

Lasso 알고리즘 적용을 통한 교통환경의 주택가격 영향력 분석

Impact of Transportation Environment on Housing Prices with Lasso Algorithm

채오성* · 박진우** · 김형주*** · 김정화***

Chae, Osung · Park, Jinwoo · Kim, Hyungjoo · Kim, Junghwa****

Abstract

Looking at domestic economic trends, per capita gross national income increased by approximately 8% in five years, but the actual transaction price index of Seoul condominiums doubled from approximately 88.6% to 155% for the first time in five years. If such a steep upward trend continues, it may cause the polarization of income groups, regional imbalances, and slowdown of economic growth, so studies and policy proposals on the factors influencing the rise in housing prices are needed. In this study, the factors that impact house prices are classified by characteristics, and their influence is quantified. Additionally, we conducted basic research on the development of a house price model that considers traffic environment factors as well as urban characteristics, in comparison with the results of the Hedonic model, often used as an existing house price determination model. The spatial range in this study is based on the city/county area of 31 Gyeonggi Province, and the first Gyeonggi Province quality of life questionnaire survey material, were used to quantify traffic environment and population characteristics. For geographic information, we collected the administrative area status data of the Prayer Agency (2019), and in the case of housing prices, the housing price data presented by the Ministry of Land, Infrastructure, Transport and Tourism through the public system (2020). From the collected data, the optimal house price model was selected through the turbine principle, the house price determinants were selected, and the influence of the variables were compared. As a result, the second phase new urban area and other regional characteristic variables were revealed as variables that had a major influence on house prices. Of the various traffic environmental characteristics, only traffic safety did not

* 경기대학교 도시·교통공학과 학부과정 (Kyonggi University · codhtjd@kyonggi.ac.kr)

** 경기대학교 도시·교통공학과 석사과정 (Kyonggi University · ref4545@kyonggi.ac.kr)

*** 차세대융합기술연구원 선임연구원 (Advanced Institute of Convergence Technology · hyungjoo@snu.ac.kr)

**** 경기대학교 도시·교통공학과 조교수 (Corresponding Author · Kyonggi University · junghwa.kim@kyonggi.ac.kr)

impact house prices. As the traffic congestion increases toward the city center, the satisfaction level for private use, which is more inconvenient than public transportation, tends to be inversely proportional to the house price. The better the accessibility to the facilities, the higher the house price. Finally, as a result of comparing the performance of the Hedonic model and the Lasso model via R, the performance of the Lasso model was approximately twice as good as that of the Hedonic model. However, since this study analyzed the house price determinants only with the house price data of a specific area, additional studies are needed to analyze the house price determinants of a wide range of areas where the limits are clear.

| Key words | 주택가격, R, Lasso 알고리즘, 교통환경, 만족도, 신도시

House Price, R, Lasso Algorithm, Transport Environment, Aatisfaction, New Town

1. 서론

1.1 연구의 배경과 목적

주택가격의 급변은 경제에 큰 영향을 미치며, 급격한 상승세가 지속될 경우 소득계층 양극화 및 지역적 불균형, 경제성장의 둔화를 일으킬 수 있다. 반대로 예상하지 못한 부동산 가격의 급락은 경제위기를 일으킨다고 한다(안지희, 2020). 이처럼 부동산 가격은 경제에 민감하게 작용할 수 있어 부동산 가격을 구성하는 요소를 정확히 이해하고 정책을 결정하는 것이 중요하다. 이 중 부동산의 큰 부분을 차지하는 주택의 가격 변동은 국민들의 복지 인식에 깊은 영향을 미칠 수 있다. 주택가격이 상승할수록 현 복지상태에 만족하고 복지확대에 대한 증세에 부정적으로 느끼는 반면, 주택가격이 낮아질수록 현 복지상태에 불만족하고 복지확대에 대한 증세에 긍정적인 태도를 보이는 것이다(조남경 외, 2020). 상쇄이론에서부터 알 수 있듯 예로부터 주거입지를 선택하는 데 있어 주택가격이 굉장히 중요한 요소로 작용하며, 주택가격은 다양한 결정요인들에 의해 구성된다. 과거에는 주택가격에 영향을 미칠 수 있는 요소가 적었지만, 사회가 발전함에 따라 더욱더 다양한 요인들이 주택가격에 영향을 미칠 수 있게 되었고, 일반적인 지리적 환경적 요소 뿐만아니라, 주변 교통환경 또한 중요한 요소로 작용할 수 있다는 것이다(Alonso, 1964; Mills, 1967; Muth, 1969).

최근 서울의 주택시장 현황을 살펴보면, 매년 주택가격의 변동은 상승 폭이 다름지언정 꾸준히 우상향하고 있는 것을 알 수 있다. 서울을 비롯한 경기 일부 지역과 주요 광역시들의 집값도 상승하는 추세를 보이는 반면, 상대적으로 지역 기반산업이 열악하고 교통환경이 열악한 지방의

경우 대부분 집값이 떨어지고 있는 추세이다. 집값 변동으로 인한 지역적 불균형과 소득계층의 양극화는 소득, 자산, 교육, 주거, 문화, 건강 등 다양한 차원의 불평등으로 연결되고 이는 개별적으로 작용하는 것이 아닌 상호작용함으로써 서로를 가속 시킨다. 이러한 사회적 불평등문제는 날이 갈수록 심각해지고 있다(황규성 · 이재경, 2014).

공간적 불평등을 야기할 수 있는 주택가격에 영향을 미치는 요인들을 정의하고 비교한 연구는 꾸준히 되어왔다(이준용, 2007; 김희호 · 박상운, 2013; Funke & Paetz, 2013; Chihiro et al., 2016; Yu Wei, 2017). 주택가격모형을 고려할 때 공간적 특성, 시계열 등의 특성을 반영하여 보다 구체적인 주택가격에 미치는 영향을 분석한 연구 또한 존재 한다(Kim and Park, 2008; Youn and Lee, 2013; Park, 2013; Ju, 2018; Wang et al., 2019). 빅데이터 시대가 도래하면서 많은 양의 데이터가 축적되고 있지만 이를 사람이 일일이 분류하고 분석하는 데는 한계가 존재한다. 또한, 설문조사데이터의 경우 데이터의 양이 많아짐에 따라 변수간 다중공선성이 있는 경우가 다반사이며, 선형회귀모형이 선형을 띄지 않을 수 있다. 이러한 데이터 분석의 한계를 극복하기 위해 머신러닝이라는 방법이 개발되었고 다양한 분야에서 사용되고 있다. 하지만 머신러닝의 경우 이용자에게 중간과정을 보여주지 않는 방식이기에 통계적인 의미를 포함하고 있지 않아 인과관계를 설명하기 힘들다. 본 연구에서는 Lasso 모델을 기반으로 한 주택가격모형을 통해 주택가격에 영향을 미치는 요인들을 분석하고 이를 주택가격의 일반적 모델로 많이 쓰이는 Hedonic 모형과 결과를 비교해보고자 한다.

1.2 연구 내용

현재 정책을 고려하거나 향후 계획을 설계하기 위해 주택가격 연구들(Ko and Park, 2010; Park, 2013)을 기반으로 주택가격에 교통적 영향을 고려한 연구들(Lee, 2010; Lee, 2012; Kim et al., 2016; Hou, 2017; Wen. et al., 2018; Ping, 2018; Bae et al., 2018)을 통해 주택가격에 영향을 미치는 요인들을 특성별로 분류하고 영향력을 수치화하였다. 또한, 기존 Hedonic 모형 결과와 비교를 통해 교통환경요소와 도시의 특성이 고려된 주택가격모형 개발의 기초 연구를 시행하고자 하였다.

본 연구에서의 공간적 범위는 경기도 31개 시·군 지역을 기준으로 하였으며, 교통환경과 인구 특성을 수치화하기 위해 제1차 경기도민 삶의 질 설문조사 자료를 사용하였다. 지리정보는 기도청의 행정구역 현황자료(2019), 주택가격의 경우, 국토교통부 실거래가 공개시스템(2020)을 통해 국토교통부에서 제시하는 연면적 그룹 중 하나인 면적 $85\text{m}^2 \sim 102\text{m}^2$ 에 해당되는 주택가격 데이터만을 수집하였다. 21년도 이후 주택가격 자료의 경우, 주택가격 변동의 폭이 다른 해에

비해 비정상적으로 클 뿐만 아니라, 코로나로 인한 경기 악화 및 정책적 변동이 심한 탓에 이러한 부분들까지 주택가격 모형에 전부 반영할 수 없어 20년도 자료를 활용하였다. 경기도민 데이터의 평균 가구원 수가 약 2.78명으로 대한민국 평균 1인 주거면적인 32.9m^2 곱하게 되면 약 91.4m^2 으로 면적 $85\text{m}^2 \sim 102\text{m}^2$ 그룹에 포함되기 때문에 이를 수집하였다. 수집한 데이터로는 주택가격에 영향을 미칠만한 요인들을 선정하여 거주자특성, 지역 특성, 교통환경특성으로 분류하였다. 최종적으로 R을 사용하여 가장 적합한 주택가격 Lasso 알고리즘 기반의 모형 구축을 진행하였으며 기존에 많이 활용되어 오던 Hedonic 접근법과도 비교분석을 시행하였다.

2. 선행연구 고찰

기존 주택가격모형의 한계를 제시하고 극복책을 제안한 선행연구로 최명섭 & 변세일(2016)은 서울시 아파트 1,991채 주택의 평균 매매가격을 기반으로 최소자승법, 공간시차모형, 공간오차모형, 지리적 가중회귀모형 그리고 지리적 가법모형과 같이 다양한 주택가격 추정모형들의 RMSE를 도출하여 예측력을 비교하였다. “주택가격의 예측을 높이기 위해서는 공간자기상관 구조와 연속형 변수의 비선형구조를 동시에 반영한 모형의 개발이 필요하다”라고 주장하였다.

주택가격을 결정하는 요소를 선정하고 분류하기 위해 이강(2016)은 서울시 이천동 지역의 주택을 대상으로 Hedonic 모형을 기반으로 한 주택가격 결정요인을 분석하였으며 그 결과, 토지의 부동성과 외부환경의 변화로 인해 주택가격이 변동하는데 토지의 부동성으로 인한 주택의 고정성이 주택가격 변동에 상당한 영향을 미치는 것으로 나타났다. 실제 주택가격 변화양상 원인을 해석하기 위해 김희호(2013)는 서울시 강남구와 노원구 주택을 대상으로 가격에 대한 Hedonic 모형과 정매 데이터를 이용하여 GLS 추정기법과 Hedonic 모형을 기반으로 한 회귀분석을 제시하였다. 이 과정에서 지역 간의 특성에 따라 주택가격에 영향을 미치는 요인의 영향력 차이가 발생할 수 있음을 밝혀냈다. 주택가격모형에 대한 예측력과 성능을 높이기 위해 이준용 외 2인(2007)은 서울시 강남구와 서초구의 1,584채 아파트 매매가격을 기반으로 SAS EMINER을 사용하여 의사결정나무와 회귀분석 그리고 신경망 분석을 실시하였다. 이를 통해 새로운 모형을 구축하여 아파트 가격의 결정요인을 선정하고 RMSE를 이용하여 모형의 예측력을 파악하였다. 아파트 가격에 대한 예측력을 높이기 위해서는 데이터마이닝의 한가지 기법만이 아닌 다양한 기법들의 장점을 사용하여 복합적으로 이용해야 한다는 것을 증명하였다.

주택가격모형에 대한 해외의 연구동향을 검토하였으며 Yu Wei(2017)은 중국의 주요 30개 도시의 집값 상승률을 보다 정확하게 예측하기 위해 DMA(Dynamic 모형 averaging)산출 방식

을 수용하였다. 시간에 따른 해당 계수와 예측모형이 실시간으로 변할 수 있음을 전제로 모형을 구축하였고, MCS Test를 실시하여 다른 정적인 모델들보다 높은 예측력을 가짐을 증명하였다. Funke & Paetz(2013)은 Iacoviello(2004), Parias and Notarpietro(2008), Calza et al.(2009) 등 다양한 선행연구들을 기반으로 2 Agent(기업, 가구), 2 Sector(직장, 집), 거시경제를 전제로 한 모형을 구축한 후 베이즈 법칙을 이용하여 구축한 모형으로부터 홍콩의 주택시장가격을 예측하였다. 분석결과, 가격상승으로 인한 매입의 두려움이나 경제적인 마찰로 인해 가격이 변동하기 보다는 시간적 선호도에 의해 움직인다는 것을 보여주었다. Sean et al.(2006)은 1975년부터 2003년까지의 알래스카와 하와이를 제외한 미국의 연간 주택가격 데이터를 기반으로 실제 주택가격이 소득으로 인한 마찰 등 충격에 의해 움직이는 정도를 시장 간의 상호작용을 고려하여 분석하였다. 최종적으로 지역 특성이나 거시경제에 따른 주택가격속도를 도출하였다. 이를 통해 집값과 소득 사이의 상호관계가 존재함을 증명하였다. Chihiro et al.(2010)은 일본의 부동산버블과 붕괴의 시기를 포함한 1986년부터 2008년까지 47만 건의 주택가격 데이터를 각각 Hedonic 모형과 반복판매지수를 이용하여 분석하였다. 분석결과 2002년 초에 콘도미니엄 가격의 Hedonic 측정치는 바닥을 친 반면, 반복 판매 수치는 2004년 봄에만 바닥을 보였다. 이처럼 주택가격 지수는 추정방법에 따라 다르게 나올 수 있으며, 이러한 불일치는 완전히 제거할 수 없다고 주장하였다. Zhen et al.(2018)은 주택가격 모형을 의사결정나무를 결합하여 더 강력한 모델을 만들어내는데 사용되는 GBRT(Gradient Boosted Regression Tree)의 과적합 방지를 위해 사용되는 XGBoost 알고리즘을 이용하여 10가지 요인에 대한 주택가격 예측모형을 제시하였다. 10가지 요인으로서는 지리적 위치, 가구 비율, 층수, 주거형태, 면적, 집방향, 데코레이션, 엘리베이터 유무, 인근 지하차 유무, 가구 수를 사용하였다. 분석 결과, 쓰촨성 청두 및 진창구 주택가의 중고주택 가격이 가장 높게 나타났고 직장인들의 거주지역일수록 높게 나타났다. 의사결정 모델과 기존의 다중선형회귀 헤도닉 모형 보다 여러 회차 부문에서 성능이 높게 평가됐으며, 불균형한 데이터 집단의 과적합을 충분히 해결하였다고 주장하였다.

전술한 선행연구 조사를 통해 국내외로 주택가격모형에 대한 연구는 꾸준히 진행되어 왔음을 확인하였다. 해외의 경우 다양한 주택가격모형을 제시하고 적용하며 기존 방법론들과 비교 분석해보는 연구가 많았지만, 국내의 경우는 Hedonic 모형을 기반으로 한 분석 연구가 대부분이었고 모형의 개발에 관한 연구는 다소 이루어지지 않았다. Hedonic 모형은 주택가격과 결정요인들 간의 인과관계를 쉽고 명확하게 보여줄 수 있지만 다른 모형들에 비해 가격에 대한 예측력이 낮을뿐더러 현실에 존재하는 다양한 변수를 고려할 순 없다. 따라서 본 연구는 공간자기상관 구조와 연속형 변수의 비선형구조를 동시에 반영하는 주택가격모형 개발의 기초 연구로써 의의를 가진다고 볼 수 있다. 주택가격 결정요인들은 전술의 선행연구들을 참고하여 주택가격 형성에

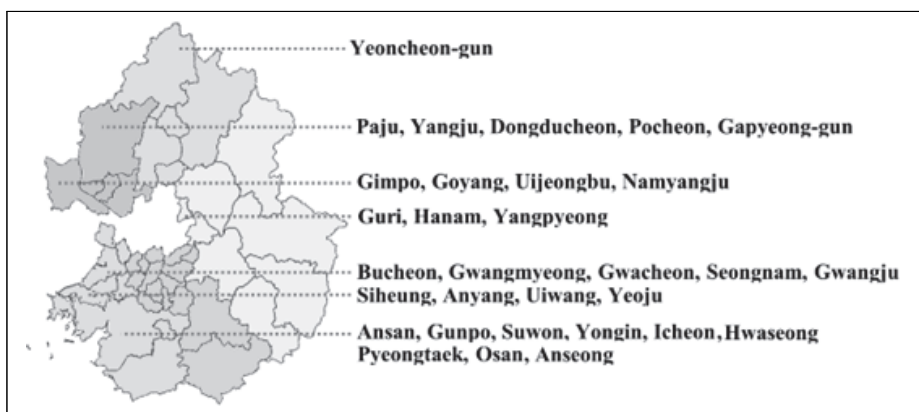
중요한 영향을 미치는 지역특성, 시간적선호도, 가구소득 등을 선정하였다. 추가적으로 교통환경에 대한 만족도 및 안전성을 추가하여 경기도민들의 심리요소까지 고려하고자 한다,

3. 연구방법론

3.1 데이터 수집

경기도 31개 지역(그림 4 참조)에 거주하는 경기도민들의 특성을 지리, 주택가격, 교통환경, 인구특성별로 분류하여 데이터를 수집하였다. 각각의 데이터는 다음의 항목들을 포함하고 있다.

- 지리데이터: 경기도 행정구역 현황(법정동 코드 데이터: 시/군/구 번호 앞 6자리 + 법정동 번호 뒷 4자리, ex.1111016000, 서울시 종로구 연지동)
- 주택가격데이터: 국토교통부 실거래가(아파트, 연립/다세대, 단독/다가구 - 단지, 전용면적, 거래금액 등)
- 교통환경데이터: 제1차 경기도민 삶의 질 조사(거주지 교통환경만족도, 편의시설 편도 소요시간, 거주지 교통안전도 등)
- 인구데이터: 제1차 경기도민 삶의 질 설문조사(행정구역 코드, 주소, 주택형태, 가구원특성, 월평균 가구 소득, 학력 등)



〈그림 1〉 연구 공간적 범위

연구에 필요한 자료를 종합한 결과, 약 2만 명의 가구 데이터와 30,706건의 거래데이터가 집계되었다. 수집된 자료는 연구 진행 시점에서 구득 가능한 가장 최신의 데이터를 사용하였으며, 자료는 경기도 행정구역 현황(2020), 국토교통부 실거래가(2017~2020), 제1차 경기도민 삶의 질 조사(2017)를 사용하여 분석을 실행했다.

3.2 데이터 전처리 및 변수선정

설문조사 데이터는 조사 시기가 2016년이기에 당시 행정구역의 현황이 행정구역의 통합 등으로 인해 현재 상태와 달라 이용이 어려워, 법정동 코드데이터를 사용하여 이를 보완하였다. 종속 변수인 주택가격은 실거래가를 평(당 가격으로 환산한 수치를 사용하였다. 또한, 경제성장률에 따른 주택가격의 변화를 간접적으로 반영할 수 있도록 3년 동안의 거래데이터를 사용하였다. 거주자 특성 중 성별, 혼인 상태, 직업 유·무, 학력, 주택형태, 전체 가구원 수들을 명목형 변수로 설정하였으며, 연령, 가구 소득은 수치형 변수로 범주화하였다. 연령은 그룹을 6개로 분류(10대, 20대, 30대, 40대, 50대, 60대 이상), 가구소득 또한 그룹을 6개로 분류(100만원 미만, 100만원 이상 200만원 미만, 200만원 이상 300만원 미만, 300만원 이상 400만원 미만, 400만원 이상 500만원 미만, 500만원 이상) 하였다. 지역특성은 거주지의 행정구역을 1기 신도시, 2기 신도시, 그 외 지역으로 분류하여 각각 명목형 변수로 설정하였다. 교통환경특성 중 편의시설까지 소요시간은 일상용품 구매점, 의료기관, 공공기관, 생활문화 기반시설, 근린시설, 복지시설까지 소요시간의 평균값을 산출하여 사용하였다. 교통 안전도는 주거지역 보행로, 상가지역 보행로, 학교주변 통학로, 어린이 보호구역, 횡단보도 및 건널목 등에서 느끼는 교통 안전도의 평균값을 사용하였다. 교통수단 이용만족도는 자가용과 대중교통 그룹으로 분류하여 대중교통 이용만족도의 경우 버스와 지하철 만족도의 평균값을 사용하였다. Lasso 모형의 최종 데이터 처리 양식은 <표 1>과 같다.

본 연구는 R을 통해 실거래가 데이터와 설문조사 자료를 활용하여 주택가격 결정요인을 선정하고, 변수들의 영향력을 비교한다. 최종적으로 Hedonic 모형과 Lasso 모형을 비교하여 최적의 주택가격 결정 모형을 제시하고자 한다.

〈표 1〉 Lasso 모형 데이터 프레임 셋

구 분			처 리
주택평균가격			(만원/평)
거주자 특성	성 별		남 = 1, 여 = 0
	연 령		10대 = 1, 20대 = 2, 30대 = 3 40대 = 4, 50대 = 5, 60대 이상 = 6
	혼인상태		혼인 = 1, 미혼 = 0
	직업 유·무		직업 있음 = 1, 직업 없음 = 0
	학 력		고졸 이상 = 1, 고졸 미만 = 0
	가구소득		0~99 = 0, 100~199 = 1 200~299 = 2, 300~399 = 3 400~499 = 4, 500이상 = 5
	주 택 형 태	단독	거주 = 1, 미거주 = 0
		다가구	거주 = 1, 미거주 = 0
		아파트	거주 = 1, 미거주 = 0
		기타	거주 = 1, 미거주 = 0
	전체 가구원수		다인가구 = 1, 1인 가구 = 0
지 역	1기 신도시		포함 됨 = 1, 포함 안 됨 = 0
	2기 신도시		포함 됨 = 1, 포함 안 됨 = 0
	그 외 지역		포함 됨 = 1, 포함 안 됨 = 0
교 통 환 경	편의시설 까지 소요시간		(분)
	교 통 안전도		(주거지역 보행로 + 상가지역 보행로 + 학교주변 통학로 + 어린이 보호구역 + 횡단보도 및 건널목)/ 5
	이 용 만족도	자가용	매우 불만족 = 1, 약간 불만족 = 2 약간 만족 = 3, 매우 만족 = 4
		지역 대중교통	(버스 이용 만족도 + 지하철 이용 만족도)/2

4. 모형추정

4.1 Lasso Algorithm의 적용

Lasso는 Least Absolute Shrinkage and Selection Operator의 약자로, 지구물리학에서 모형 생성에 있어 변수선택 한계를 극복하고자 만든 알고리즘이다(Tibshirani, 1986). 본 연구에서는 Lasso Algorithm을 이용하여 Hedonic 모형을 구성하였고, 이를 Lasso 모형이라고 지칭하였다. Lasso 모형의 기본형태는 식 1과 같다.

$$Y = \beta_0 + X\beta_1 + e$$

Y: 종속변수
 X: 설명변수
 β_1 : 설명변수의 계수
 β_0 : 상수항의 계수
 e: 오차

식 (1)

Lasso 모형의 프레임워크를 살펴보면 RSS선정, Shrinkage 조정, Subset selection, 최적 회귀모형 선정의 총 4단계의 과정을 통해 이루어진다. RSS은 Residual Sum of Square의 약자로, 잔차의 제곱의 합을 선형으로 만드는 단계이다. <그림 2> Shrinkage 조정은 Lasso 회귀분석의 회귀계수 추정원리를 따르는데, LASSO는 회귀계수 β 에 대한 제약 조건이 있으므로, 회귀계수(β)추정은 식 2와 같이 나타낼 수 있다(Duffyand Santner,1989).

$$\widehat{\beta}^{lasso} = \arg \min_{\beta} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^P x_{ij}\beta_j)^2$$

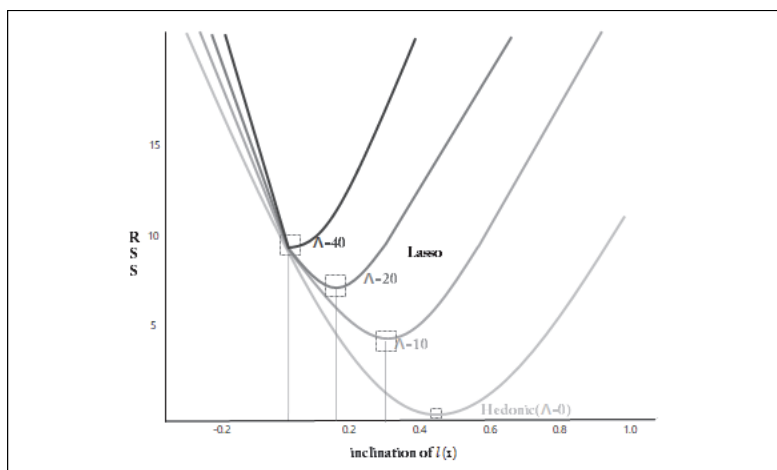
subject to $\sum_{j=1}^P |\beta_j| \leq s (s \geq 0)$

$\widehat{\beta}^{lasso}$: residual error를 최소화하는 적정 회귀계수
 y_{ij} : 종속변수
 x_{ij} : 설명변수
 β_j : 설명변수의 계수
 β_0 : 상수항의 계수
 s: 임계기준 모수

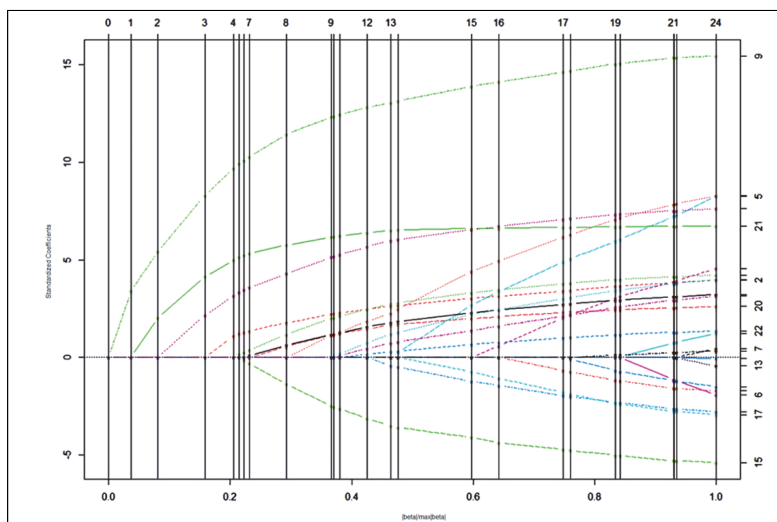
식 (2)

$\arg \min_{\beta} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^P x_{ij}\beta_j)^2$ 은 $\arg \min l(\beta)$ 로 나타낼수 있으며, 잔차의 제곱을 최소화하는 선형모형을 의미한다. shrinkage란 변수들의 영향력을 조정하는 행위를 말한다. β_j 값이 s보다 작거나 같아야 하는데, s의 값이 매우 큰 경우 일반적인 선형 모형을 띄고, s의 값이 작은 경우 적절히 shrinkage된 모형이라 할 수 있다. 정규화 매개변수 λ 에 의해 shrinkage 정도가 조절되는데, 이때 0이 되는 변수들이 존재하기 때문에, Subset selection이 된다고 할 수 있다.

결국, 이 s를 결정한다는 것은 Lasso 모형에 있어서, ‘몇 개의 독립변수를 쓸 것이다.’와 같은 문제로 s는 Shrinkage를 시행한 횟수(사용된 변수+1)와 같다. Lasso 모형의 Subset selection 과정은 <그림 3>에서 나타내고 있다. y축은 Shrinkage를 시행한 횟수를 의미하고, x축은 shrinkage의 비율(사용된 변수/총 변수)을 나타낸다.



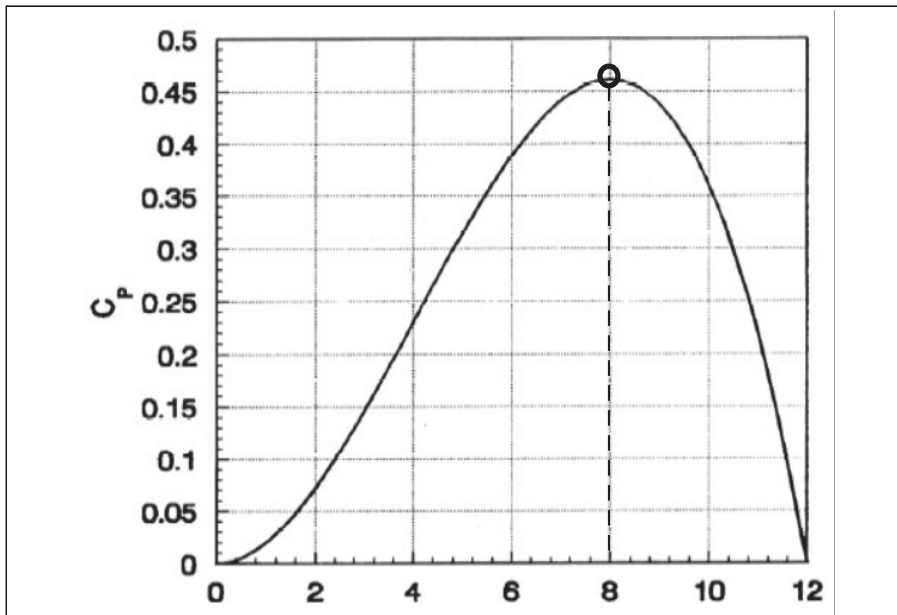
〈그림 2〉 Residual sum of squares step applied in Lasso



〈그림 3〉 Subset selection step applied in Lasso

본 연구의 Lasso 모형은 마지막 회차 시 모델을 선정하는 기존 Lasso 모형(Robert, T. 1997 “The lasso Method for Variable Selection in the Cox Model” Statistics in Medicine .)과는 달리, $CP-\lambda$ Curve의 터빈원리를 이용한 최적 모형선정단계를 추가한다. 터빈원리에서는 터빈의 날개가 특정 설계 끝단속도비(TSR, 날개끝단속도와 풍속 사이의 비율로 정의)에서 최적의 작동을 위해 설계될려면 예상되는 모든 끝단속도비에 대한 로터의 성능을 결정해야 한다고 주장한다(Marwa Ayadi et al. 2015). 터빈원리에서 사용되는 $CP-\lambda$ Curve를 이용하여 Lasso 모형

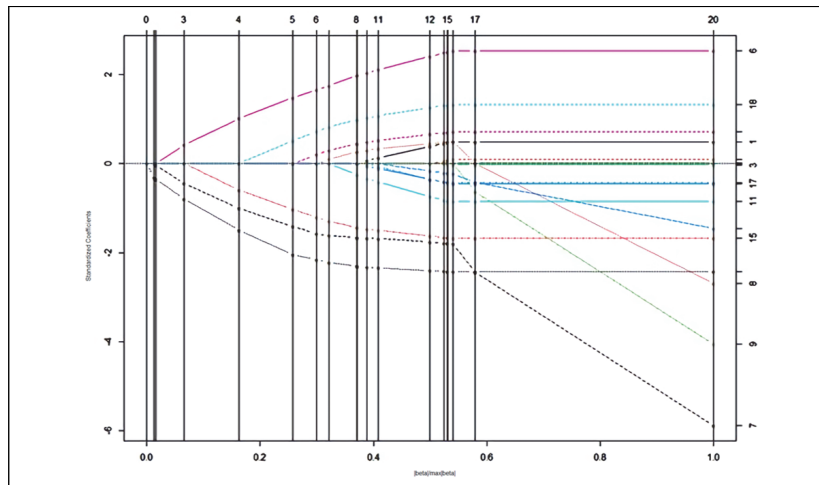
에서는 λ 값 사이 어딘가에 존재하는 최소 CP와 최소 CP가 추출되는 최적 λ 값을 유추할 수 있게 된다. <그림 4>는 터빈원리를 구체적으로 설명한다. x축은 람다, y축은 CP(Coefficient power)를 나타내며, $\lambda = 0$ 에서는 로터가 회전하지 않으므로 바람에서 동력을 추출할 수 없다. 또한, $\lambda = 12$ 에서 로터는 너무 빠르게 작동하여 바람에 의해 완전히 막히므로 동력을 추출할 수 없다. $\lambda = 0$ 과 $\lambda = 12$ 사이 어딘가에 최대 전력이 추출되는 최적값 $\lambda = 8$ 이 존재한다. 이러한 원리를 통해 잔차의 제곱을 최소화 하는 최적 CP를 도출할 수 있다.



<그림 4> Representative CP- λ curve (turbine principle)

4.2 속성 선택

본 연구의 Lasso 모형은 shrinkage 과정에서 비선형 $l(x)$ 의 잔차를 최소화 만드는 $l'(x)$ 를 찾는 기존 Lasso 회귀 식을 따른다. 정규화 매개변수 λ 에 따라 shrinkage 정도가 달라지는데, shrinkage의 회차가 증가함에 따라 선택되는 변수의 개수는 단계적으로 증가한다. 이 과정에서 변수의 영향력은 0에서 증가하거나 감소할 수 있지만 변동하지 않을 수 있다. 또한, 변수가 영향을 미치는 상태에서 영향을 미치지 않는 상태로 조정될 수 있다. 이러한 과정을 거쳐 $l(x)$ 의 최대 shrinkage를 시행한 후 변수들의 영향력 변화양상을 분석하였다(그림 5).



〈그림 5〉 Attributes selection based on 20 shrinkage rounds using Lasso

그래프의 x축은 $\beta / \text{Max} |\beta|$ ($0 \leq \beta / \text{Max} |\beta| \leq 1$), y축은 표준화된 계수들의 영향력을 나타낸다. shrinkage 정도는 정규화 매개변수 λ 에 따라 그 정도가 변화하기 때문에 λ 에 따른 계수들의 영향력 변화양상을 분석하는 것이 일반적이나, 본 연구에서는 $\beta / \text{Max} |\beta|$ 즉, 현재 회차 계수들의 합/최대 회차 계수들의 합을 λ 대신 사용하였다. $\beta / \text{Max} |\beta|$ 를 사용할 경우 변수들의 영향력 변화양상을 λ 를 사용했을 때보다 세부적으로 분석할 수 있기 때문이다.

shrinkage 1회차 시에는 어떠한 변수도 주택 평균 실거래가에 영향을 미치지 않는다. 2회차 시 신도시 외 지역 특성이 부정적인 영향을 미치기 시작하며, 회차가 증가함에 따라 다양한 특성들이 주택가격에 영향을 미치기 시작한다. 다가구 특성의 경우, 17회차까지 주택가격에 긍정적 영향을 미치다가 20회차까지 아무런 영향을 미치지 않는다. 마지막 회차 시에는 부정적인 영향을

: Non-zero coefficient

Trial	Selected	Sex	Age	Married	Job	Education	Income	1 House	2+ House	Apartment	Etc	Furniture	1 New City	2 New City	Other Site	Convenience TT	Traffic Safety Car	Satisfaction	Public Satisfaction	Constant	House Price
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
4	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
5	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
6	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
7	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
8	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
9	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
10	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
11	10	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
12	11	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
13	12	75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
14	13	88	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
15	14	91	0.3	-3	-81	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
16	15	91	0.3	-4	-81	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
17	16	95	0.3	-10	-84	-2	130	-259	92	0	-111	-192	0	469	-807	-25	-3	-47	144	1	1
18	17	95	0.3	-10	-84	-2	130	-351	0	-92	-203	-192	0.1	469	-807	-25	-3	-47	144	1	1
19	18	95	0.3	-10	-84	-2	130	-351	0	-92	-203	-192	0.1	469	-807	-25	-3	-47	144	1	1
20	19	95	0.3	-10	-84	-2	130	-351	0	-92	-203	-192	0.1	469	-807	-25	-3	-47	144	1	1
21	20	95	0.3	-10	-84	-2	130	-847	-496	-588	-699	-192	0.1	469	-807	-25	-3	-47	144	1	1

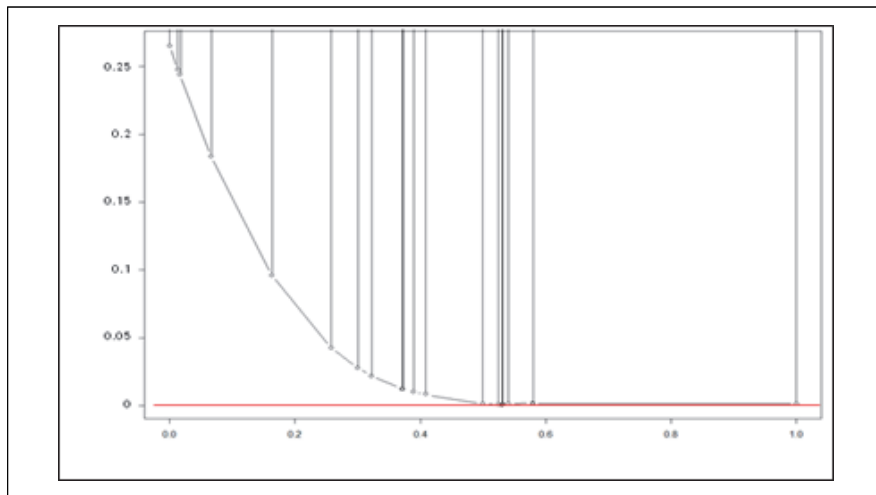
〈그림 6〉 Effects of variables on housing prices during shrinkage rounds (blue background indicates non-zero coefficients)

미치게 된다. 상수항의 영향력은 18회차까지 변동하지 않다가 19회차부터는 영향력이 변동하기 시작한다. shrinkage 회차에 따른 변수들의 영향력 수치는 다음 <그림 6>과 같다.

4.3 모델 계산

shrinkage 과정을 통해 비선형 $l(x)$ 의 잔차를 최소로 만드는 $l'(x)$ 들을 찾아 총 21개의 shrinkage된 Lasso 모형을 회차별로 도출하였다. 본 Lasso 모형에서는 잔차의 크기로 CP를 표현하였기 때문에 CP가 최소를 가져야 가장 적절하게 shrinkage된 모형이라 할 수 있다. Marwa Ayadi의 터빈원리를 적용해 CP- λ Curve를 이용하여 λ 값 대신 $\beta/\text{Max}|\beta|$ 을 통해 사이어딘가에 존재하는 최소 CP와 최소 CP가 추출되는 최적 지점들을 산출하였다.

<그림 7>을 통해 shrinkage 회차에 따라 CP 값이 변동함을 알 수 있다. shrinkage 회차가 지날수록 CP가 감소하는 추세를 보이지만, 10회 차까지는 CP가 감소하다가 11회차부터는 CP가 증가하거나 감소하기 시작한다. 최소 CP를 도출한 결과 shrinkage 15회차 시, 14개의 속성을 선택했을 때, 최소 CP(0.09982)를 가짐을 알 수 있다.

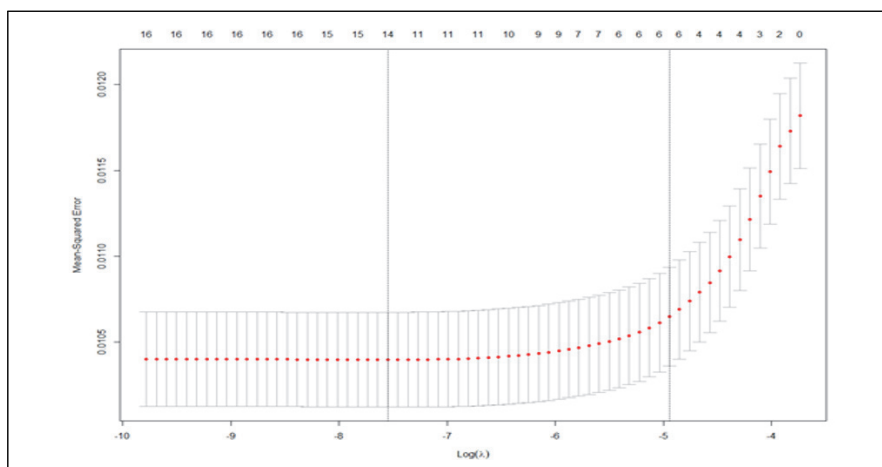


<그림 7> CP - $\beta/\text{Max}|\beta|$ curve

MSE(Mean Square Error)란 잔차의 제곱에 대한 평균을 취한 값으로 통계적 추정의 정확성에 대한 질적인 척도로 많이 사용되고 그 값이 적을수록 추정의 정확성이 높다고 할 수 있다. MSE의 변화양상을 곡선으로 보기 위해 본 모델에서는 λ 값 대신 $\text{Log}(\lambda)$ 를 사용하여 shrinkage 회차

별 λ 값에 따른 MSE 변화 폭을 산출하였다.

〈그림 8〉의 x축은 $\text{Log}(\lambda)$, y축은 MSE를 의미한다. 가로 회색 선은 shrinkage 회차에 따른 MSE 수치를 나타내며, 세로 회색 선은 λ 값에 따른 MSE 변화폭을 나타낸다. 빨간 점은 변동하는 MSE들의 평균 수치를 의미한다. 15회차 시, 최소 MSE(0.010409091)를 가지는 $\text{Log}(\lambda)$ 를 산출하였다. 그 결과 $\text{Log}(\lambda) = -7.62999999913...$, $\lambda = 0.0000000234422882...$ 임을 확인할 수 있었다. 최종적으로 산출한 최소 MSE를 통해 RMSE를 산출하여 0.10202495..로 0.5보다 작음을 확인하였다. RMSE는 MAE보다 이상점에 대해 더 민감하게 반응하기 때문에 오차를 검증할 때 더욱 효과적이라 할 수 있으며 0에 가까울수록 완벽한 모델이라 할 수 있다. 해당 모델에서는 RMSE는 0.5를 기준으로 반올림하여 0.5 이상의 수치는 1, 미만의 수치는 0으로 처리하여 해당 모델을 유효하다고 판단 하였다.



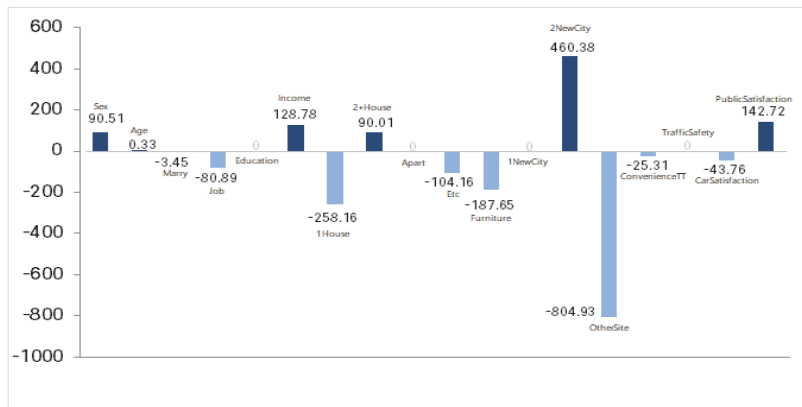
〈그림 8〉 Lasso 모형 MSE calculation

4.4 Lasso 모형 영향력 분석

총 표본 2만 개 중 11개의 결측값을 모델에서 제외한 후 속성들의 영향력을 분석하였다. RSS선정, Shrinkage 조정, Subset selection, MSE평가의 총 4단계의 절차를 거친 후 가장 성능이 좋은 주택가격 결정 모델의 속성별 영향력을 도출하였다. 최적 모델에 대한 속성들의 영향력은 다음 〈그림 9〉와 같다.

2기 신도시 지역과 그 외 지역 속성들이 각각 460.38, -804.93(만원/평)으로, 주택가격에 큰 영향을 미치는 것으로 나타났다. 2기 신도시의 주택 위치 여부는 주택가격에 긍정적 영향을, 그

외 지역의 주택 위치 여부는 부정적 영향을 미쳤다. 1기 신도시의 주택 위치 여부는 1기 신도시 건설로부터 시간이 많이 흐름에 따라 1기 신도시가 주변 지역의 특성을 포함하게 되면서 주택가격에 큰 영향을 미치지 않는 것으로 분석됐다. 교통환경특성 중 교통안전도의 주택가격에 대한 영향력은 없는 것으로 나타났는데 이는 도시지역의 교통안전에 대한 불감증이 반영된 결과라 할 수 있다. 교통안전을 강조하면 사회적 비용이 발생하게 된다. 안전요소의 특성상 편익은 금방 눈에 드러나는 성과가 없으므로 투자가 필요함에 불구하고 개인투자자들의 경우 투자저해요소로 받아들이는 경우가 많다. 교통수단 이용만족도의 경우, 도심으로 갈수록 교통혼잡도가 높아지기 때문에, 대중교통보다 이용이 불편한 자가용의 이용만족도가 주택가격과 반비례하는 경향을 보였다. 편의시설로의 접근성은 걸리는 소요시간이 적을수록 주택가격에 긍정적인 영향을 미쳤다.



〈그림 9〉 Analysis of the influence of determinants on housing prices

〈표 2〉 Analysis of the influence of determinants on housing prices

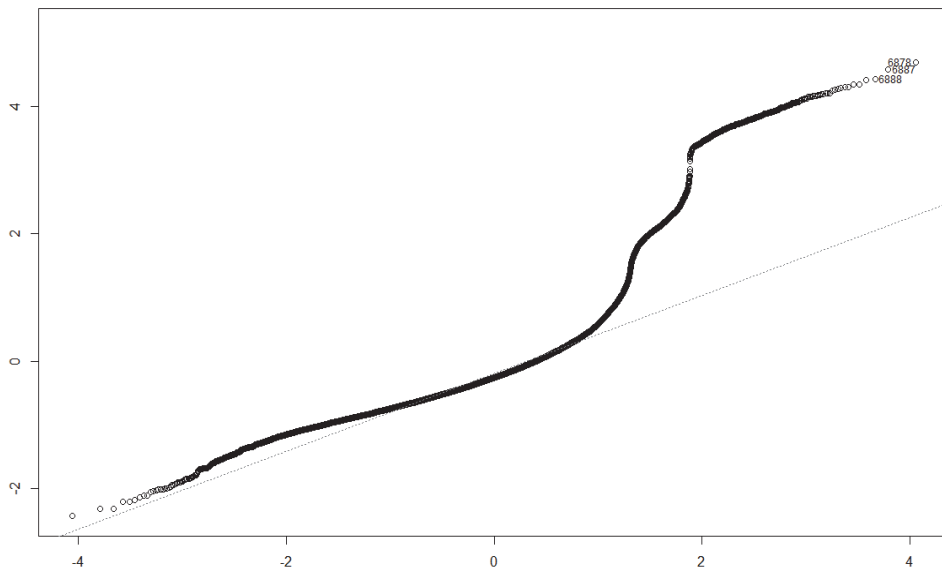
SEX	AGE	MARRY	JOB	EDUCATION	INCOME	1 HOUSE	2+ HOUSE	APART
90.51	0.33	-3.45	-80.89	0	128.78	-258.16	90.01	90.01
ETC	FURNITURE	1 NEW CITY	2 NEW CITY	OTHER SITE	CONVENIENCE TT	TRAFFIC SAFETY	CAR SATISFACTION	PUBLIC SATISFACTION
-104.16	-187.65	0	460.38	-804.93	-25.31	0	-43.76	142.72

4.5 Hedonic 모형과 비교

Hedonic 모형은 환경재에 대한 시장이 뚜렷하게 존재하지 않기 때문에, 시장 재인 주택, 노동 등을 이용하여 간접적으로 환경재에 대한 가치를 추정하는 방법이다. 본 Lasso 모형은 간접적으

로 환경재의 가치를 추정하는 면에서 Hedonic 모형과 같지만, 변수 간 존재할 수 있는 영향력을 고려하여 현실을 반영한다는 점이 다르다 할 수 있다. Hedonic 모형은 모형의 전제 특성상 변수를 독립적인 존재로 인식하고, 변수 간에 존재할 수 있는 영향력을 배제하기 때문에, 직관적이고 또렷한 해석은 할 수 있지만, 이는 시시각각 다양한 요소에 의해 변동하는 현실의 주택가격을 설명하는 데 있어 명확한 한계점이 존재한다.

본 주택가격모형의 Hedonic 모형 생성을 위해 본 연구의 Hedonic 모형은 Q-Q Plot을 이용하여 정규화된 표본들의 근사선을 통해 최적의 기울기를 찾아왔습니다. 이를 선형모형으로 만들었다. (figure 9) 그래프에서 x축은 표준화된 잔차를 나타내고 y축은 이론적 분위수를 의미한다. 모든 표본의 평균적인 기울기를 구하게 되면, 특정 속성을 과대해석하거나 과소해석할 수 있으므로 표본의 수가 많은 지점을 위주로 우선순위를 배열하였다. 또한, 표본이 근사선에 가까울수록 헤도닉 모형의 정규성을 만족하기 때문에 이를 최적 Hedonic 모형으로 선정하였다. Hedonic 모형의 MSE를 산출한 결과 0.03997120735로, RMSE가 0.199928005419로 측정되었다. RMSE가 0.5 미만이므로 유효한 모델이라 할 수 있지만, 변수 간에 존재할 수 있는 영향력을 배제하였기 때문에 <그림 10>과 같이 오른쪽 위 그래프 지점을 과소해석한 것으로 나타났다. Hedonic 모형과 Lasso 모형의 모형 성능을 비교한 결과 RMSE가 2배가량 감소한 것으로 나타나, 모형의 성능이 올라간 것을 알 수 있었다<표 3>. <그림 11> 그래프에서 두 모형 간의 영향력을 비교하였을 때, Hedonic 모형은 지역 변화에 따른 주택가격의 차이를 크게 보지 않는 반면, Lasso 모형에서



<그림 10> Hedonic 모형

는 지형변화에 따른 주택가격의 차이를 크게 보고 있다. 또한, 거주자 특성 중에서 가구주의 연령과 나이의 영향력을 서로 반대로 해석하고 있다. 또한, Hedonic 모형은 교통안전 만족도가 높을수록 집값이 저렴하게 나타났지만, Lasso 모형에선 교통안전 만족도의 영향은 없는 것으로 나타났다. 이처럼 변수들의 영향력을 개별적으로 설명할때 Hedonic은, 부정적인 측면의 성향을 극대화하고, 긍정적인 수치를 최소화 하여 이를 비교 하는 성향이 강하지만, Lasso 모형의 경우, Lasso algorithm을 통해 선형을 띄지 않는 집단을 사전에 전처리하여 비선형 모형을 구축하기에 긍정적인 영향이나 부정적인 영향을 최소화하거나 극대화해서 해석하지 않는다.

〈표 3〉 Comparison between Lasso and hedonic models

	Hedonic 모형	Lasso 모형
MSE	0.039	0.01
RMSE	0.199	0.102



〈그림 11〉 Hedonic & Lasso Coefficient Comparison

5. 결론 및 향후 연구과제

본 연구는 주택가격에 영향을 미치는 요인들을 특성별로 분류하고 영향력을 수치화하여 기존 hedonic 모형과는 다른 교통환경요소와 도시의 특성이 고려된 주택가격모형을 제시하였다. 분석 결과, Lasso 모형의 성능이 Hedonic 모형보다 약 2배 정도 우수한 것으로 나타났다. 속성들의

영향력을 살펴보면, 2기 신도시의 지리적 위치는 주택가격을 상승시키는 긍정적 요인으로 작용하였고, 그 외 지역의 지리적 위치는 주택가격을 하락시키는 부정적 요인으로 작용하였다. 1기 신도시의 지리적위치의 경우, 신도시 건설 이후 오랜 시간이 지남에 따라 주변 지역의 특성을 포함하게 되면서 주택가격에 큰 영향을 미치지 않는 것으로 나타났다. 교통환경특성 중 교통안전도가 주택가격에 대한 영향력이 없는 것으로 나타났는데, 이는 도시지역 거주민들의 교통환경에 대한 안전불감증을 간접적으로 보여줄 수 있는 지표로 교통환경에 대한 이해와 안전교육이 필요할 것으로 보인다. 교통수단 이용만족도의 경우, 자가용 이용만족도가 주택가격과 반비례하는 경향을 보였는데, 이는 복잡한 도심부일수록 혼잡한 교통상황으로 인해 자가용보다 대중교통을 선호하는 현상 때문이라 할 수 있다. 또한, 편의시설 접근성은 통행시간이 줄어들수록 주택가격에 긍정적인 영향을 미친 것으로 보아 집약적인 서비스 형태를 선호하는 것을 알 수 있었다.

국내 경제 동향을 살펴보면, 1인당 국민 총소득은 5년 동안 약 8% 증가했지만, 서울 아파트 실거래가격 지수는 약 88.6%에서 5년 만에 155%로 약 2배가량 증가했다(KOSIS, 2020). 이러한 급격한 상승세가 지속될 경우 소득계층 양극화 및 지역적 불균형, 경제성장의 둔화를 일으킬 수 있기 때문에 주택가격상승요인에 관한 연구와 정책적 방안이 필요한 상황이다. 정책적 방안을 위한 주택가격상승요인을 파악하기 위해선 단순히 주택가격의 변동양상만을 분석하는 것이 아닌, 다양한 외부요소를 고려한 회귀모형이 필요하다. 하지만 기존 Hedonic모형의 경우, 변수를 독립적인 존재로 인식하고, 변수 간에 존재할 수 있는 영향력을 배제하기 때문에, 시시각각 변동하는 현실의 주택가격을 설명하는 데 있어 한계점이 명확하다. 본 Lasso모형은 간접적으로 환경재에 대한 가치를 추정하고 변수들의 상호작용으로 발생할 수 있는 현실 주택가격의 변동성을 고려하기 때문에, 다양한 분야의 정책요소가 고려되어야 하는 주택정책에 도움을 줄 수 있는 모델이라 할 수 있다. 예를 들어, 주택가격 안정화를 위해서는 주택가격 자체를 안정시키는 수단보다 지역 편중화되어있는 대중교통시설이나 편의시설 등을 공평하게 배분함으로써 보다 효과적으로 주택가격을 안정시킬 수 있다는 것이다.

본 연구는 경기도 31개 지역의 주택가격데이터만을 가지고 주택가격 결정요인을 분석하였기에, 다른 지역에 Lasso 모형을 적용할 경우 그 한계가 존재하고 또한 교통환경요소를 반영하기 위해 거주민들의 설문조사 데이터만을 사용하였기에 통행시간 데이터, 교통규제 현황 등의 추가적인 교통환경요소들이 고려될 필요가 있다. 또한, 21년도 이후의 주택가격 데이터를 다루지 않았기에 코로나 이후 주택가격 정책으로 인한 주택가격 변동 요인을 분석하지 않았다. 주택가격을 결정하는 요인에는 교통환경 이외에도 반영되어야 할 부분이 많다는 것으로, 다양한 요인의 검토 또한 추가 되어야 한다. 마지막으로, Lasso 모형은 현실에 나타나는 다양한 요인들의 효과를 구체적으로 비교 분석할 수 있지만, 요인들 간의 인과관계를 대략적으로 비교하여 유추할 수밖에

없다는 점에서 그 한계가 명확하다. 요인들 간의 인과관계를 명확히 하는 것은 정책을 결정할 때 있어 중요한 근거로써 작용하기에 향후 연구에서는 공간적 자기분석을 포함한 알고리즘과의 결합을 통해 인과관계를 설명할 수 있는 효과적인 예측력을 가진 주택가격모형 개발에 관한 연구를 다룰 예정이다.

Acknowledgement

참고문헌

1. Ahn J-H.(2020), "A Study on Housing Price Movement", Korea University Graduate School, Korea.
2. Bae, S-Y., Chung, E-C., Lee S-Y.(2018), "Effects of Urban Railway Transportation Services on Housing Prices: Case of Apartments in Gyeonggi Province", The Journal of the Korea Real Estate Society, vol.24, no.3, pp.85~98.
3. Cho N-K., Song K-H & Kim B-I.(2020), "Maybe I'll be an asset too: the impact of rising housing prices on welfare awareness", social welfare policy practice, vol.6, no.2 pp.45~81, Korea.
4. Choi M-S., Byun S-I.(2016), "Comparison of Predictive Power of the Housing Price Estimation Model in Seoul", Seoul Metropolitan Research, 17.3: pp.75~89, Korea.
5. Conway, D., Li, C-Q., Wolch, J., Kahle C., Michael J.(2010), "A Spatial Autocorrelation Approach for Examining the Effects of Urban Greenspace on Residential Property Values", J Real Estate Finan Econ 41, pp.150~169.
6. Hou Y.(2017), "Traffic congestion, accessibility to employment, and housing prices: A study of single-family housing market in Los Angeles County", Urban Studies. vol.54(15), pp.3423~3445.
7. Hwang G-S., Lee Jae-kyung(2014), "Deeper Financialization and Inequality in Income and Housing", Trends and Prospects, No. 91, pp.199~231. Korea.
8. Ju, H-J.(2018), "A Study on the Influence of Road Traffic Noise on the House Price Based on Geographically Weighted Regression Model: The Case of Daelim-dong and Singil-dong Apartment in Yeongdeungpo-gu, Seoul", Thesis for Degree of Master, Graduate School of Sangmyung University, Seoul.

9. Kim H-H., Park S-W.(2013), "The Determinants of Housing Price in Seoul: Regression of the Fountain", Korea Housing Association, 21.2: pp.141~168, Korea.
10. Kim, B-H., Ryu, S-G., Hong, S-G.(2016), "The Effect of Medical Service Accessibility on the Housing Price - Focused on Apartment Complex in Gyeonggi - do, Korea -", The Journal of the Korea Real Estate Society, no.66, pp.188~201.
11. Kim, T-G. Park, H-S.(2008), "A Time Series Analysis of Spatial Factors Affecting Housing Prices", The Journal of the Korea Planning Association, vol. 43(3), pp. 145~166.
12. Ko, P-S., Park C-S.(2010), "A Study about the Real Estate' Policy Impact on house prices (Focusing on the time series analysis and regression)", The Journal of the Korea Institute of Electronic Communication Sciences, v.5, no.2, pp.205~213.
13. Lee J-Y., Choi M-H., Lee Sang-yeop(2007), "A Study on the Estimation of Apartment Prices through Data Mining", National Land Planning, 42.4 : pp.135~148, Korea.
14. Lee, J-H.(2010), "A study on the effect of Intra-City Accessibility Improvement on housing price", Thesis for Degree of Master, Graduate Engineering School of Hanyang University, Seoul.
15. Lee, S-H., Chun, K-K.(2012), "A Study of the Influences of Urban Infrastructure on Housing Prices : A Hierarchical Linear Model", The Journal of Korea Planning Association, vol. 47(4), pp.193~204.
16. Lee.(2016), "A Study on the Determinants of Housing Price Using the Hedonic Price Model", The Korea Urban Administration Society, Journal of Academic Presentation Conference. : pp.317~333, Korea.
17. MainLane, S.(2006), "A spatio-temporal Model of house prices in the US", US CESifo Working Paper, No. 1826, US.
18. Michael, F., Michael, P.(2013), "Housing prices and the business cycle: An empirical application to Hong Kong", Journal of Housing Economics, Vol.22, I.1, pp.62~76, Germany.
19. Mills, E-S.(1967) "An Aggregative Model of Resource Allocation in a Metropolitan Area", The American Economic Review, no.57, pp.97~210. US.
20. Muth, R-F.(1969), "Cities and Housing: The Spatial Pattern of Urban Residential Land Use", The University of Chicago Press, US.
21. Park, H-M.(2013), "An Analysis of the Factors Affecting on the Prices of the Urban-life Housing Type of the Structure of the Urban Spaces in Seoul - Focused on the System of Structured Spaces in the basic city planning of Seoul 2030", The Journal of the Korea Planning Association, vol. 48(3), pp.419~435.
22. Ping, p.(2018), "The Impact of Traffic Accessibility on Housing Prices", Thesis for Degree of Master, Graduate School of Konkuk University, Seoul.

23. Robert, T.(1997), "The lasso Method for Variable Selection in the Cox Model", *Statistics in Medicine*. 16(4), pp.385~395.
24. Shimizu, C., Nishimura, K., Watanabe, T.(2010), "Housing Prices in Tokyo: A Comparison of Hedonic and Repeat Sales Measures", *Journal of Economics and Statistics(Jahrbuecher fuer Nationaloekonomie und Statistik)*, Vol.10. no.230, pp.792~813, Japan.
25. Trevor, H., Robert, T.(1986), "Generalized Additive Models", *Statistical Science*, vol.1(3) pp.297~310, US.
26. Wang, W-C., Chang, Y-J, Wang, H-C(2019), "An application of the spatial autocorrelation method on the change of real estate prices in Taitung City", *ISPRS International Journal of Geo-Information*, vol. 8, no. 6, p. 249.
27. Wei, Y.(2017), "Forecasting house prices using dynamic Model averaging approach: Evidence from China", *Economic Modelling*, Vol.61, Pages 147~155, China.
28. Wen, H., Gui, Z., Yue Xiao, Y. Fang, L.(2018), "Subway Opening, Traffic Accessibility, and Housing Prices: A Quantile Hedonic Analysis in Hangzhou, China", *Sustainability, MDPI*, vol. 10(7), pp.1~23.
29. William, A.(1960), "A THEORY OF THE URBAN LAND MARKET", *Papers in Regional Science*, vol.6, no.1, pp.149~157. US.
30. Youn, H-M., Lee, S-H.(2013), "A Study on Interaction Effects of Regional and Individual Transportation Environment Factors on Price of Apartment Houses", *The Journal of The Korean Regional Development Association*, vol. 25(3), pp.129~144.
31. Z. Peng, Q. Huang and Y. Han(2019), "Model Research on Forecast of Second-Hand House Price in Chengdu Based on XGboost Algorithm", *2019 IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT)*, 2019, pp.168~172.

논문접수 : 2022.07.14.

1차 심사 : 2022.08.02.

게재확정 : 2022.12.22.

국문요약

Lasso 알고리즘 적용을 통한 교통환경의 주택가격 영향력 분석

국내 경제 동향을 살펴보면, 1인당 국민 총소득은 5년 동안 약 8% 증가했지만, 서울 아파트 실거래가격 지수는 약 88.6%에서 5년 만에 155%로 약 2배가량 증가했다. 이러한 급격한 상승세가 지속될 경우 소득계층 양극화 및 지역적 불균형, 경제성장의 둔화를 일으킬 수 있기 때문에 주택가격상승요인에 관한 연구와 정책적 방안이 필요한 상황이다. 본 연구는 주택가격에 영향을 미치는 요인들을 특성별로 분류하고 영향력을 수치화하여 교통환경 요소와 도시의 특성이 고려된 주택가격모형을 구축하였다. 공간적 범위는 경기도 31개 시·군 지역을 기준으로 하였으며, 교통환경과 인구특성을 수치화하기 위해 제1차 경기도민 삶의 질 설문조사 자료를 사용하였다. 수집한 데이터에 Lasso 알고리즘 적용을 통해 최적 주택가격모형을 선정하여 주택가격 결정요인을 선정하고, 변수들의 영향력을 비교하였다. 그 결과, 2기 신도시 지역과 그 외 지역 특성 변수가 주택가격에 영향력이 큰 변수로 나타났다. 다양한 교통환경 특성 중 교통안전도만 주택가격에 영향을 미치지 않았는데, 이는 도시지역 거주민들의 안전불감증을 간접적으로 반영한 결과이다. 도심으로 갈수록 교통혼잡도가 높아지기 때문에, 대중교통보다 이용이 불편한 자가용 이용만족도가 주택가격과 반비례하는 경향을 보였다. 편의시설로의 접근성은 좋을수록 주택가격이 상승하였다. 또한, 본 연구에서 구축된 Lasso 모형의 성능이 기존 널리 활용되던 Hedonic 모형보다 약 2배가량 우수함을 검증하였다.

주제어: 주택가격, R, Lasso 알고리즘, 교통환경, 만족도, 신도시