 다음과 모형의 한계가 있다.

첫째, 주변 환경 시설의 영향 분석 시 인접 거리(Euclidean distance)에 따른 가격의 변화로 영향 정도를 파악한다. 그러나 다수의 비선호 및 선호시설이 존재할 경우에는 유사한 시설끼리 높은 상관성을 보이게 되는 문제가 있다. 통계학에서는 이를 다중공선성이라 하며, 이에 따라 헤도닉 모형을 적용할 경우 다중공선성에 의해 분석의 오류가 생기지 않도록 주

의해야 한다(Thériault and Rosiers 2013).

둘째, 선형 회귀 분석에서는 유의한 인자를 분별하는 데에는 효과적이지만, 해당 시설의 영향이 어디까지 미치는지는 그 경계(boundary)를 파악하기가 쉽지 않아, 해석력이 떨어지는 문제가 있다.

2.3 능형 회귀분석을 이용한 다중공선성 해결

2.3.1 분산팽창지수

다중회귀분석에서는 독립변수가 2개 이상 존재하게 되는데 이들 독립변수 간 서로 상관성이 높은 경우 다중공선성이 존재한다고 정의한다(Kim and Kim 2014).

다중회귀분석에서 추정하는 회귀계수(β)는 잔차의 최소 제곱합을 가장 작게 하는 최소제곱법(Ordinary Least Squares, OLS)을 이용하게 된다. 다중공선성이 존재하는 경우 최소제곱 추정치의 분산이 커지게 되며, 이에 따라 회귀모델의 신뢰성이 떨어지게 된다.

다중 공선성은 독립변수들 간 선형관계가 존재하거나 모델이 과적합(overfitting)된 경우에 발생하게 되며, 기본적으로 독립변수들 간 쌍별 상관분석(pairwise correlation)을 통해 진단할 수 있다. 그러나 변수들이 많아지면 변수들 간의 상관계수를 파악하는 데에는 무리가 있으며 이때 분산팽창지수(Variance Inflation Factor, 이하 VIF라 함)를 활용하기에 적합하다.

VIF는 의존변수를 제외한 독립변수 간의 R-squared 값을 이용하는 것이다. 즉, 개별 독립변수를 나머지 독립변수를 가지고 회귀분석을 실시하여 정확도 척도인 R-squared를 구하고 이를 가지고 상관성 정도를 파악하는 것이다. R-squared 값이 크면 이는 설명력이 큰 것이므로, 특정 독립변수가 나머지 독립변수에 의해 설명이 잘된다는 것이며, 이는 상관성이 큰 것으로 해석된다. 실제 VIF는 식 (2)와 같이 산출되며, R-squared가 크면 VIF도 큰 값을 갖게 된다.

$$VIF = \frac{1}{1-R^2} \quad (2)$$

일반적으로 VIF가 3-10이상인 경우 다중공선성이 존재하는 것으로 보며, VIF가 10이라면 모델에 포함된 다른 변수들에 의해 변동(Variation)의 90%가 설명된다는 것을 의미한다. 즉, VIF가 큰 변수가 포함되어 있다면 오히려 모델의 정확성이나 신뢰성이 떨어질 수 있다.

2.3.2 능형회귀분석

VIF는 다중공선성의 여부를 밝히는데 효율적이지만 해당 변수의 제거 여부를 알려주지는 않는다. 이를 위해서는 능형회귀분석을 추가로 수행해줘야 하며, Hoerl and Kennard(1970)이 처음 제안하였다.

일반적으로 회귀분석에서 개별 독립변수는 분산이 크고 공분산은 작은 것이 좋다. 개별 독립변수의 분산이 커야 의존변수의 영향정도를 파악하기 좋기 때문이다. 반면, 변수들 간의 공분산이 크다는 것은 상호 간 유사성이 높다는 것을 의미하는데, 공분산이 크면, 어느 독립변수가 의존변수에 영향을 주는지 파악하기 힘들어 진다. 그러므로 공분산은 작은 것이 좋다.

변수들의 분산 및 공분산을 행렬로 표현한 것을 variance-covariance matrix라 하며 식 (3)과 같은 형태를 띤다. 행렬에서 사선은 독립변수의 분산을, 양쪽은 동일한 값으로 변수들 간의 공분산을 표현한다. 설명력이 높은 회귀모델을 구축하기 위해서는 사선의 분산은 크고, 양가의 공분산은 작은 것이 좋다. 여기서 공분산이 커지면, 독립변수들 간의 상관성이 많다는 것으로 다중공선성의 존재를 의미한다.

$$var(Y|X=x) = \begin{pmatrix} \sigma_{Y_1 \cdot x}^2 & \sigma_{12 \cdot x} & \cdots & \sigma_{1p \cdot x} \\ \sigma_{21 \cdot x} & \sigma_{Y_2 \cdot x}^2 & \cdots & \sigma_{2p \cdot x} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1 \cdot x} & \sigma_{p2 \cdot x} & \cdots & \sigma_{Y_p \cdot x}^2 \end{pmatrix} \quad (3)$$

어느 회귀모델이나 다중공선성은 어느 정도 존재한다. 문제는 다중공선성이 회귀모델 계수(β)의 안전성(stability)에 얼마나 영향을 주는지를 알아야 한다.

이를 파악하기 위해서 고안해 낸 방법이 능형회귀기법이다. 능형회귀기법에서는 다중공선성이 존재할 경우, 식(4)에서와 같이 원 회귀계수(β^{OLS}) 대신 능형회귀계수(β^{ridge})로 대체하는데, 이는 각 독립변수의 분산을 λ 만큼 점진적으로 증가시켜 편향(bias)을 증가시킨다. 이 작은 변화에 회귀계수(β^{OLS})의 값이나 부호가 크게 바뀐다면, 이는 해당 독립변수가 다른 변수와 매우 큰 다중공선성이 존재한다는 것을 의미하며, 따라서 본 변수를 제거하는 것이 모델의 정확성에 도움을 준다.

$$\beta^{ridge} = \frac{\beta^{OLS}}{1+\lambda} \quad (2)$$

반면, λ 를 증가시켜도 회귀계수 값이 크게 변하지 않고 안정되어 있다면, 본 변수는 전체 모델에서 포함시켜도 무방하다. 특히 능형회귀분석에서는 λ 값에 따른 회귀계수들의 값을 도식화할 수 있어, 직관적으로 문제가 되는 변수들을 선별하는 데에 매우 유용하다.

3. 분석 방법 및 절차

3.1 분석 방법 및 대상

앞서 소개한 바와 같이 본 연구에서는 우선 헤도닉 가격 모형 구축을 위해 다중회귀분석을 실시한 후 VIF를 이용해 다

중공선성 여부를 파악하고 능형회귀분석을 통해 최종 변수들을 선별하였다.

분석 대상은 2014년 1분기 서울시 구로구 184개의 아파트에 대한 개별 가구 1,039건으로 하였다. 구로구에는 대표적인 비선호시설인 서울 남부 구치소가 있으며, 동시에 선호시설인 전자단지가 입주해 있고, 가까운 지하철 역까지의 거리 등 입지시설의 영향을 파악하는데 좋은 대상이 되었다(Fig. 1).

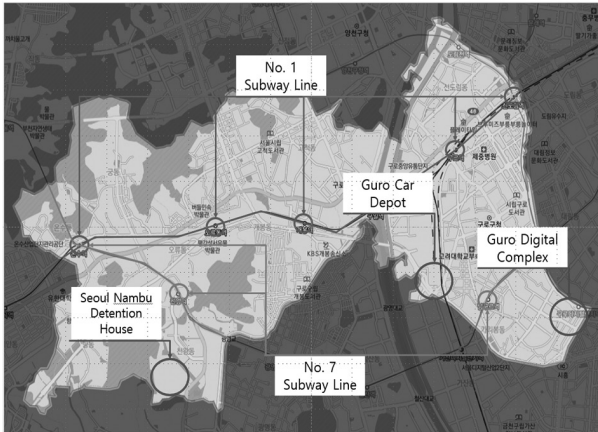


Fig. 1. Map of the Guro-gu and main facilities

3.2 헤도닉 가격 모형 구축

3.2.1 변수 선정

헤도닉 가격 모형 구축을 위해서는 종속변수인 개별 아파트 가격과 이 가격에 영향을 미치는 다수 인자들과 이들의 개별 데이터를 수집해야 한다.

2장의 기존 문헌고찰과 데이터 취득 여부에 따라 Table 2에서와 같이 변수들을 구축하였다. 가격에 영향을 미치는 변수는 내외부적 변수로 구분하였다. 여기서 내부적 변수는 연면적, 층수, 건축년도 등 아파트 자체가 가지고 있는 성질의 변수이며, 외부적 변수는 주변 환경적 특성으로서 아파트의 위치 좌표, 인접 시설과의 거리 등을 의미한다. 특히, 입지시설과 개별 아파트와의 거리를 dis_prison(서울남부구치소), dis_sub(아파트에서 가장 가까운 지하철역까지의 직선거리), dis_complex(구로 전자단지), dis_depot(전철 차량 기지)로 각각 표현하였다. 여기서, 남부구치소 및 전철 차량기지는 비선호시설(undesirable)이며, 지하철 및 전자단지는 선호시설(desirable)이다.

내부적 특성에 대한 자료는 부동산 포털사이트¹⁾에서 단지 정보(세대수, 건설사, 주차대수 등)에 대한 자료를 이용하였고, 외부적 변수는 GIS 전문회사인 'Biz-GIS'²⁾에서 제공하는

온라인 GIS 툴을 이용하여 좌표 및 인접 시설까지의 거리에 대한 데이터를 얻었다(Fig. 2). 마지막으로 종속변수인 아파트 가격은 국토교통부³⁾에서 고시하고 있는 아파트 실거래가 정보를 활용하였다.

Table 2. Variables for Hedonic Pricing Model

Type	name	unit	explanation
dependent variable	price	10,000 won	Real sale price of apartments
	type	-	Address "dong" (legal district)
	year	-	Construction year
	area	m ²	area for exclusive use
independent variable (internal)	story	-	story
	porch	-	Porch structure (terracing:1, corridor:0)
	households	-	total number of households
	brand	-	assessment system of construction ability rank of constructor
	pph	per household	Number of parking lots per household
	highest	-	Highest of apartment story
	lowest	-	Lowest of apartment story
	x	m	x-coordinate of apartment
	y	m	y-coordinate of apartment
	independent variable (external)	dis_prison (undesirable)	km
dis_sub (desirable)		km	linear distance to the nearest subway station
dis_complex (desirable)		km	linear distance to guro digital complex
dis_depot (undesirable)		km	linear distance to guro car depot

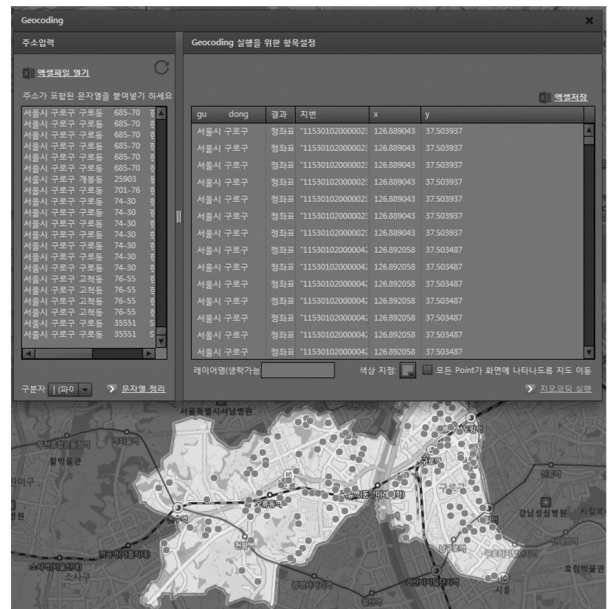


Fig. 2. Biz-gis tool for geo-coding apartment locations

1) <http://land.naver.com>
 2) <http://www.biz-gis.com/>

3) <http://rt.molit.go.kr/>

3.2.2 헤도닉 가격 모형 결과 및 문제점

Table 3은 모든 독립 변수를 이용하여 아파트 가격을 예측한 모형의 회귀분석 결과이다. 여기서 유효하지 않은 변수들인 최저 층고(lowest)와 경도(y)이다. 이들은 p-value (Pr)가 0.05 이상인 변수들로서 순차적으로 제거하면 Table 4와 같이 유효한 변수들만 남게 된다. 여기서 t value를 보면 알 수 있듯이, 전용면적(area), 건축년도(year), 아파트형식(type) 등과 같이 아파트 자체가 가지는 내부적 특성이 가격에 영향을 크게 미치는 것을 알 수 있다. 주변 입지시설은 모두 유효하게 나왔으며, 그 중 남부구치소(dis_prison)가 외곽에 위치해 있어 영향의 크기가 가장 작은 것을 알 수 있다. 두 번째 회귀 모형의 R-squared와 Adjusted R-squared는 각각 0.846, 0.843으로 높은 설명력을 보이고 있으며 초기 회귀모형보다 향상된 것을 알 수 있다.

Table 3. Regression results (all variables)

variable	Liner Regression (All data)				
	Estimate	Std. Error	t value	Pr(> t)	
Intercept	-2639000.00	269200.00	-9.81	0.0000	***
year	408.00	36.03	11.32	0.0000	***
area	291.40	7.09	41.10	0.0000	***
story	95.32	21.64	4.41	0.0000	***
porch	880.50	361.70	2.43	0.0151	*
type	5718.00	603.90	9.47	0.0000	***
households	1.16	0.29	3.99	0.0001	***
brand	1694.00	306.70	5.52	0.0000	***
pph	2887.00	412.50	7.00	0.0000	***
highest	-115.90	36.08	-3.21	0.0014	**
lowest	2.64	31.23	0.08	0.9327	
x	6.71	0.98	6.84	0.0000	***
y	0.30	0.55	0.54	0.5889	
dis_prison	1.23	0.68	1.82	0.0693	.
dis_sub	-5.62	0.66	-8.55	0.0000	***
dis_complex	3.72	1.24	2.99	0.0028	**
dis_depot	-5.99	0.71	-8.44	0.0000	***
R ²	0.8455		Adj. R ²	0.8430	

Table 4. Regression results (significant variables)

variable	Liner Regression 2				
	Estimate	Std. Error	t value	Pr(> t)	
Intercept	-2285000.00	185800.00	-12.30	0.0000	***
year	413.30	35.76	11.56	0.0000	***
area	290.90	7.09	41.04	0.0000	***
story	99.38	21.64	4.45	0.0000	***
porch	851.70	357.90	2.38	0.0175	*
type	5922.00	588.70	10.06	0.0000	***
households	1.21	0.27	4.41	0.0000	***
brand	1735.00	306.10	5.67	0.0000	***
pph	2909.00	409.90	7.10	0.0000	***
highest	-115.80	32.41	-3.57	0.0037	**
x	7.64	0.84	9.13	0.0000	***
dis_prison	0.89	0.44	2.03	0.0427	*
dis_sub	-5.76	0.65	-8.84	0.0000	***
dis_complex	5.75	0.33	10.81	0.0000	***
dis_depot	-6.22	0.70	-8.90	0.0000	***
R ²	0.8460		Adj. R ²	0.8433	

그러나 비선호시설인 전철 차량기지(dis_depot)의 계수가 음의 값을 갖는데, 이를 해석하면 아파트와 차량기지가 가까울수록 가격이 높다는 것이다. 이는 부호가 잘못된 것으로 볼 수 있으며, 다중공선성의 존재를 엿볼 수 있는 부분이다. 실제로 변수들 간 상관분석을 실시할 경우, 전철 차량기지(dis_depot)와 디지털 단지(dis_complex)의 상관계수가 0.869로 매우 높게 나오며 이는 두 변수의 산포도를 통해서도 확인할 수 있다(Fig. 3). 이와 같은 경우, 다중공선성 분석을 통해 올바른 변수의 선별이 필요하다.

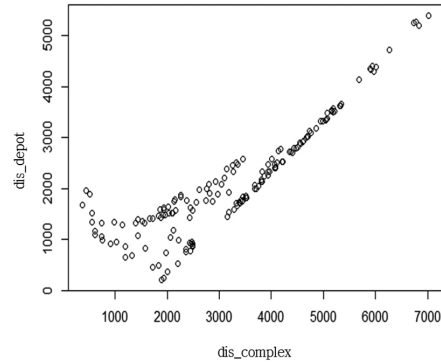


Fig. 3. Scatter plot of variables dis_depot and dis_complex

3.2.3 VIF 및 능형회귀 분석

Table 5는 모든 독립변수에 대해 VIF 및 능형회귀를 적용한 결과를 보여준다.

Model 1에서 보듯이, 디지털 단지(dis_complex), 위도(x) 및 차량기지(dis_depot)의 VIF가 가장 큰 것을 볼 수 있다. 이는 디지털 단지와 차량 기지가 구로구 내에서 비슷한 위치에 있으므로 아파트 가격에 미치는 영향이 유사하기 때문인 것으로 파악된다. 위도(x) 역시 구로구내에서 동-서 방향으로의 위치를 나타내기 때문에, 상기 두 변수와 다중공선성을 띄는 것으로 분석된다.

Table 5. Variable selection by VIF

variable	VIF		
	Model 1	Model 2	Model 3
year	3.21	3.07	3.04
area	1.38	1.38	1.36
story	1.31	1.31	1.31
porch	1.7	1.69	1.68
type	2.23	1.52	1.32
households	2.23	2.22	2.16
brand	1.51	1.48	1.43
pph	2.57	2.57	2.56
highest	3.28	3.25	2.77
lowest	2.26	2.19	2.17
y	20.96	3.92	2.14
dis_prison	35.6	29.72	4.38
dis_sub	1.67	1.58	1.4
dis_depot	33.72	30.44	4.14
x	171.98	67.01	
dis_complex	203.66	-	-
R-squared	0.8460	0.8446	0.8351

앞서 2.3.2절에서 소개한 바와 같이, 다중공선성을 줄이기 위해 독립변수의 제거 여부를 판단하기 위해서는 능형 회귀 분석을 추가로 수행해야 한다. 여기서는 λ 를 0-100까지 설정하여 점진적으로 증가시키면서 회귀계수(β)를 관찰하여 개별 변수 값의 변화 및 부호전환 여부를 파악하였다. 그 결과 Fig. 4와 같은 결과를 얻을 수 있었으며, 이는 단지, 위도, 기지 및 구치소가 불안정한 값들인 것을 가시화해준다.

이에 따라 디지털 단지(dis_complex) 및 위도(x)를 제거하면 Table 5의 Model 3처럼 VIF 값들이 낮아지고 모두 10이하로 줄어드는 것을 볼 수 있다. 또한 Model 3의 변수들로 다시 능형회귀를 실시하면 Fig. 4와 같이 도식화되며, 모든 변수들이 안정된 것을 파악할 수 있다. 즉, 다중공선성의 영향이 제거되었다고 할 수 있다.

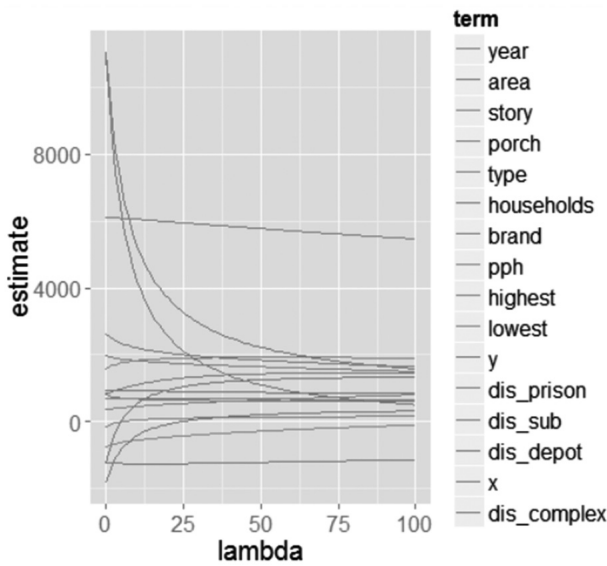


Fig. 4. Ridge regression for Model 1

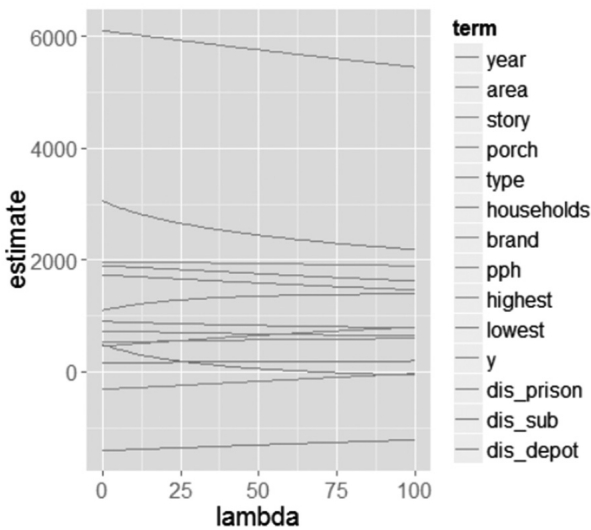


Fig. 5. Ridge regression for Model 3

3.2.4 최종 회귀 모형 결과

선별된 변수들을 대상으로 순차적 회귀분석을 다시 수행했을 때의 최종 결과는 Table 6과 같으며, R-squared 및 Adjusted R-squared는 각각 0.8275, 0.8251이다. 이 값은 능형회귀분석 전의 회귀모델, 즉 Table 4보다는 낮은 값이다. 그러나, 더욱 중요한 것은 입지시설인 남부구치소, 지하철 및 전철 차량기지 변수들의 부호가 정확한 것을 볼 수 있으며, 이는 설명력이 올바른 모델인 것을 알 수 있다.

3.2.5 최종 회귀 모형 결과 해석

Table 6에서 제시된 바와 같이, 아파트의 가격 결정에 가장 큰 영향을 미치는 내부적 인자는 건축년도(year), 전용면적(area)와 아파트 형식(type)이다.

인접시설인 남부구치소, 지하철 및 차량 기지 역시 유의하게 가격에 영향을 미치고 있다. 특히 지하철은 가까울수록, 구치소 및 차량기지는 멀수록 가격 상승효과가 있어 선호 및 비선호시설의 영향을 올바르게 모델링하고 있다.

Table 6. Regression results (Final)

variable	Liner Regression 3				
	Estimate	Std. Error	t value	Pr(> t)	
Intercept	-1056000.00	121300.00	-8.71	0.0000	***
year	292.40	34.01	8.60	0.0000	***
area	290.30	7.41	39.15	0.0000	***
story	108.40	22.79	4.76	0.0000	***
porch	1150.00	378.90	3.04	0.0025	**
type	6678.00	584.40	11.43	0.0000	***
households	1.41	0.30	4.69	0.0000	***
brand	1003.00	313.80	3.20	0.0014	**
pph	3416.00	422.80	8.08	0.0000	***
highest	-115.80	32.41	-3.57	0.0037	***
y	1.05	0.21	5.07	0.0000	***
dis_prison	1.62	0.12	12.98	0.0000	***
dis_sub	-4.39	0.64	-6.91	0.0000	***
dis_depot	2.16	0.28	7.76	0.0000	***
R ²	0.8275		Adj. R ²	0.8251	

인접 시설의 계수를 비교할 경우, 지하철(dis_sub)이 구치소(dis_prison)에 비해 가격에 약 4배 정도로 크게 작용한다는 것을 알 수 있다. 또한 지하철의 거리가 차량기지에 비해 약 2배 이상 영향을 미치고 있다. 즉, 비선호 시설이 미치는 영향이 선호시설인 지하철에 비해 그 영향이 크지 않다는 것을 알 수 있다.

이를 가격으로 환산하면 구치소(dis_prison)변수의 경우 평균적으로 100m씩 멀어질 때마다 아파트의 가격은 169만원씩 상승한다는 것을 의미하고, 지하철(dis_sub)변수는 100m씩 가까워질수록 평균적으로 약 550만원씩 해당 아파트의 가격이 상승한다는 것을 의미한다. 그만큼 대표적 선호시설인 역세권이 비선호시설보다 아파트 가격에 큰 영향을 준다는 것을 알 수 있다.

5. 결론

헤도닉 가격 모형은 부동산 가격과 이에 미치는 영향 인자들을 다중 회귀 모형의 형태로 구축하여 가격을 산정 내지 예측하는데 활용되는 대표적인 부동산 모형이다. 기존 연구에서 보는 바와 같이 부동산과 인접한 선호 및 비선호 시설의 관계는 해당 부동산과의 인접거리를 변수로 하여 그 영향 정도를 파악하고 있다.

그러나 헤도닉 가격 모형에서는 입지 시설이 선호 또는 비선호 시설인지 인지하거나 구분하지 못한다. 이로 인해 다수의 인접시설이 서로 근거리에 위치할 경우, 해당 변수 간 다중공선성이 발생할 수 있고, 이에 따라 설명력이 떨어지는 헤도닉 모형 결과를 초래할 수 있다.

본 연구에서는 이 경우, VIF 및 능형회귀분석을 통해 문제가 되는 변수들을 선별하고 이를 보완할 수 있는 방법을 소개하였으며, 실제 사례를 통해 그 방안의 유효성을 입증하였다. 또한 이와 같은 추가 분석을 통해 선호 및 비선호 시설의 분석을 동일 모형에서 수행할 수 있다는 것을 보여줬다.

본 연구에서 제시한 접근 방법과 모형은 NIMBY, PIMFY와 같은 문제에 직면한 공공 및 민간사업주체들에게 입지시설의 영향 정도를 정량화해 줌으로서, 균형 있고 객관적인 의사결정을 하는데 도움이 될 것으로 판단된다.

감사의 글

이 연구는 서울과학기술대학교 교내연구비의 지원으로 수행되었습니다(2015-1167).

References

Alan Agresti and Barbara Finlay., (2008). *Statistical Methods for the Social Sciences*, 4th ed, Pearson.

Biz-GIS., X-Ray Map for Web <<http://www.biz-gis.com/XRayMap/>> (Dec. 15. 2014).

Hoerl, E. and Kennard, R., (1970). "Ridge Regression: Biased Estimation for Nonorthogonal Problems" *Technometrics*, 12(1), pp. 55-67.

Hwang S. D., Jeong M. O and Lee S. Y., (2015). "A Study on the Analysis of Apartment Price affected by Urban Infrastructure System," *Korean journal of Construction Engineering and Management*, KICEM, 16(1), pp. 127-136.

Kim J. E. and Kim Y. S. (2014). "An Analysis of Profitability Study of Overseas Construction Projects

using Multiple-Regression" *Korean journal of Construction Engineering and Management*, KICEM, 15(2), pp. 95-103.

Kim, M. H., (2010). "A study on the Effect of Environment-related Facilities on the Surrounding Housing Price : Focusing on Gangnam-Gu in Seoul" MS thesis, Dankook University.

Marius Thériault, François Des Rosiers., (2013). *Modeling Urban Dynamics*, John Wiley & Sons.

MOLIT, Apartment Real Transaction Price <<http://rt.molit.go.kr/>> (Dec. 15. 2014).

Moon, J. H., (2012). "A study on the Effect of Improvement in Accessibility to Capital Area on the Sale Price of Apartment : Focusing the Chuncheon City in Gangwon province", MS thesis, Konkuk University.

Naver corp., Naver Real Estate <<http://land.naver.com/>> (Dec. 15. 2014).

Park J. K., Cho Y. K. and Lee S. Y., (2014). "PSM based Price Estimating for Local Mixed-Use Apartment Development." *Korean journal of Construction Engineering and Management*, KICEM, 15(4), pp. 86-94.

Rosen, S., (1974). "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition" *Journal of Political Economy*, The University of Chicago Press, 82(1), pp. 34-55.

Rudolf J. Freund, William J. Wilson, Ping Sa (2006). "Regression Analysis", *Academic Press, Mathematics*.

Shin, J. T., (2008). "Urban Non-preference Facilities Effect on Residential Real Estate Price : Focusing on Incineration Plant in Capital Area," *Real Estate Focus*, 4, pp. 22-34.

Yang O. H., Kim M. S., Hwang U. S and Kim Y. S., (2011) "The Analysis of Importance Ratio Factors of Sales Price for Apartment Building Construction Projects," *Korean journal of Construction Engineering and Management*, KICEM, 12(5), pp. 127-136.

Yang, S. D. and Choi. N. Y., (2003). "The Hedonic Impacts Accrued from the Han-ver Civic Park on the Surrounding Residential Areas : The Case of Tuk-Som River Park" *Journal of Korea Planners Association*, 38(3), pp. 275-285.

요약 : 헤도닉 가격 모형은 부동산 가격에 영향을 미치는 여러 요소를 모델링하는데 활용되는 대표적 방법이다. 부동산 가격은 전용 면적, 방의 개수, 주차공간과 같은 내재적 속성 뿐 아니라 주변 신호/비선호시설의 존재여부에 따라 영향을 받는다. 주변 입지시설의 경우, 그 영향을 파악하기 위해서는 해당 부동산과의 인접거리를 설명변수로 사용하게 된다. 그러나 다수의 입지시설이 인접해 있는 경우에는 설명 변수 간 다중공선성이 발생하는 문제가 존재한다. 본 연구에서는 분산팽창지수 및 능형회귀분석을 이용해 다중공선성을 파악하고 유의한 설명변수를 선별하는데에 활용하였다. 이들 기법을 서울시 구로구 아파트들에 적용한 결과, 전철 차량 기지, 디지털 단지 및 위도에 해당하는 변수간의 다중공선성을 파악하였으며, 능형회귀분석을 통해 적합한 변수들을 체계적으로 선정할 수 있었다. 본 사례를 통해 상기 기법들이 더 정확하고 적절한 헤도닉 가격 모형을 구축하는데 중요한 보완적 기능을 해 준다는 것을 알 수 있다.

키워드 : 헤도닉 가격 모형, 다중공선성, 분산팽창지수, 능형회귀 분석
