

주택금융연구

VOL.4

1. 아파트가격 지역 간 전이효과 분석 **방두완 · 권혁신**
2. 기계학습 알고리즘을 이용한 주택가격감정
시스템의 구축 및 평가: XGBoost, LightGBM,
CatBoost 알고리즘에 기반하여 **홍정의**
3. 자산심사제도 도입 전후 내집마련 디딤돌 대출의
특성 변화 분석 **김성진**
4. 주택특성 및 가구특성이 주택구입자금 부채비율에
미치는 영향 **이성원**
5. 인구구조 변화와 주택가격전망을 고려한
주택연금의 중장기 수요추정 **이종훈**

부록

주택금융연구

VOL.4

- | | |
|---|-----------|
| 1. 아파트가격 지역 간 전이효과 분석 | 방두완 · 권혁신 |
| 2. 기계학습 알고리즘을 이용한 주택가격감정
시스템의 구축 및 평가: XGBoost, LightGBM,
CatBoost 알고리즘에 기반하여 | 홍정의 |
| 3. 자산심사제도 도입 전후 내집마련 디딤돌 대출의
특성 변화 분석 | 김성진 |
| 4. 주택특성 및 가구특성이 주택구입자금 부채비율에
미치는 영향 | 이성원 |
| 5. 인구구조 변화와 주택가격전망을 고려한
주택연금의 중장기 수요추정 | 이종훈 |

부록

- 주택금융연구의 내용을 인용할 때에는 반드시 “한국주택금융공사 주택금융연구 통권 제〇호 〇〇페이지에서 인용”이라고 표시하여 주십시오.
- 주택금융연구에 대한 질의 또는 제안은 주택금융연구원(☎ 051-663-8155)으로 연락하여 주시기 바랍니다.

목 차

아파트가격 지역 간 전이효과 분석 A Study on the Spillover Effect of the Regional Housing Price 방두완 · 권혁신	5
기계학습 알고리즘을 이용한 주택가격감정 시스템의 구축 및 평가: XGBoost, LightGBM, CatBoost 알고리즘에 기반하여 An Application of XGBoost, LightGBM, CatBoost Algorithms on House Price Appraisal System 홍정의	33
자산심사제도 도입 전후 내집마련 디딤돌 대출의 특성 변화 분석 An Analysis of the Characteristics of Didimdol Loans with the Introduction of the Asset Review System 김성진	65
주택특성 및 가구특성이 주택구입자금 부채비율에 미치는 영향 The Effect of Housing and Household Characteristics on the Purchase Loan to Value 이성원	85
인구구조 변화와 주택가격전망을 고려한 주택연금의 중장기 수요추정 Long-Term Demand Estimation for Reverse Mortgage Considering Changes in Population Structure and House Price Prospect 이종훈	111
부록	141

아파트가격 지역 간 전이효과 분석

방두완* · 권혁신**

요 약

본 연구는 Diebold와 Yilmaz (2009)가 제안한 방법론을 활용하여 지역별 주택가격의 전이효과를 분석하였다. 7개 지역에 대해 1986년에서 2020년까지, 16개 지역에 대해서는 2003년 7월부터 2020년 7월까지 주택가격 전이효과를 분석하였다. 7개 지역 지역별 아파트 가격 수익률 분석결과, 지역별로 상호 영향을 미치는 전이효과를 확인하였으며, 서울 아파트 가격 수익률은 자기자신의 영향을 가장 많이 받으며, 인천 아파트 가격 수익률에 큰 영향을 미치는 것으로 분석되었다. 경제위기를 고려한 하위기간 지역별 아파트 가격 전이효과 실증분석 결과, 서울 아파트 가격이 다른 지역에 미치는 영향이 크지는 않지만, 외환위기와 금융위기 기간 이후 점차 감소하는 것으로 나타난다. 마지막으로 16개 지역의 아파트 가격 수익률과 변동성의 50% 이상이 전이효과로 설명되는 것으로 분석되었고, 따라서 한국 지역별 아파트 가격은 서로 영향을 많이 주고받는 것으로 결론지을 수 있다.

핵심어 : 지역별 아파트 가격 전이효과, 일반화 예측오차 분산분해, 전이효과 지수, 순전이효과

1. 서론

일반적으로 부동산가격의 변동은 임의보행과정(random walk process)을 따르지 않고 일정한 변동주기를 가지며 상승과 하락을 반복하는 순환변동과정으로 알려져 있다.

따라서 부동산가격은 호황기와 불황기가 존재하며, 특정 지역의 부동산가격이 다른 지역의 부동산가격에 영향을 미친다는 선행연구들이 존재한다(이항용, 이진, 2014; 전형철, 형남원, 2018; 방두완, 권혁신, 김명현, 2019).

이러한 이유로 정부와 부동산시장의 시장참여자들은 현재의 부동산시장이 호황인지 불황인지,

* 방두완, 주저자, 주택도시보증공사 주택도시금융연구원 선임연구위원, doowoan@khug.or.kr

**권혁신, 교신저자, 주택도시보증공사 주택도시금융연구원 차장(행정학박사), hskwon@khug.or.kr

© Copyright 2020 Housing Finance Research Institute. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

그리고 한국의 경우 서울이나 강남지역 등 특정 지역의 주택가격이 다른 지역의 주택가격에 영향을 미치는지에 대해 민감하게 반응하는 경향이 있다.

구체적으로 정부는 수립된 정책 목적 달성을 위하여 부동산시장에 개입할 필요성이 있으며, 가계나 기업도 부동산시장에 참가하여, 합리적인 의사결정을 내리기 위해서 부동산시장 및 지역 간 전이효과(spillover effect)에 관한 정보가 필요하다고 할 수 있다.

그러나 현실적으로 부동산 가격지수(주거용, 상업용 부동산지수)는 발표되고 있으나, 현재 부동산시장이 호황인지 불황인지를 부동산 참여자들이 정확히 파악하기는 쉽지 않으며, 지역 부동산 시장간의 전이효과 연구도 도 단위의 지역의 경우, 연구결과가 보고되고 있지 않다.

그리고 기존 선행연구들이 제시한 지역 전이효과 또한 대부분의 연구들이 주택가격의 변동성을 중심으로 지역 간 전이효과를 분석하고 있다. 따라서 주택가격 자체의 수익률을 중심으로 지역 간 전이효과를 분석할 필요성이 있다.

따라서 본 연구는 Diebold and Yilmaz (2009)의 연구 방법론을 준용하여, 지역별 주택가격 전이효과를 분석하기로 한다. 한걸음 더 나아가 본 연구는 서울, 광역시, 그리고 도 단위 지역을 포함한 16개 지역을 대상으로 지역 간 전이효과를 추가로 분석하기로 한다.

이를 위해 본 연구는 분석 기간을 전체기간 및 3개의 하위기간으로 구분하여 분석하기로 한다. 구체적으로 전체기간(1986년 1월~2020년 7월)의 지역 간 주택가격 전이효과를 분석한 후, 분석기간 동안 발생한 외환위기와 금융위기 기간을 반영하여, 분석기간을 3개 즉, 외환위기 발생이전(1986년 1월~1997년 12월), 외환위기 발생이후 금융위기 발생 이전(1998년 1월~2008년 8월), 금융위기 발생 이후(2008년 9월~2020년 7월)로 구분하여 분석하기로 한다.

분석기간을 세부적으로 구분하는 이유는 경제적 충격(economy shock) 이전과 이후의 지역 간 주택가격 전이효과에 차이가 있을 것으로 기대되기 때문이다.

이후의 본 연구 구성은 다음과 같다. 제 I 장 서론에 이어 제 II 장에서는 선행연구를 살펴본다. 그리고 제 III 장에서는 분석모형 및 데이터를 고찰하고, 제 IV 장에서는 지역 간 주택가격 전이효과를 실증분석한다. 마지막으로 제 V 장에서는 본 연구의 결론 및 정책 시사점에 대해 논의한다.

II. 선행연구

주택가격이 지역 간에 상호 영향을 미칠 수 있다는 지역 전이효과에 관한 연구는 부동산시장과 정책을 담당하는 실무자 입장에서는 중요한 부분이라고 할 수 있으며, 최근 지속적인 연구결과들이 보고되고 있다.

김의준, 김양수, 신명수(2000)는 수도권 아파트 매매가격의 지역 연관성을 연구하였는데, 수도권 아파트 가격변동을 일으키는 중심 지역은 강남구, 서초구 등이 속해 있는 동남지역이며, 동남지역의 가격은 양천구, 동작구 등의 서남 지역과 고양시 등의 가격변동에 영향을 미치는 것으로 나타났다.

이상경(2003)은 GARCH(1,1), EGARCH(1,1) 모형을 사용하여 강남지역으로부터 수원시와 부산시로의 가격 및 변동성 확산효과를 분석한 결과, 서울 강남지역으로부터 수원·부산으로 가격 및 변동성의 전이효과가 존재함을 발견하였다.

서승환(2007)은 강남구를 출발점으로 하는 물결효과가 존재하는지, 그리고 축차적 수요소진 가설이 물결효과와 관련이 될 수 있는지에 대해, 그랜저-심즈 인과관계 검정(Granger-Sims Causality Test) 및 VECM(벡터오차수정모형)의 추정결과를 이용하여 실증분석하였다. 분석결과, 강남구 아파트 매매가격은 인접 지역을 중심으로 아파트 매매가격에 영향을 주고 있다는 분석결과를 보고하였다.

박헌수, 안지아(2009)는 개별 지역에서의 주택가격변화에 상호 영향을 주는 요인들을 거시경제 요인, 주변 지역과 지역 내 매매시장 및 전세시장 요인들로 세분화하였다. 최종적으로 박헌수, 안지아는 이러한 요인들이 미치는 영향들을 분석하였다. 분석 결과, 서울 아파트 매매가격은 단기적으로 주변 지역에서 영향을 미치는 매매요인에 의해 크게 영향을 받고 있다는 결과를 제시하고, 지역별로 주택시장의 움직임을 파악하여 주택가격 변동요인에 따른 주택시장의 움직임을 파악하였다. 그 결과, 지역 여건에 따른 차이를 감안하여 정책을 수립해야 한다고 보고하고 있다.

최경욱, 형남원(2010)은 물가, 금리 혹은 환율과 같은 가격변수의 변동성이 증가하였을 때, 국내 경기에 어떤 영향을 상호 미치는지에 대한 상호효과를 분석했다. 실증분석의 결과, 최경욱, 형남원은 경제성장률의 불확실성이 증가하면, 물가상승률이 높아지는 현상을 보고하고 있다. 그리고 고빈도 자료(high frequency data) 간의 확산효과(spillover)를 분석한 Diebold and Yilmaz(2009) 모형을 저빈도 자료(low frequency data)인 거시경제변수를 대상으로 분석이 가능하도록 보완한 모형으로 확산효과를 분석하였다. 최경욱, 형남원은 거시경제변수의 변동성을 살펴본 결과, 2001년 이후 동 변수들의 변동성 확산 효과가 지속적으로 증가하고 있는 것으로 보고하고 있다. 그리고 변동성은 2008년 하반기 국제금융위기 시에 정점에 이르고, 이후 빠른 속도로 감소하는 것으로 보고하고 있다.

문규현, 이동희(2011)는 1986년부터 2009년까지의 강남아파트 매매가격지수, 강남아파트 전세가격지수, 전국 아파트매매가격지수, 전국 아파트전세가격지수의 월별자료를 이용하여 강남아파트시장과 전국아파트 시장 사이의 관계를 분석하였으며, 결과적으로 이러한 변수들 간의 선도(lead)와 지연(lag)관계를 분석하고자 하였다. 이들은 분석모델로 그랜즈 인과관계분석,

충격반응함수분석 및 분산분해분석과 같은 동태적 VAR분석기법을 사용하였는데, 분석결과, 전체표본 분석에서 강남아파트 매매시장·전세시장은 전국아파트매매시장·전세시장에 선행하는 것으로 보고하고 있다.

전해정(2013)도 GARCH, EGARCH모형을 이용해 2003년부터 2013년까지 서울, 강남, 강북 아파트 매매가격·전세가격 간의 가격변동성 및 전이효과를 실증분석하였다. 분석결과, 강남지역에서 강북·서울지역으로 매매가격과 전세가격의 변동성 이전 효과가 어느 정도 존재함을 확인하였다.

박해선, 김승년(2014)은 DAG 방법을 적용하여 주택가격의 지역 간 상호의존성을 밝히고자 하였으며, 부동산시장과 정책당국의 관심이 집중된 강남지역의 아파트매매가격이 인접 지역의 아파트 매매가격에 미치는 영향을 분석하였는데, 서울 강남구 주택가격이 전국 주택가격과 인근 지역 주택가격을 선도한다는 결과를 보고하였다.

한편, 강임호, 조영진(2014)은 전국 8대 지역의 전세지수·매매지수·전세매매비율 사이의 공적분, 그리고 전세지수·매매지수와의 공적분 관계를 분석하였고, 강남, 강북, 인천이 포함된 수도권 지역의 전세지수 간, 매매지수 간, 전세매매비율 간에는 공적분이 있음을 보고하고 있다.

이항용, 이진(2014)은 서울 지역과 전국 6대 광역시의 아파트 매매가격을 이용하여 지역 사이에 존재할 수 있는 전이효과를 분석하였다. 실증분석결과, 서울과 6대 광역시 사이의 상호 전이효과는 상당히 높은 수준으로 나타났음을 보고하였고, 전체 지역의 가격변동의 50% 이상이 전이효과로 설명되는 것으로 분석하였다.

박영준, 김기호(2017)는 Diebold and Yilmaz (2012) 방법을 이용하여 하위지역별 관점에서, 수도권을 8개 지역으로 분류하여 지역별 주택가격 변동성의 동조화 및 변동성 전이에 대해 분석하였다. 잠재 요인별 분산분해 결과, 수도권 대부분 지역의 주택시장 변동은 동태적 잠재요인 모형으로 추정된 수도권 공통요인의 설명력이 매우 높은 것으로 분석되었다. 따라서 이들은 수도권 주택시장에 동조화 현상이 존재함을 보고하고 있다.

전형철, 형남원(2018)은 주택시장에 대한 서울시 강남지역의 영향력을 분석하였다. 주택시장을 서울시 강남지역, 강북지역, 6대 광역시의 3개 권역으로 나누어 분석하였다. 전형철, 형남원은 분석대상 지역의 매매와 전세변수를 이용하여 Diebold and Yilmaz (2009)의 방법론에 따라 분석하였다. 실증분석 결과, 전체 주택시장의 확산 효과 지수는 67%로 상당히 높은 것으로 보고하고 있다. 특히 서울 강남의 영향력이 큰 것으로 보고하고 있는데, 1997년~2008년 동안 서울 강남의 다른 지역에 대한 영향력이 매우 높은 것으로 보고하고 있다. 이러한 분석결과를 근거로 전형철, 형남원은 서울 강남의 주택가격의 급등이 다른 지역의 매매시장과 전세시장에 큰 파급을 주어 주택시장에서 강남효과가 실재한다고 결론짓고 있다. 추가로 확산 효과의 추이 분석에서, 1997년

이후 강남발 충격의 영향력이 급등하였으나, 2008년 글로벌 금융위기 이후 지속해서 하락하고, 2011년 이후에는 안정적 추세로 바뀌어, 강남발 충격에 의한 주택시장의 교란 요인이 점차 약화하는 것으로 보고하고 있다.

방두완, 권혁신, 김명현(2019)는 FAVAR 모형을 이용하여 지역별 주택경기지수를 작성하고, 작성된 주택경기지수를 이용하여 Diebold and Yilmaz (2012) 방법론으로 지역별 주택가격의 전이효과를 분석하였다. 실증분석 결과, FAVAR 모형을 이용하여 작성한 전국아파트 경기지수와 통계청에서 발표하는 경기동행지수 순환변동과 유사한 시계열 추이를 보이는 것으로 보고하고 있다. 특히 2008년 금융위기와 그 이후의 경제 충격을 전국 아파트 경기지수가 잘 반영하며, 전국 아파트 경기지수는 경기동행 순환변동을 약 4~5개월 정도 선행한다고 분석하고 있다. 지역별 아파트 경기지수 분석결과, 2008년 세계금융위기의 영향을 공통으로 받은 것을 확인되었지만, 금융위기 이후에는 지역별 아파트 경기지수가 차이를 보이는 것으로 분석되었다. 최종적으로 아파트 경기지수는 서울과 수도권이 서로 영향을 주고받고 있으며, 부산에도 상당한 영향을 미치는 것으로 보고하고 있다.

지금까지 살펴본 지역별 전이효과 선행연구와 본 연구의 차이점은 선행연구들이 대부분은 지역별 전이효과를 분석하기 위해 주택가격의 월별 변동성을 산정하여 분석했지만, 본 연구는 Diebold and Yilmaz (2009)의 연구에서 제시한 방법론을 적용하여 지역별 주택가격 수익률을 적용하여 지역별 주택가격 전이효과를 분석한 점에서 선행연구와 구별되며, 분석지역을 도 단위로 확장한 점에서도 차별성이 있다고 하겠다.

III. 분석모형 및 데이터

1. 분석모형

전이효과를 분석한 국내외 대부분의 연구들은 대부분 주식시장이나 금융시장을 대상으로 분석하였다. 이후 일부의 연구들이 주택가격의 지역 간 전이효과를 연구하기 시작하였다.(최경욱, 형남원, 2010; 이항용, 이진, 2014; 전형철, 형남원, 2018; 방두완, 권혁신, 김명현, 2019).

그리고 전이효과를 연구한 다수의 선행연구들은 1단계로 VAR(Vector Auto Regression) 모델을 사용하고, 2단계로 추정결과를 바탕으로 그랜저 인과관계 검정(Granger Causality Test), 충격반응분석(Impulse Response Analysis), 예측오차 분산분해(Forecast Error Variance

Decomposition) 등을 사용하여 전이효과를 분석하였다.

본 연구에서는 Diebold and Yilmaz (2009)의 방법을 준용하여 지역 간 아파트 가격 전이효과를 분석하였는데, 구체적으로 자산가격의 수익률이나 변동성을 중심으로 전이효과를 추정하는 방법을 제안하였다. 이변량 VAR₍₁₎ 모형을 이용하여 전이효과를 추정하는 방법은 다음과 같다.

$$X_t = \Phi X_{t-1} + \epsilon_t \quad \langle \text{식 1} \rangle$$

여기서 X_t 는 2×1 매트릭스, Φ 는 2×2 매트릭스이다. 이를 이동평균(moving average) 방식으로 표현하면,

$$X_t = \theta(L)\epsilon_t \quad \langle \text{식 2} \rangle$$

이때, $\theta(L) = (I - \Phi)^{-1}$ 이다. ϵ_t 의 분산·공분산 행렬 콜레스키 요인(Cholesky factor)인 Q^{-1} 을 이용하여 다시 기술하면 다음과 같이 표현할 수 있다.

$$X_t = A(L)\mu_t \quad \langle \text{식 3} \rangle$$

여기서 $A(L) = \theta(L)Q^{-1}$, $\mu_t = Q_t\epsilon_t$, $E(\mu_t\mu_t') = I$ 이다. 따라서 1-기간 예측오차(1-step ahead forecast error)는 다음과 같이 표현할 수 있다.

$$\epsilon_{t+1,t} = X_{t+1} - X_{t+1,t} = A_0\mu_{t+1} = \begin{bmatrix} a_{0,11} & a_{0,12} \\ a_{0,21} & a_{0,22} \end{bmatrix} \begin{bmatrix} \mu_{1,t+1} \\ \mu_{2,t+1} \end{bmatrix} \quad \langle \text{식 4} \rangle$$

다음으로 예측오차의 공분산 행렬(covariance matrix)은 다음과 같다.

$$E(\epsilon_{t+1,t}\epsilon_{t+1,t}') = A_0A_0' \quad \langle \text{식 5} \rangle$$

여기에서 $X_{1,t}$ 의 1-기간 예측오차의 분산은 $a_{0,11}^2 + a_{0,12}^2$, $X_{2,t}$ 의 1-기간 예측오차의 분산은 $a_{0,21}^2 + a_{0,22}^2$ 이다. VAR 분석에서 많이 사용하는 분산분해(variance decomposition)를 이용하면, 자신의 충격과 다른 변수로부터 오는 충격으로 각각 분해할 수 있다.

따라서 전이효과 지수(spillover index)는 다른 변수로부터 받는 영향인 $a_{0,12}^2 + a_{0,21}^2$ 과 전체 예측오차의 변동성인 $a_{0,11}^2 + a_{0,12}^2 + a_{0,21}^2 + a_{0,22}^2 = \text{trace}(A_0 A_0')$ 의 비중으로 계산할 수 있다.

$$S = \frac{a_{0,12}^2 + a_{0,21}^2}{\text{trace}(A_h A_h')} \times 100 \quad \langle \text{식 6} \rangle$$

그리고 N개의 변수를 가진 VAR(p) 모형의 전이효과 지수는 다음과 같이 일반화할 수 있으며, 식(7)을 이용하여 지역별 아파트 가격 전이효과 지수를 추정할 수 있다.

$$S = \frac{\sum_{h=0}^{H-1} \sum_{i,j=1, i \neq j}^N a_{h,ij}^2}{\sum_{h=0}^{H-1} \text{trace}(A_h A_h')} \times 100 \quad \langle \text{식 7} \rangle$$

2. 데이터 및 분석변수

본 연구에서 지역별 아파트 가격과 변동률 분석에 사용한 변수는 국민은행에서 발표하는 전국 및 지역별 아파트 월별 가격지수를 사용하였다. 본 연구의 전체 분석기간은 1986년 1월부터 2020년 7월까지이며, 전체 분석기간에 적용한 분석지역은 시계열 자료가 전 기간 존재하는 서울, 부산, 대구, 인천, 광주, 대전, 울산 지역을 대상으로 하였다.

본 연구는 분석대상 기간에 포함된 외환위기와 금융위기를 전후하여 지역별 아파트 가격지수의 전이효과에 변화가 있는지를 분석하였다. 구체적인 하위분석 기간의 설정은 전형철, 형남원(2018)의 연구와 같이 외환위기 이전 기간(1986년 1월~1997년 12월), 외환위기 이후부터 금융위기 이전기간(1998년 1월~2008년 8월), 금융위기 이후 기간(2008년 9월~2020년 7월)으로 구분하여 분석하였다.

마지막으로 본 연구는 서울, 광역시, 도를 모두 포함하는 16개 지역(서울, 부산, 대구, 인천, 광주, 대전, 울산, 경기도, 강원도, 충청남도, 충청북도, 전라남도, 전라북도, 경상남도, 경상북도, 제주도)의 주택가격 전이효과를 분석하였다. 이를 위해 <표 1>에서 제시한 것처럼 16개 지역의 주택가격 시계열을 모두 확인할 수 있는 2003년 7월부터 2020년 7월까지를 분석 기간으로 분석하였다.

〈표 1〉 분석변수

변수	세부내용	출처
아파트 가격 수익률	국민은행 지역별 아파트 가격지수 월별 수익률	국민은행
아파트 가격 변동성	국민은행 지역별 아파트 가격지수 월별 변동성	국민은행

3. 기초통계량 분석

본 연구에서 선정한 전체기간 분석지역은 7개 지역이며, 구체적으로 서울, 부산, 대구, 인천, 광주, 대전, 울산이다. 선행연구에서 주장하는 강남지역의 파급효과 즉, 강남효과가 있는 지를 분석하기 위해 서울을 대용할 변수로 강남지역의 아파트 가격을 추가하여 비교 분석하였다.

지역별 아파트 가격변수들의 기초통계량은 〈표 2〉에 자세히 나타나 있다. 먼저 패널 A는 지역별 아파트 가격 수익률의 기초 통계량이 제시되어 있다. 전체 관측월수는 414개월이며, 월별 평균 변동률은 0.003%에서 0.004% 사이로 분석되었고, 월별 최고 하락률은 광주 아파트 가격 -0.075%로 분석되었다.

패널 B는 지역별 아파트 가격 변동성의 기초 통계량이 제시되어 있다. 서론에서 언급한 것처럼 대부분의 선행연구에서 사용한 전이효과 분석은 대부분 수익률이 아닌 변동성을 계산하여 분석하였다. 본 연구에서 사용한 변동성은 Black-Scholes model에서 제시된, 주식가격의 변동성은 수익률의 표준편차로 계산된다는 정의에 따라 계산하였다(Hull, 2005).

구체적인 작성방법을 살펴보면, 먼저 분석기간 전체의 월별 로그차분 수익률을 계산한 후, 전체기간 로그차분 수익률 평균-월별 로그차분 수익률의 절댓값으로 계산하였다.¹⁾

패널 B는 분석결과를 살펴보면, 월별 평균 변동성은 0.005%에서 0.009% 사이로 분석되어, 서울 아파트 가격의 변동성이 큰 것으로 나타났다.

〈그림 1〉은 지역별 아파트 가격 수익률 및 변동성 추이를 보여주고 있다. 먼저 패널 A의 수익률 추이를 보면 1990년대에 상대적으로 수익률의 변화가 크고, 1997년 외환위기 시 최대폭의 수익률 하락을 보여주고 있으며, 2008년 금융위기의 수익률 하락은 상대적으로 낮은 것으로 나타난다. 패널 B의 변동성 추이도 수익률과 큰 차이를 보이지 않는 것으로 보여진다.

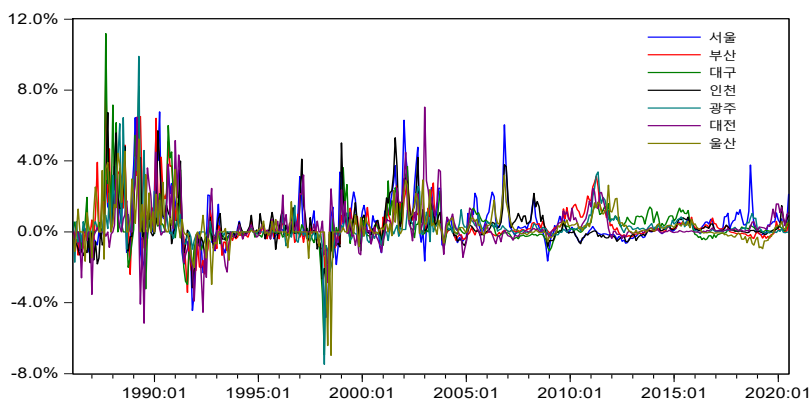
1) 지역별 주택가격 변동성은 ABS(전체기간 로그차분 수익률 평균-월별 로그차분 수익률)로 계산하였다. 보다 자세한 계산 방법은 Hull(2005)를 참고하기 바란다.

〈표 2〉 기초통계량

패널 A. 지역별 아파트 가격 수익률 (1986년 2월~2020년 7월)					
변수명	관측치	평균	최대값	최소값	SD
서울	414	0.004	0.068	-0.048	0.014
부산	414	0.004	0.090	-0.034	0.013
대구	414	0.003	0.112	-0.069	0.013
인천	414	0.003	0.067	-0.032	0.012
광주	414	0.003	0.099	-0.075	0.010
대전	414	0.003	0.070	-0.052	0.012
울산	414	0.003	0.052	-0.069	0.011

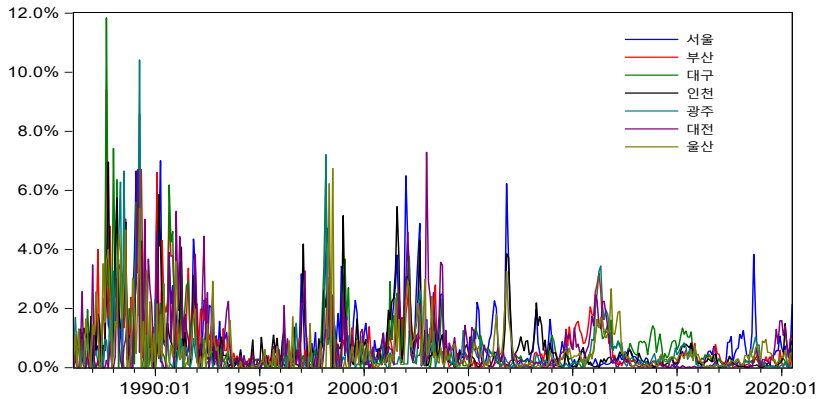
패널 B. 지역별 아파트 가격 변동성 (1986년 2월~2020년 7월)					
변수명	관측치	평균	최대값	최소값	SD
서울	414	0.009	0.070	0.000	0.012
부산	414	0.008	0.094	0.000	0.011
대구	414	0.007	0.118	0.000	0.012
인천	414	0.007	0.070	0.000	0.010
광주	414	0.005	0.104	0.000	0.010
대전	414	0.007	0.073	0.000	0.010
울산	414	0.006	0.067	0.000	0.009

패널 A. 지역별 아파트 가격 수익률 추이 (1986년 2월~2020년 7월)



〈그림 1〉 지역별 아파트 가격 수익률 및 변동성 추이

패널 B. 지역별 아파트 가격 변동성 추이 (1986년 2월~2020년 7월)



〈그림 1〉 계속

IV. 지역 간 주택가격 전이효과 실증분석

본 연구는 지역 간 주택가격 전이효과를 분석하기 위해 1986년 1월부터 2020년 7월까지 전체기간 분석이 가능한 7개 지역의 지역별 아파트 가격 전이효과를 먼저 분석한 후, 외환위기와 금융위기 기간을 구분하여 경기충격을 전후로 지역 간 주택가격 전이효과에 차이가 있는지를 주택가격 수익률과 변동성 모형으로 구분하여 비교 분석하였다. 그리고 나아가 본 연구는 전국 16개 지역의 주택가격 전이효과 비교분석이 가능한 2003년 7월부터 2020년 7월까지 이들 지역의 지역 간 주택가격 전이효과를 비교분석하고 결과를 제시하였다. 분석에 사용한 방법론은 Diebold and Yilmaz (2009)가 제시한 방법론을 사용하였으며, 수익률 모형과 변동성 모형을 비교 분석하여 결과를 제시하였다.

분석을 위해 본 연구는 지역별 아파트 가격지수의 시계열 안정성을 검증하였다. 일반적으로 시계열 자료에 단위근(unit root)을 가지는 불안정 시계열인 경우가 많다. 단위근이 존재하는 시계열을 최소자승회귀분석(ordinary least square, OLS)으로 분석하면 가성회귀(spurious regression) 문제가 발생하기 때문에 통상 단위근 검정을 먼저 수행하는 것이 필요하며, 본 연구에서 사용한 단위근 검정법은 ADF(augmented Dickey-Fuller) 검정법이다.

1. 단위근 검정

본 연구에서 분석에 사용한 국민은행의 지역별 아파트 가격지수의 단위근 분석결과는 〈표 3〉에

〈표 3〉 단위근 검정결과

패널 A. 아파트 가격 수익률 단위근 검정결과 (1986년 2월~2020년 7월)			
변수명	수준변수	5% 유의수준	적분 차수
서울	-5.648	-3.421	I(0)
부산	-6.379	-3.421	I(0)
대구	-6.817	-3.421	I(0)
인천	-5.124	-3.421	I(0)
광주	-5.102	-3.421	I(0)
대전	-11.443	-3.421	I(0)
울산	-5.309	-3.421	I(0)
패널 B. 아파트 가격 변동성 단위근 검정결과 (1986년 2월~2020년 7월)			
변수명	수준변수	5% 유의수준	적분 차수
서울	-6.074	-3.421	I(0)
부산	-7.728	-3.421	I(0)
대구	-6.765	-3.421	I(0)
인천	-5.877	-3.421	I(0)
광주	-6.084	-3.421	I(0)
대전	-5.489	-3.421	I(0)
울산	-9.364	-3.421	I(0)

제시되어 있다. 먼저 패널 A 지역별 아파트 가격수익률 단위근 검정결과, 서울, 부산, 대구, 인천, 광주, 대전, 울산 지역 아파트 가격지수 수익률 변수는 예상한 것처럼, 모두 I(0) 시계열 즉, 안정 시계열로 나타난다.

먼저 패널 B 지역별 아파트 가격 변동성 단위근 검정결과, 서울, 부산, 대구, 인천, 광주, 대전, 울산 지역 아파트 가격지수 변동성 변수는 수익률 변수와 같이 모두 I(0) 시계열 즉, 안정 시계열로 나타난다. 따라서 본 연구에서는 지역별 아파트 가격지수 수익률과 변동성 변수를 사용하여 분석하기로 한다.

2. 지역별 주택가격 그랜저 인과관계(Granger causality) 분석

〈표 4〉는 지역별 주택가격 수익률 변수의 그랜저 인과관계 분석결과를 제시하고 있다. 그랜저

〈표 4〉 그랜저 인과관계 분석결과(수익률)

귀무가설	관측치	F 통계량	유의확률
부산 ⇨ 서울	409	5.301	0.001
서울 ⇨ 부산		16.544	0.001
대구 ⇨ 서울	409	1.346	0.244
서울 ⇨ 대구		7.369	0.001
인천 ⇨ 서울	409	1.795	0.113
서울 ⇨ 인천		44.521	0.000
광주 ⇨ 서울	409	4.375	0.001
서울 ⇨ 광주		3.591	0.003
대전 ⇨ 서울	409	2.489	0.031
서울 ⇨ 대전		20.871	0.001
울산 ⇨ 서울	409	1.431	0.212
서울 ⇨ 울산		6.589	0.001

주: 1. 적정시차 분석결과 적정 시차는 5로 분석되어 그랜저 인과관계의 시차도 5로 가정하였다.
2. 지역별 전체 그랜저 인과관계 분석결과는 〈부록 표 1〉에 제시하였으며, 분석변수는 아파트 수익률 변수이다.

인과관계 분석은 시차에 따라 분석결과에 차이를 보이기 때문에 본 연구에서는 VAR(p) 모형 시차분석 결과에 따라 적정 시차로 나타난 시차 5를 사용하기로 한다.

인과관계 분석 결과, 서울과 부산, 서울과 광주, 서울과 대전은 서로 그랜저 인과관계가 존재하는 것으로 분석되었다. 이외에도 서울의 아파트 가격 수익률은 대구, 인천 그리고 울산에 통계적으로 유의적인 그랜저 인과관계가 영향을 미치는 것으로 나타나 서울은 모든 지역의 아파트 가격 수익률에 그랜저 인과관계가 있는 것을 확인할 수 있다.

그러나 그랜저 인과관계는 두 변수간의 상호영향에 중점을 두기 때문에, 전체적인 시장 관점에서 전이효과를 분석하기에는 일정한 한계가 있다. 이런 관점에서 본 연구는 Diebold and Yilmaz (2009) 방법론을 이용하여 지역 간 전이효과의 영향과 크기를 보다 세부적으로 분석하기로 한다.

3. 7개 지역의 지역 간 아파트 가격 전이효과

본 연구는 7개 지역(서울, 부산, 대구, 인천, 광주, 대전, 울산)의 지역별 아파트 가격지수 수익률과

변동성을 이용하여 시차가 5인 VAR₍₅₎ 모형을 추정하였다.²⁾ 분석기간은 7개 지역 아파트 가격을 확인할 수 있는 1986년 1월부터 2020년 7월까지이다. 전이효과 분석모형을 이용한 지역별 아파트 가격 전이효과 분석결과는 <표 5>에 제시되어 있다.

<표 5>의 패널 A를 자세히 설명하면, 행렬의 i 번째 행과 j 번째 열에 해당하는 셀을 $\theta_{i,j}$ 라고 한다면, $\theta_{i,j}$ 의 의미는 j 지역 아파트 가격이 i 지역 아파트 가격에 미치는 영향이다. 그리고 대각행렬인(diagonal) $\theta_{i,i}$ ($i=1, \dots, 7$)는 각 지역의 아파트 가격이 자기 자신에게 미치는 영향을 의미한다. 그리고 비대각(off-diagonal) 요인인 $\theta_{i,j}$ 는 j 지역 아파트 가격이 i 지역 아파트 가격에 미치는 전이효과를 나타낸다.

<표 5>의 패널 A의 9번째 열은 개별 행에서 자신의 값을 제외한 비대각 셀들의 값을 모두 더한 것으로 각 지역이 다른 지역으로부터 받게 되는 전이효과의 합을 나타낸다. 즉, 서울 지역의 경우, 9번째 열에 표시된 값은 서울의 값을 제외한 다른 열들의 합 즉, 다른 지역들이 서울에 미치는 합이다.

이와 마찬가지로 9번째 행은 각각의 열에서 자기 자신의 지역을 제외한 비대각 셀들의 합이며, 각 지역이 다른 지역에 미치는 전이효과의 크기를 의미한다. 그리고 서울의 9행의 값은 171.5이며, 이는 서울지역 아파트 가격 수익률이 다른 지역에 미치는 전이효과의 크기를 의미한다. 그리고 괄호 안의 값 64.9%는 서울지역 아파트 가격 수익률이 다른 지역 아파트 가격에 미치는 전이효과 전체 비중을 나타낸다.

<표 5>의 패널 A의 각 행렬을 살펴보면, 지역별로 상호 영향을 미치는 전이효과를 확인할 수 있다. 서울의 경우, 자기 자신의 영향을 가장 많이 받으며, 그 다음으로 서울의 아파트 가격 수익률은 인천의 아파트 가격 수익률에 큰 영향을 미치는 것으로 분석되었다.

이외에도 그랜저 인과관계 분석결과 동일하게, 서울의 아파트 가격 수익률은 부산, 대구, 대전, 울산, 광주 순으로 모든 지역 아파트 가격 수익률에 영향을 미치는 것으로 나타난다.

반면, 서울의 아파트 가격 수익률에 다른 지역이 미치는 영향은 7.1로 나타나 매우 낮은 것으로 분석이 되었으며, 부산의 아파트 가격이 다른 지역에 비해 상대적으로 가장 큰 영향을 미치는 것으로 나타난다.

부산의 아파트 가격 수익률도 자신의 영향이 가장 큰 것으로 나타났으며, 다음으로 대구 아파트 가격 수익률에 영향을 미치는 것으로 분석되었다. 대구 아파트 가격 수익률이 가장 큰 영향을 미치는 지역은 광주이며, 다른 지역 인천, 광주, 대전, 울산의 아파트 가격 수익률은 다른 지역에 큰 영향을

2) 전체기간의 지역 간 아파트 가격 전이효과 분석을 위한 VAR(p) 모형 시차분석 결과, SC (Schwarz information criterion), HQ(Hannan-Quinn information criterion) 통계량의 적정 시차가 5로 분석되어, VAR₍₅₎로 전이효과를 분석하였다. 추정된 SC와 HQ의 기준 시차 5의 통계량은 각각 -6.269, -6.305로 분석되었으며, 일반화 예측오차 분산분해를 위한 예측기간은 12개월로 설정하였다.

〈표 5〉 지역 간 아파트가격 전이효과 분석(서울)

패널 A. 주택가격 수익률 모형 (1986년 2월~2020년 7월)								
지역	서울	부산	대구	인천	광주	대전	울산	전이효과: from others
서울	92.9	2.9	0.7	1.0	0.9	1.2	0.4	7.1
부산	30.1	53.3	6.4	0.9	2.0	6.3	0.9	46.6
대구	23.2	13.4	54.0	1.4	1.3	5	1.6	45.9
인천	59.7	4.2	2.1	32.9	0.2	0.5	0.5	67.2
광주	17.2	6.6	14.8	0.6	54.3	5.7	0.9	45.8
대전	20.8	2.5	0.8	1.8	1.9	71.9	0.4	28.2
울산	20.5	6.0	8.5	1.8	3.0	3.1	57.1	42.9
전이효과: to others	171.5 (64.9%)	35.6 (40.4%)	33.3 (38.1%)	7.5 (18.6%)	9.3 (14.6%)	21.8 (23.3%)	4.7 (7.6%)	283.7 (40.5%)

주: VAR(p) 모형의 시차는 5임.

패널 B. 주택가격 변동성 모형 (1986년 2월~2020년 7월)								
지역	서울	부산	대구	인천	광주	대전	울산	전이효과: from others
서울	91.7	3.7	0.5	3.3	0.1	0.5	0.2	8.3
부산	21.4	64.6	5.7	0.5	5.7	1.8	0.2	35.3
대구	16.0	17.5	62.2	0.3	1.2	2.1	0.8	37.9
인천	51.5	4.5	3.2	39.7	0.1	0.5	0.5	60.3
광주	14.3	8.3	11.9	0.9	61.3	2.7	0.7	38.8
대전	15.7	6.4	0.5	0.7	0.3	76.4	0.1	23.7
울산	16.3	5.5	8.2	1.0	3.9	0.9	64.2	35.8
전이효과: to others	135.2 (59.6%)	45.9 (41.5%)	30.0 (32.5%)	6.7 (14.4%)	11.3 (15.6%)	8.5 (10.0%)	2.5 (3.7%)	240.1 (34.3%)

주: VAR(p) 모형의 시차는 5임.

미치지 못하는 것으로 분석되었다.

이러한 결과를 종합하면, 아파트 가격지수 수익률은 서울이 전국 아파트 가격에 영향을 미치며, 부산과 대구 아파트 가격수익률도 일부 지역에 영향을 미치는 것으로 결론지을 수 있다.

마지막 행에 표시된 다른 지역으로의 전이효과(to others)에서, 마지막 열에 나타난 다른

지역으로부터의 전이효과(from others)를 제외하면, 지역별로 순전이효과(net spillover effect)를 계산할 수 있다. 지역별 순전이효과 크기를 비교해 보면 서울이 $164.4(=171.5-7.1)$ 로 가장 큰 것을 확인할 수 있다.

마지막으로 전체 아파트 가격지수 수익률 전이효과 즉, 전이효과 지수(spillover index: SI)는 40.5%로 나타났는데, 이 값은 마지막 행의 총합을 모든 행렬을 더한 값으로 나눈 값이다. 이는 7개 지역 아파트 가격지수 수익률은 40.5% 정도를 전이효과로 설명할 수 있다는 의미이다. 결론적으로 한국의 아파트 가격 수익률은 서울 지역의 영향이 절대적이며, 부분적으로 지역 간 상호영향을 주고 받는 것을 확인하였다.

〈표 5〉의 패널 B는 지역별 아파트 가격 변동성 모형 분석결과를 제시하고 있다. 실증분석결과는 지역별 아파트 가격 수익률 모형 분석결과와 크게 다르지 않음을 알 수 있다. 패널 B 변동성 모형의 전이효과 지수(spillover index)는 34.3%로 나타났는데, 이는 주택가격 변동성의 34.3%를 전이효과로 설명할 수 있다는 의미이다.

〈표 6〉은 〈표 5〉에서 제시한 전이효과 모형에서 서울 아파트 가격 수익률과 변동성 대신 강남의 아파트 가격을 이용해서 강건성 분석을 진행한 결과를 보여주고 있다.

패널 A는 지역별 아파트 가격 수익률 모형 분석결과이며, 패널 B는 지역별 아파트 가격 변동성 모형 분석결과이다. 실증분석결과, 강남 포함한 7개 지역 전이효과 분석결과가 서울을 포함한 전이효과 분석결과와 큰 차이를 보이지 않는 것으로 분석되었다. 따라서 이후의 분석에서는 서울 아파트 가격을 사용하여 분석하기로 한다.

다만, 전체기간을 분석한 지역별 아파트 가격 전이효과 분석결과와 경제위기를 전후한 분석결과가 다를 수 있기 때문에 본 연구에서는 분석기간에 포함된 외환위기와 금융위기를 중심으로 3개 기간으로 나누어 지역별 아파트 가격 전이효과에 차이가 있는지를 추가로 분석하기로 한다.

4. 경제위기를 고려한 지역 간 아파트 가격 수익률 전이효과

〈표 7〉은 경제위기를 중심으로 하위기간1 (1986년 2월~1997년 12월), 하위기간2(1998년 1월~2008년 8월), 하위기간 3 (2008년 9월~2020년 7월)으로 나누어 분석한 결과를 제시하고 있다.³⁾

경제위기를 고려한 하위기간 지역별 아파트 가격 전이효과 실증분석결과, 서울 아파트 가격이 다른 지역에 미치는 영향이 크기는 하지만 외환위기와 금융위기 기간 이후 점차 감소하는 결과를

3) 경제위기(외환위기, 금융위기) 기간 분류는 선행연구인 전형철, 형남원(2018)의 연구와 동일한 기간으로 설정하고 분석하였다.

〈표 6〉 지역 간 아파트가격 전이효과 분석(강남)

패널 A. 주택가격 수익률 모형 (1986년 2월~2020년 7월)								
지역	강남	부산	대구	인천	광주	대전	울산	전이효과: from others
강남	92.3	2.7	0.7	1.0	0.8	1.9	0.5	7.6
부산	27.5	54.8	6.8	0.6	2.3	6.9	1.0	45.1
대구	22.3	14.1	54.2	1.5	1.0	5.4	1.6	45.9
인천	56.3	4.6	2.4	35.6	0.1	0.4	0.5	64.3
광주	15.0	6.8	15.2	0.5	55.6	5.9	1.0	44.4
대전	19.7	2.6	0.8	1.9	1.7	73.0	0.4	27.1
울산	19.4	6.3	8.6	1.6	3.3	3.3	57.5	42.5
전이효과: to others	160.2 (63.4%)	37.1 (40.4%)	34.5 (38.9%)	7.1 (16.6%)	9.2 (14.2%)	23.8 (24.6%)	5.0 (8.0%)	276.9 (39.6%)

주: VAR(p) 모형의 시차는 5임.

패널 B. 주택가격 변동성 모형 (1986년 2월~2020년 7월)								
지역	강남	부산	대구	인천	광주	대전	울산	전이효과: from others
강남	91.0	3.8	0.7	3.1	0.2	1.0	0.3	9.1
부산	17.7	67.3	6.2	0.1	6.3	2.1	0.2	32.6
대구	15.7	18.2	61.8	0.3	1.1	2.2	0.8	38.3
인천	48.1	5.5	3.5	41.9	0.2	0.4	0.5	58.2
광주	12.2	8.9	12.2	0.7	62.6	2.8	0.7	37.5
대전	14.9	6.8	0.6	0.5	0.4	76.7	0.1	23.3
울산	15.2	6.1	8.1	1.0	4.3	1.0	64.3	35.7
전이효과: to others	123.8 (57.6%)	49.3 (42.3%)	31.3 (33.6%)	5.7 (12.0%)	12.5 (16.6%)	9.5 (11.0%)	2.6 (3.9%)	234.7 (33.5%)

주: VAR(p) 모형의 시차는 5임.

보여주고 있다.

즉, 외환위기 이전에는 서울 아파트 가격의 다른 지역 전이효과가 71.6%를 보이고 있으나, 외환위기 이후 금융위기 전까지의 기간에는 65.9%로 낮아졌다가, 금융위기 이후에는 43.3%로 낮아지는 것을 확인할 수 있다. 금융위기 이후에는 부산 아파트 가격이 다른 지역에 미치는 영향이 커지는 것으로 분석되었다.

〈표 7〉 경제위기 전후 지역별 아파트 가격 수익률 전이효과 분석

하위기간1 (1986년 2월~1997년 12월) 분석결과								
지역	서울	부산	대구	인천	광주	대전	울산	전이효과: from others
서울	78.9	5.3	0.9	5.0	0.6	5.1	4.3	21.2
부산	42.0	36.2	6.9	5.1	1.7	5.2	2.9	63.8
대구	26.5	14.3	45.6	3.5	1.2	4.8	4.1	54.4
인천	51.0	5.6	2.9	35.4	0.5	2.0	2.6	64.6
광주	31.6	4.2	10.0	1.9	43.8	8.2	0.3	56.2
대전	26.2	1.1	3.2	4.7	4.4	56.5	3.9	43.5
울산	21.2	6.9	7.8	5.0	7.5	4.4	47.3	52.8
전이효과: to others	198.5 (71.6%)	37.4 (50.8%)	31.7 (41.0%)	25.2 (41.6%)	15.9 (26.6%)	29.7 (34.5%)	18.1 (27.7%)	356.0 (50.9%)

주: VAR(p) 모형의 시차는 2임⁴⁾.

하위기간2 (1998년 1월~2008년 8월) 분석결과								
지역	서울	부산	대구	인천	광주	대전	울산	전이효과: from others
서울	79.4	5.1	12.0	1.2	0.5	1.6	0.3	20.7
부산	23	50.9	18.8	4.1	1.0	1.1	1.1	49.1
대구	21.9	5.3	60.5	3.9	6.1	1.4	0.9	39.5
인천	56.8	7.9	4.8	25.2	0.9	3.0	1.4	74.8
광주	13.3	2.9	35.3	2.5	42.8	0.3	2.9	57.2
대전	12.3	4.6	12.0	9.4	2.6	57.2	1.9	42.8
울산	25.9	1.9	17.8	1.5	7.7	1.5	43.7	56.3
전이효과: to others	153.2 (65.9%)	27.7 (35.2%)	100.7 (62.5%)	22.6 (47.3%)	18.8 (30.5%)	8.9 (13.5%)	8.5 (16.3%)	340.0 (48.6%)

주: VAR(p) 모형의 시차는 2임.

4) 하위기간의 지역 간 아파트 가격 전이효과 분석을 위한 VAR(p) 모형 시차분석 결과, SC와 HQ 통계량 적정 시차는 전체기간과 다르게 시차 2로 분석되었다.

〈표 7〉 계속

하위기간3 (2008년 9월~2020년 7월) 분석결과								
지역	서울	부산	대구	인천	광주	대전	울산	전이효과: from others
서울	88.6	6.4	0.6	1.6	0.4	0.8	1.6	11.4
부산	3.6	65.4	2.7	0.8	13.8	12.4	1.4	34.7
대구	6.4	13.1	66.1	1.5	0.1	9.6	3.2	33.9
인천	31.9	5.7	2.7	54.5	2.4	2.5	0.3	45.5
광주	11.4	34.0	6.5	2.2	32.1	11.4	2.5	68
대전	12.0	8.2	6.1	1.2	0.8	67.7	4.1	32.4
울산	2.4	25.5	14.2	0.5	5.3	14.7	37.4	62.6
전이효과: to others	67.7 (43.3%)	92.9 (58.7%)	32.8 (33.2%)	7.8 (12.5%)	22.8 (41.5%)	51.4 (43.2%)	13.1 (25.9%)	288.0 (41.2%)

주: VAR(p) 모형의 시차는 2임.

그리고 7개 지역 아파트 가격지수 수익률 전이효과 지수는 하위기간 1에는 50.9%, 하위기간 2에는 48.6%, 하위기간 3에는 41.2%로 점차 감소하는 것으로 분석되었다.

결론적으로 경제위기를 고려한 지역 간 아파트 가격 전이효과 분석결과를 요약하면, 경제위기를 고려한 하위분석 결과는 전체 분석결과와 대체로 유사한 결과를 보이고 있지만, 전체적인 전이효과 설명력은 하위기간 1에서 하위기간 3 그리고 하위기간 3으로 갈수록 감소하는 것으로 나타났다. 이는 서울 아파트 가격이 전국에 미치는 영향도 낮아지는 것으로, 서울의 집중도가 다소 낮아지는 것으로 해석할 수 있다.

5. 전국 16개 지역 지역별 아파트 가격 전이효과

본 연구는 서울, 광역시 및 도를 포함한 16개 지역별 아파트 전이효과를 분석하였다. 분석기간은 아파트 가격 시계열 자료를 확인할 수 있는 2003년 7월부터 2020년 7월까지로 한다. 분석방법은 수익률 모형과 변동성 모형을 모두 사용하기로 한다.

〈표 8〉의 패널 A의 지역별 아파트 가격 수익률 모형 분석결과를 살펴보면 지역별로 아파트 가격 수익률이 상호 영향을 미치는 전이효과를 확인할 수 있다.

16개 지역별 아파트 가격 수익률 전이효과 분석 결과, 7개 지역 분석과 같이 서울의 경우, 자기

자신의 영향을 가장 많이 받지만, 7개 지역 분석결과와는 다르게, 서울의 아파트 가격 수익률은 인천이 아닌 경기도 아파트 가격 수익률에 가장 큰 영향을 주는 것으로 나타났다. 인천은 경기도 다음으로 인천 아파트 가격 수익률에 영향을 미치는 것으로 분석되었다.

이러한 분석 결과는 패널 B의 지역별 아파트 가격 변동성 모형 분석 결과도 동일한 것으로 나타나, 7개 지역을 대상으로 한 분석 결과와는 다르게 서울 아파트 가격은 경기도 아파트 가격에 가장 큰 영향을 미치는 것을 다시 확인할 수 있다.

패널 A와 B의 서울의 마지막 행의 값은 각각 174.9와 123.4이며, 이는 서울지역 아파트 가격 수익률이 다른 지역에 미치는 전이효과의 크기를 의미하는 것이다.

패널 A의 부산의 아파트 가격 수익률도 자신의 영향이 가장 큰 것으로 나타났으며, 다음으로 대구·경북 아파트 가격 수익률에 영향을 미치는 것으로 분석되었다. 그리고 광주·전남·전북, 대전·충남·충북 아파트 가격 수익률에도 영향을 미치는 것으로 분석되었다.

패널 A의 지역별 순전이효과 크기를 비교해 보면 서울이 156.2로 가장 큰 것으로 나타났으며, 다음으로 지역별 순전이효과가 큰 지역은 부산으로 분석되었다.

그리고 패널 A의 16개 지역 아파트 가격지수 수익률 전이효과 지수(spillover index)는 59.9%로 나타났으며, 패널 B의 16개 지역 아파트 가격지수 변동성 전이효과 지수는 55.9%로 분석되었다.

이러한 결과는 16개 지역의 아파트 가격 수익률과 변동성의 50% 이상이 전이효과로 설명된다는 의미로 이항용, 이진(2012), 방두완 외(2019)의 연구결과와 일치하는 것으로 볼 수 있으며, 따라서 지역별 아파트 가격은 서로 영향을 많이 주고받는다 고 결론지을 수 있다.

V. 결론

한국은 부동산 자산의 비중이 높아 부동산 가격변동에 관한 관심이 매우 높고, 특히 서울이나 강남지역 등 특정 지역의 주택가격이 다른 지역의 주택가격에 영향을 미치는 지에 대해 민감하게 반응하는 경향이 있다.

이러한 주택가격의 전이효과를 분석하기 위해 본 연구는 기존 선행연구들이 제시한 주택가격의 변동성을 중심으로 분석 이외에 주택가격 자체의 수익률을 중심으로 지역 간 전이효과를 분석하였다.

이를 위해 본 연구는 Diebold and Yilmaz (2009)의 연구 방법론을 준용하여, 지역별 주택가격 전이효과를 분석하였다. 한결음 나아가 본 연구는 선행연구와 다르게 서울, 광역시, 그리고 도 단위 지역을 포함한 16개 지역을 대상으로 지역 간 전이효과를 분석하기로 하고, 경제위기 전후를 중심으로 지역 간 전이효과에 차이가 있는 지를 분석하였다. 분석기간을 세부적으로 구분한 이유는

〈표 8〉 전국 16개 지역 아파트 가격 전이효과 분석

패널 A. 주택가격 수익률 모형 (2003년 7월~2020년 7월)																	
지역	서울	부산	대구	인천	광주	대전	울산	경기도	강원도	충청 남도	충청 북도	전라 남도	전라 북도	경상 남도	경상 북도	제주도	전이 효과: from others
서울	81.5	4.2	0.4	0.2	0.5	0.8	4.5	1.2	0.1	0.6	1.8	0.8	1.3	0.9	0.2	1.2	18.7
부산	2.0	57.8	2.1	0.5	12.1	3.3	0.4	2.1	5.3	3.1	3.6	4.0	1.6	1.3	0.6	0.3	42.3
대구	10.9	10.8	44.8	3.2	1.9	4.6	0.7	2.9	3.9	7.2	2.4	0.6	2.4	1.1	2.3	0.4	55.3
인천	37.0	3.1	1.5	33.7	0.2	1.9	5.9	4.2	1.7	0.2	3.3	1.7	2.5	1.5	0.6	1.0	66.3
광주	3.4	18.8	4.8	4.3	36.2	3.5	0.8	1.9	4.0	8.5	7.2	0.8	2.2	0.6	3.0	0.1	63.9
대전	10.4	15.6	6.8	4.4	2.2	43.1	0.8	0.7	4.5	0.4	7.2	0.3	0.4	0.8	1.2	1.2	56.9
울산	19.4	4.2	1.5	1.1	1.9	5.0	44.2	3.4	0.6	9.6	1.0	4.3	0.4	2.2	0.2	0.8	55.6
경기도	63.1	2.9	0.3	2.4	0.4	0.5	12.0	12.2	0.6	0.4	1.3	1.2	1.0	1.0	0.1	0.5	87.7
강원도	0.5	14.7	3.9	1.9	5.2	10.4	0.2	1.2	40.4	7.7	2.6	2.2	5.9	1.2	1.8	0.1	59.5
충청남도	4.0	12.1	6.4	1.6	0.8	14.0	1.2	1.9	3.0	40.3	0.8	5.0	3.4	3.9	1.1	0.5	59.7
충청북도	2.9	11.0	7.6	0.3	3.0	7.4	3.8	2.0	0.4	8.7	47.5	2.0	1.0	1.6	0.5	0.3	52.5
전라남도	5.6	10.5	2.1	2.4	1.6	8.7	0.9	0.5	2.6	11.9	2.4	44.8	0.7	3.0	1.7	0.8	55.4
전라북도	5.5	19.3	3.0	2.7	2.5	7.8	0.9	1.2	0.9	6.7	3.6	4.5	39.1	0.8	1.2	0.4	61.0
경상남도	3.2	3.8	26.6	0.9	1.0	4.3	7.0	2.9	1.7	16.6	0.9	5.2	0.6	24.3	0.5	0.4	75.6
경상북도	4.2	25.8	3.0	0.8	6.7	12.0	0.1	2.6	6.7	7.3	1.3	8.1	7.3	2.5	11.7	0.1	88.5
제주도	2.8	18.3	0.7	1.2	5.2	3.8	2.9	1.0	4.0	1.8	1.6	3.2	3.6	4.9	4.9	40.0	59.9
전이효과: to others	174.9 (68.2%)	175.1 (75.2%)	70.7 (61.2%)	27.9 (45.3%)	45.2 (55.5%)	88.0 (67.1%)	42.1 (48.8%)	29.7 (70.9%)	40.0 (49.8%)	90.7 (69.2%)	41.0 (46.3%)	43.9 (49.5%)	34.3 (46.7%)	27.3 (52.9%)	19.9 (63.0%)	8.1 (16.8%)	958.8 (59.9%)

주: VAR(p) 모형의 시차는 2임

〈표 8〉 계속

패널 B. 주택가격 변동성 모형 (2003년 7월~2020년 7월)																	
지역	서울	부산	대구	인천	광주	대전	울산	경기도	강원도	충청 남도	충청 북도	전라 남도	전라 북도	경상 남도	경상 북도	제주도	전이효과: from others
서울	79.0	5.7	2.9	0.1	2.6	0.5	4.4	0.5	1.1	0.3	1.4	0.6	0.2	0.2	0.4	0.1	21.0
부산	0.8	60.7	3.7	1.5	6.4	7.3	1.2	0.1	2.8	2.1	3.0	5.1	3.6	0.8	0.7	0.3	39.4
대구	3.3	13.2	58.3	1.0	3.8	3.8	1.0	3.4	1.4	3.0	3.0	0.2	1.2	2.2	0.5	0.5	41.5
인천	31.0	1.7	1.1	40.9	4.0	1.3	4.7	3.0	0.2	4.9	0.4	1.0	2.6	1.2	1.3	0.6	59.0
광주	1.1	27.3	5.0	4.1	33.5	5.0	0.3	0.2	4.2	10.6	1.1	0.6	2.9	0.1	3.8	0.2	66.5
대전	1.5	13.1	9.2	3.0	4.5	54.6	1.2	1.6	0.5	0.6	2.5	1.1	2.6	0.9	2.8	0.5	45.6
울산	15.3	5.0	0.6	4.6	6.5	3.7	51.8	0.3	0.9	4.7	1.2	1.5	0.5	1.1	0.1	2.3	48.3
경기도	54.5	3.3	1.6	2.2	2.7	0.9	13.7	14.4	1.3	0.3	0.4	1.7	0.7	0.3	1.3	0.4	85.3
강원도	0.8	15.0	3.0	1.8	11.1	6.8	0.8	1.1	43.3	6.2	1.5	0.6	5.0	0.9	1.8	0.2	56.6
충청남도	0.1	17.5	3.3	8.9	9.5	4.1	0.4	0.1	2.6	47.5	0.4	2.0	1.4	0.3	1.8	0.1	52.5
충청북도	1.2	7.6	4.4	1.3	16.5	2.8	1.8	3.9	1.0	8.4	47.6	0.5	0.5	2.0	0.2	0.1	52.2
전라남도	3.6	18.3	3.0	2.1	0.8	6.7	1.5	0.6	3.6	4.1	2.4	47.8	1.1	1.5	0.9	2.1	52.3
전라북도	2.1	26.3	1.2	1.4	1.8	6.9	2.1	1.8	0.7	10.3	1.7	6.5	33.8	1.3	1.7	0.3	66.1
경상남도	1.4	4.5	24.2	4.3	3.4	3.4	5.2	1.2	0.3	13.5	0.2	1.3	0.2	34.8	0.1	2.0	65.2
경상북도	1.8	30.5	7.7	3.4	5.0	10.3	1.2	0.4	4.2	8.3	1.4	4.1	5.4	0.1	15.8	0.4	84.2
제주도	4.9	15.5	1.6	0.4	9.6	3.0	2.6	2.4	3.5	0.5	1.3	1.3	2.6	5.4	3.4	42.1	58.0
전이효과: to others	123.4 (61.0%)	204.5 (77.1%)	72.5 (55.4%)	40.1 (49.5%)	88.2 (72.5%)	66.5 (54.9%)	42.1 (44.8%)	20.6 (58.9%)	28.3 (39.5%)	77.8 (62.1%)	21.9 (31.5%)	28.1 (37.0%)	30.5 (47.4%)	18.3 (34.5%)	20.8 (56.8%)	10.1 (19.3%)	893.7 (55.9%)

주: VAR(p) 모형의 시차는 2임

경제적 충격(economy shock) 이전과 이후의 지역 간 주택가격 전이효과에 차이가 있을 것으로 기대되었기 때문이다.

실증분석을 위해 본 연구는 먼저 지역별 주택가격 수익률 변수의 그랜저 인과관계를 분석하였다. 인과관계 분석결과, 서울과 부산, 서울과 광주, 서울과 대전은 서로 그랜저 인과관계가 존재하는 것으로 분석되었다. 이외에도 서울의 아파트 가격 수익률은 대구, 인천 그리고 울산에 통계적으로 유의적인 그랜저 인과관계가 영향을 미치는 것으로 나타나 서울은 모든 지역의 아파트 가격 수익률에 그랜저 인과관계가 있는 것으로 분석되었다.

본 연구는 7개 지역(서울, 부산, 대구, 인천, 광주, 대전, 울산)의 지역별 아파트 가격지수 수익률과 변동성을 이용하여 시차가 5인 VAR₍₅₎ 모형을 추정하였다. 분석기간은 7개 지역 아파트 가격을 확인할 수 있는 1986년 1월부터 2020년 7월까지이다.

지역별 아파트 가격 수익률 분석결과, 지역별로 상호 영향을 미치는 전이효과를 확인하였으며, 서울 아파트 가격 수익률은 자기자신의 영향을 가장 많이 받으며, 다음으로 인천의 아파트 가격 수익률에 큰 영향을 미치는 것으로 분석되었다. 그리고 그랜저 인과관계 분석결과, 동일하게 서울의 아파트 가격 수익률은 부산, 대구, 대전, 울산, 광주의 순으로 모든 지역 아파트 가격 수익률에 영향을 미치는 것으로 분석되었다.

반면, 서울의 아파트 가격 수익률에 다른 지역이 미치는 영향은 7.1로 나타나 매우 낮은 것으로 분석이 되었으며, 부산의 아파트 가격이 다른 지역에 비해 상대적으로 가장 큰 영향을 미치는 것으로 나타났다.

부산의 아파트 가격 수익률도 자신의 영향이 가장 큰 것으로 나타났으며, 다음으로 대구 아파트 가격 수익률에 영향을 미치는 것으로 분석되었다. 이러한 결과를 종합하면, 아파트 가격지수 수익률은 서울이 전국 아파트 가격에 영향을 미치며, 부산과 대구 아파트 가격수익률도 일부 지역에 영향을 미치는 것으로 결론지을 수 있다. 지역별로 순전이효과(net spillover effect) 분석결과, 서울이 164.4로 순전이효과가 가장 큰 것으로 분석되었다.

마지막으로 전체 아파트 가격지수 수익률 전이효과 지수(spillover index)는 40.5%로 나타났는데, 이는 7개 지역 아파트 가격지수 수익률은 40.5% 정도를 전이효과로 설명할 수 있다는 의미이다. 지역별 아파트 가격 변동성 모형 분석결과도 수익률 모형 분석결과와 크게 다르지 않은 것으로 분석되었다.

서울 아파트 가격 대신 강남의 아파트 가격을 이용해서 강건성 분석한 결과도 서울 아파트 가격 분석 결과와 큰 차이를 보이지 않는 것으로 분석되었다.

추가로 본 연구에서는 분석기간에 포함된 외환위기와 금융위기를 중심으로 3개 기간으로 나누어

지역별 아파트 가격 전이효과에 차이가 있는지를 분석하였다.

경제위기를 고려한 하위기간 지역별 아파트 가격 전이효과 실증분석결과, 서울 아파트 가격이 다른 지역에 미치는 영향이 크기는 하지만, 외환위기와 금융위기 기간 이후 점차 감소하는 결과를 보여주고 있다.⁵⁾

즉, 외환위기 이전에는 서울 아파트 가격의 다른 지역 전이효과가 71.6%를 보이고 있으나, 외환위기 이후 금융위기 전까지의 기간에는 65.9%로 낮아졌다가, 금융위기 이후에는 43.3%로 낮아지는 것을 확인하였다. 이는 서울아파트 가격이 전국에 미치는 영향도 낮아지는 것으로, 서울의 집중도가 다소 낮아지는 것으로 해석할 수 있다.

그리고 7개 지역 아파트 가격지수 수익률 전이효과 지수(spillover index)는 하위기간 1에는 50.9%, 하위기간 2에는 48.6%, 하위기간 3에는 41.2%로 점차 감소하는 것으로 분석되었다.

마지막으로 본 연구는 서울, 광역시 및 도를 포함한 16개 지역별 아파트 전이효과를 분석하였는데, 분석기간은 2003년 7월부터 2020년 7월까지이다. 16개 지역별 아파트 가격 수익률 모형 전이효과 분석결과 7개 지역 분석과 같이 서울의 경우, 자기 자신의 영향을 가장 많이 받지만, 7개 지역 분석결과와는 다르게, 서울의 아파트 가격 수익률은 경기도 아파트 가격 수익률에 가장 큰 영향을 주는 것으로 나타났으며, 그 다음으로 인천 아파트 가격 수익률에 영향을 미치는 것으로 분석되었다. 지역별 아파트 가격 변동성 모형 분석결과도 수익률 모형과 유사한 것으로 분석되었다.

16개 지역 아파트 가격지수 수익률 모형 전이효과 지수는 59.9%로 나타났으며, 변동성 모형 전이효과의 설명력은 55.9%로 분석되었는데, 이는 16개 지역의 아파트 가격 수익률과 변동성의 50% 이상이 전이효과로 설명된다는 의미로 이항용, 이진(2012), 방두완 외(2019)의 연구결과와 일치하는 것으로 볼 수 있으며, 따라서 한국 지역별 아파트 가격은 서로 영향을 많이 주고받는 것으로 결론지을 수 있다.

본 연구는 Diebold and Yilmaz(2009)의 수익률 및 변동성 모형을 적용하여 16개 지역의 지역별 아파트 가격 전이효과를 분석하였다는 점에서는 의의가 있지만, 지역별로 정책효과와의 상관관계를 고려하지 못하였고, 이는 향후의 연구과제로 남겨둔다.

5) 한국은 지역별 주택가격 전이효과 측면에서, 지역 상호 간의 영향뿐만 아니라, 서울과 수도권의 주택가격 변동이 다른 지역 주택가격을 선도하는 것으로 알려져 왔다. 본 연구의 분석결과도 서울과 강남의 주택가격이 다른 지역의 주택가격에 많은 영향을 미치고 있지만, 금융위기 이후 이러한 주택가격 전이효과가 조금씩 약해지는 모습을 보여주고 있다.

참고문헌

- 강임호, 조영진, 2014, “시차공적분을 통해 본 수도권과 비수도권 아파트시장의 특성”, 국토연구 80, 67-81.
- 김의준, 김양수, 신명수, 2000, “수도권 아파트 가격의 지역적 인과성분석”, 국토계획 34권 4호, 109-117.
- 문규현, 이동희, 2011, “강남아파트시장은 전국아파트시장을 선도하는가?”, 산업경제연구 24권 1호, 115-136.
- 박영준, 김기호, 2017, “수도권 주택가격 변동의 동조화와 변동성 전이”, 부동산학보 69권, 131-145.
- 박해선, 김승년, 2014, “주택가격의 지역 간 상호의존성이 관한 연구: 서울지역 아파트 매매가격을 중심으로”, 산업경제연구 27권 2호, 565-583.
- 박현수, 안지아, 2009, “VAR모형을 이용한 부동산가격 변동요인에 관한 연구”, 부동산연구 19권 1호, 27-69.
- 방두완, 권혁신, 김명현, 2019, “FAVAR를 이용한 지역별 아파트 경기지수 전이효과 분석”, 주택연구 27권 3호, 147-171.
- 서승환, 2007, “주택가격 변화의 지역연관성에 관한 연구: 강남구 물결효과를 중심으로”, 서울도시연구 8권 4호, 1-13.
- 이상경, 2003, “서울 주택시장으로부터 지방주택시장으로의 가격 및 변동성 이전효과 연구”, 국토계획 38권 7호, 81-90.
- 이항용, 이진, 2014, “아파트 매매가격의 지역 간 전이효과: 일반화 예측오차 분산분해를 이용한 7개 대도시를 중심으로”, 국토연구 82권, 3-15.
- 전해정, 2013, “서울시 주택가격 변동성 및 이전효과에 관한 실증분석”, 지역연구 29권 4호, 83-98.
- 전형철, 형남원, 2018, “주택의 매매 및 전세가격의 확산효과에 대한 분석: 강남효과를 중심으로”, 주택연구 26권 제1호, 63-88.
- 최경욱, 형남원, 2010, “가격변수 불확실성과 경기변동간의 관계”, 경제분석 16권 제3호, 1-41.
- Diebold, F. and Yilmaz, K., 2009, “Measuring financial asset return and volatility spillovers, with application to global equity markets”, Economic Journal 119(1), 158-171.
- Diebold, F. and Yilmaz, K., 2012, “Better to give than to receive: Predictive directional

measurement of volatility spillovers”, International Journal of Forecasting 28(1), 57-66.

Hull, J. C., 2005, “Fundamentals of Futures and Options Markets”, Prentice Hall.

(논문 접수일: 2020.08.22. 수정논문 접수일: 2020.11.24. 논문 채택일: 2020.12.11.)

부록

〈부록 표 1〉 그랜저 인과관계 분석결과

귀무가설	관측치	F 통계량	유의확률
부산 $\Rightarrow$ 서울	409	5.301	0.000
서울 $\Rightarrow$ 부산		16.544	0.000
대구 $\Rightarrow$ 서울	409	1.346	0.244
서울 $\Rightarrow$ 대구		7.369	0.000
인천 $\Rightarrow$ 서울	409	1.795	0.113
서울 $\Rightarrow$ 인천		44.521	0.000
광주 $\Rightarrow$ 서울	409	4.375	0.001
서울 $\Rightarrow$ 광주		3.591	0.003
대전 $\Rightarrow$ 서울	409	2.489	0.031
서울 $\Rightarrow$ 대전		20.871	0.000
울산 $\Rightarrow$ 서울	409	1.431	0.212
서울 $\Rightarrow$ 울산		6.589	0.000
대구 $\Rightarrow$ 부산	409	12.247	0.000
부산 $\Rightarrow$ 대구		1.778	0.116
인천 $\Rightarrow$ 부산	409	4.251	0.001
부산 $\Rightarrow$ 인천		5.660	0.000
광주 $\Rightarrow$ 부산	409	21.947	0.000
부산 $\Rightarrow$ 광주		3.373	0.005
대전 $\Rightarrow$ 부산	409	5.071	0.000
부산 $\Rightarrow$ 대전		6.988	0.000
울산 $\Rightarrow$ 부산	409	4.848	0.000
부산 $\Rightarrow$ 울산		4.555	0.001
인천 $\Rightarrow$ 대구	409	2.676	0.022
대구 $\Rightarrow$ 인천		14.269	0.000
광주 $\Rightarrow$ 대구	409	1.473	0.198
대구 $\Rightarrow$ 광주		12.529	0.000
대전 $\Rightarrow$ 대구	409	10.674	0.000
대구 $\Rightarrow$ 대전		4.916	0.000
울산 $\Rightarrow$ 대구	409	1.459	0.202

〈부록 표 1〉 계속

귀무가설	관측치	F 통계량	유의확률
대구 ⇔ 울산		11.423	0.000
광주 ⇔ 인천	409	8.107	0.000
인천 ⇔ 광주		4.033	0.001
대전 ⇔ 인천	409	2.301	0.044
인천 ⇔ 대전		12.754	0.000
울산 ⇔ 인천	409	6.051	0.000
인천 ⇔ 울산		6.300	0.000
대전 ⇔ 광주	409	6.848	0.000
광주 ⇔ 대전		6.412	0.000
울산 ⇔ 광주	409	3.524	0.004
광주 ⇔ 울산		6.976	0.000
울산 ⇔ 대전	409	2.075	0.068
대전 ⇔ 울산		3.091	0.010

A Study on the Spillover Effect of the Regional Housing Price

Doo-Won Bang* · Hyuck-Shin Kwon**

Abstract

We estimated the spillover effect of the regional apartment prices using the methodology proposed by Dibold and Yilmaz (2009). We analyzed the transition effects of the seven regions for the period from January 1986 to July 2020, and we also estimated those of the 16 regions from July 2003 to July 2020. As a result of analyzing apartment price returns by seven regions, we found the transfer effect, that each region has a mutual influence. The return on apartment prices in Seoul was most affected by themselves and has a great influence on the Incheon apartment price return. When we consider the period of the economic crisis, the effect of apartment prices in Seoul is large on other regions, but it appears that they gradually decrease after the period of the foreign exchange crisis and the financial crisis. Finally, we find that more than 50% of apartment price returns and volatility in 16 regions were explained by the spillover effect. So, we conclude that apartment prices by region in Korea have a lot of influence on each other.

Keywords : Regional Apartment Spillover Effect, Generalized Forecast Error Variance Decomposition, Spillover Index, Net Spillover Effect

* Doo-Won Bang, First author, Korea Housing & Urban Guarantee Corporation, Housing Urban Finance Institute, Senior Research Fellow, doowoan@khug.or.kr

** Hyuck-Shin Kwon, Corresponding author, Korea Housing & Urban Guarantee Corporation, Housing Urban Finance Institute, Senior Manager (Ph. D), hskwon@khug.or.kr

© Copyright 2020 Housing Finance Research Institute. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

기계학습 알고리즘을 이용한 주택가격감정 시스템의 구축 및 평가: XGBoost, LightGBM, CatBoost 알고리즘에 기반하여

홍정의*

요 약

주택 가격 감정 모형은 아주 작은 비용으로 대량의 부동산 감정을 동시에 수행할 수 있기 때문에, 모기지 담보가치 추정, 주택가격지수 산출, 재산세 추정 등과 같이 대규모 자산 가치평가를 빈번하게 수행해야 하는 모든 활동 영역에서 다양하게 활용될 수 있다. 최근에는 급격히 성장하는 데이터 수집·분석 기법들을 주택 감정 모형의 정확성을 상승시키는 데에 활용하는 연구들이 늘어가고 있다. 본문의 목적은 효율성과 예측력이 높은 것으로 알려진 세 가지 알고리즘(XGBoost, LightGBM, CatBoost)을 통해 주택 감정 모형을 구축하고, 그 성능과 특징 및 활용방법을 분석하는 것이다. 본문은 2009년부터 2019년까지 서울에서 거래된 아파트 매매 데이터 620,617건을 통해 헤도닉 모형과 기계 학습 모형 기반의 주택 가치 감정 모형의 예측력을 비교하였다. 분석 결과는 다음과 같다; 첫째, 기계 학습 모형의 예측력은 상대적인 측면 (헤도닉 가격 모형에 비해) 뿐 아니라, 절대적인 측면 (모형의 실용적 활용 가능성)에서도 상당히 높게 나타났다. 헤도닉 모형의 경우, 시장 가격에 대한 예측의 평균 백분율 오차가 약 11.5% 내외인 반면, XGBoost·LightGBM·CatBoost는 각각 3.7%, 3.8%, 3.6%에 불과했다. 두 번째로, CatBoost 알고리즘이 평균 예측력에서나, 이상치 발생 빈도에서나 다른 두 알고리즘에 비해 더 우수한 것으로 나타났다. 세 번째로, 소프트 보팅을 통한 세 알고리즘의 앙상블 모형을 구축하는 경우, 개별 알고리즘보다 더 예측력을 상승시킬 수 있음을 확인하였다.

핵심어 : 기계학습, XGBoost, LightGBM, CatBoost, 대량 주택 감정

* 홍정의, 주저자, 한동대학교 경영경제학부 교수, 경제학 박사, hwgh024@handong.edu

© Copyright 2020 Housing Finance Research Institute. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

이른바 4차 산업혁명 시기에 접어들면서, 기계학습, 인공 신경망 기법 등을 통한 각 분야의 혁신이 가시화되고 있다. 이러한 데이터 과학 기술을 주택 시장 연구 및 관리에 접목한 프롭테크(proptech)는 당분간 흐름을 이어갈 것으로 예상된다. 프롭테크는 다양하게 적용될 수 있지만, 특히 빅 데이터와 그에 대한 다양한 기계 학습 기법을 바탕으로 전통적 모형에 비해 예측력이 뛰어난 주택 가치 감정 모형을 만드는 데에도 활용될 수 있다는 점에서 흥미롭다. 주택 가치 감정 모형 또는 대량 감정 모형이란 일정한 수리·통계 모형 또는 컴퓨터 공학적 기법을 적용하여 부동산의 해당 시점 시장 가격을 예측하는 시스템 (Standard of Mass Appraisal of Real Property)을 의미한다.¹⁾ 전통적으로 주택의 감정평가는 감정평가사를 통한 추정을 의미하는데, 이는 직접적이고 정교한 반면, 상대적으로 비용이 많이 드는 방식이다. 한편, 대량 감정 모형의 경우 매우 작은 비용으로 대량의 부동산 감정을 동시에 실시할 수 있기 때문에, 모기지 담보가치 추정, 건설 투자 입지 분석, 주택가격지수 산출, 재산세 산정 등과 같이 대규모의 가치평가를 빈번하게 수행해야 하는 모든 활동 영역에 다양하게 활용될 수 있다(Zhou, Ji, et al., 2018, Wang and Li, 2020).

최근의 데이터 분석 기술이 특히 주택 가격 감정 분야에서 두각을 드러낼 수 있는 이유는 크게 두 가지가 있다. 첫째는, 근래의 데이터 수집 및 가공 기술의 발전으로 인해 종래에는 관찰하지 못했던 변수들을 관찰할 수 있게 되었으며, 수집할 수 있는 표본의 수 또한 대폭 늘어났다는 것이다. 특히 GIS를 통한 입지 정보를 이용할 수 있게 되면서부터 이로 인한 주택 가격의 차이를 효과적으로 포착할 수 있는 다양한 모형들을 활용할 수 있게 되었다. 표본 수의 확보 문제는 모형의 복잡성이 높을수록 중요해지는데, 이는 복잡한 모형일수록 과적합 문제에 취약해지기 때문이다. 과적합이란 특정 표본 집단의 특성을 지나치게 적응하는 상황으로, 결과적으로 분석 대상 전반에 대한 모형의 예측력을 훼손시킨다. 이른바 빅데이터 기술로 불리는 데이터 수집 가공 능력의 상승은 이처럼 보다 복잡하고 비선형성이 높은 모형을 활용할 수 있는 길을 열어주고 있다. 두 번째로, 데이터 분석 모형 기법 자체의 비약적인 발전이다. 이는 앞서 언급한 데이터 수집 가공 기술의 발전과 맞물려 최근 폭발적인 추세로 성장하고 있다. 이러한 모형들은 전통적 계량모형에 비해 훨씬 높은 복잡성과 비선형성을 다룰 수 있으며, 그에 따라 특히 변수 예측 문제에서 높은 성능을 가지게 되었다. 주택가격 감정 모형에서 가장 중요한 것 역시 예측의 정확성과 효율성, 즉, 가능한 한 적은 종류의 변수를 통해 가능한 한 정확하고 안정적인 예측을 수행하는 것이다. 그러므로 최근의 데이터 분석 기법들은 주택 감정 시스템에서도 일대 혁신을 불러올 수 있을 것으로 기대된다.

1) IAAO. Standard on Mass Appraisal of Real Prop.; IAAO: Kansas City, MO, USA, 2017.

근래 다양한 데이터 분석 기법이 소개되면서, 이를 주택 가치 감정 모형에 적용한 연구들도 늘어가는 추세이다. 대표적으로, 의사결정나무 (Fan et al., 2006), 랜덤포레스트 (Antipov and Pokryshevskaya, 2012; Hong et al., 2020), 서포트 벡터 머신 (Gu, Zhu, and Jiang, 2011), 인공신경망 (McCluskey and Anand, 1999; Selim, 2009), 그래디언트 부스트 (Park and Bae, 2015) 등이 주택 감정 모형에 사용될 수 있음이 알려져 있다.

본문의 목적은 최근 그래디언트 부스트를 기반으로 효율성과 예측력을 상승시키는 것으로 알려진 세가지 알고리즘(XGBoost, LightGBM, CatBoost)을 통해 대량 주택 감정 모형을 구축하고, 그 성능과 특징 및 활용방법을 분석하는 것이다. 이러한 알고리즘들은 수행시간이 느리고, 과적합 규제가 없는 그래디언트 부스트의 한계를 극복한 것으로 평가되며, 특히 분류 문제에 있어서 다른 기계 학습 기법에 비해 예측 성능이 높은 것으로 알려져 있다. 이 세 알고리즘은 비교적 최근에 소개되었기 때문에, 랜덤 포레스트나 인공 신경망 등과 같이 더 일찍 알려진 알고리즘들에 비해 주택 가격 감정 모형 분야에서 얼마나 효율적으로 사용될 수 있는지 상대적으로 덜 알려져 있다.

또한, 본문은 이 세 알고리즘의 예측력이 전통적인 방식(특히 헤도닉 모형)에 비해 얼마나 상승하는지를 비교 분석하였다. 전통적인 선형 헤도닉 모형은 몇 가지 가정 하에서 자산의 가격을 설명변수의 회귀식으로 표현하는데(Rosen, 1974), 이러한 방식의 장점은 간단한 선형 다변량 회귀분석을 통해서 감정모형 시스템을 구축할 수 있다는 데에 있다. 즉, 변수 간 관계를 직관적으로 해석하거나 대중에게 설명하기 쉬운 특징을 가지고 있다는 것이다. 그러나, 이와 같은 단순성과 직관성이 동시에 약점이 되기도 한다. 완전 경쟁, 분리 가능한 효용구조, 완전히 통합된 시장구조 등의 엄격한 경제학적 가정이 동시에 만족되지 않는 경우, 회귀식으로 표현 가능한 단순한 함수형태는 실제의 부동산 시장이 가지고 있는 복잡성을 포착하기에 충분하지 않기 때문이다(Sheppard, 1999; Malpezzi, 2002). 결과적으로 헤도닉 모형에 기반한 부동산 감정은 함수 형태의 단순성이나 관찰 가능한 변수의 제한으로 인한 설명력 저하 문제를 갖게 된다. 한편, 기계학습 기반의 모형은 기본적으로 변수 간 관계를 하나의 함수 형태로 묘사하는 것이 아니기 때문에, 실제 시장의 비선형성을 포착하는 데에 있어 헤도닉 모형보다 유리한 특징을 가지고 있다. 무엇보다도, 헤도닉 모형에서는 입지 효과를 포착하기 위해서 주요 시설과의 거리 또는 이동 시간을 정교하게 추정해야 하는 데에 반해, 위 알고리즘들은 주택의 입지 정보를 직접 이용하여 주택 가치 간의 비선형적 차이 자체를 학습하는 데에 사용할 수 있다. 입지 효과는 주택 가치 간 차이를 설명하는 데에 있어 가장 중요한 요소일 뿐 아니라 그 복잡성도 높기 때문에, 이를 효과적으로 포착하는 만큼 전통적 방식에 비해 더 높은 예측력을 갖는 것으로 나타나게 될 것이다.

본문은 2009년부터 2019년까지 서울에서 거래된 아파트 매매 데이터 620,617건을 통해 헤도닉

모형과 세 알고리즘 (XGBoost · LightGBM · CatBoost) 기반의 주택 가치 감정 모형을 비교 분석하였다. 결과를 간략히 요약하면 다음과 같다; 첫째, 헤도닉 모형에 비해 기계 학습 알고리즘의 예측력이 월등히 높게 나타났다. 헤도닉 모형의 경우, 시장 가격에 대한 예측의 평균 백분율 오차가 약 11.5% 내외인 반면, XGBoost · LightGBM · CatBoost는 각각 3.7%, 3.8%, 3.6%에 불과했다. 이는 기계학습 기반의 예측력이 다른 모형과의 상대적인 비교뿐 아니라, 절대적 수치 측면에서도 유의미하게 낮아질 수 있다는 것을 의미한다. 즉, 절대적인 예측력이 다양한 실용적 목적에 사용되기에 적합할 정도로 정교화될 수 있음을 보여준다. 두 번째로, 약간이나마 CatBoost 알고리즘의 예측력이 다른 두 알고리즘에 비해 더 높게 나타났다. 평균적 예측력 (평균 백분율 오차나 R-squared)에서 뿐 아니라, 이상치의 발생빈도 등 모형의 안정성 측면에서도 CatBoost가 더 우수한 것을 확인할 수 있다. 세 번째로, 소프트 보팅을 통한 세 알고리즘의 앙상블 모형을 구축하는 경우, 개별 알고리즘보다 더 예측력을 상승시킬 수 있음을 확인하였다. 보팅은 서로 다른 알고리즘을 통해 얻어진 예측을 결합하는 방식으로, 본 연구에서는 평균을 취하는 방식을 택했다. 이는 개별 알고리즘을 통한 예측 과정에서 발생한 고유의 소음이 다른 알고리즘의 예측과 결합되는 과정에서 약화될 수 있음을 의미하는 것이다.

본 연구의 구성은 다음과 같다. II장에서는 주택감정모형 분야에서 최근 기계학습 기법이 적용된 사례를 소개하였다. III장은 본 연구에서 사용한 세 알고리즘의 원리와 특징을 개략적으로 서술하였다. IV장은 분석자료와 기초통계를 설명하고, 그것을 통해 구축한 주택감정모형의 예측력을 비교 · 분석하였다. V장에서는 앞의 논의를 정리하며, 추후의 연구과제 등에 관해 서술하였다.

II. 이론 및 선행연구 검토

주택가격 추정 시 사용하는 가장 전통적이고 잘 알려진 모형은 헤도닉 가격 모형이다. Lancaster (1966)와 Rosen (1974)의 모형에서 시작된 헤도닉 모형은 주택의 시장 가치를 주택의 속성에 대한 회귀분석으로 묘사할 수 있는 이론적 근거를 제공한다. Lancaster의 소비자 이론에서 소비자는 상품 자체가 아니라, 속성에서 효용을 얻는 것으로 묘사된다. 그러므로 어떤 재화의 소비는 재화에 담긴 속성의 복합을 소비하는 것으로 표현될 수 있다. Rosen (1974)은 이러한 소비자 이론을 헤도닉 가격 모델로 확장했는데, 그는 재화의 가치를 Lancaster (1966)에서 묘사된 바처럼 재화에 포함된 각 속성의 가치로 환원할 수 있다고 제시했다. 각 속성이 완전 경쟁 시장에서 각각의 암묵적 가격을 갖는다는 가정하에 재화의 시장 가치는 그 재화에 포함된 속성 가치의 합으로 해석될 수 있으며, 결과적으로 이는 재화의 시장 가치가 그 재화의 속성에 대해 회귀될 수 있음을 의미한다.

그러나, 실제 회귀분석 과정에서 가정되어야 하는 함수의 형태이나 변수 목록들은 헤도닉 가격 이론 자체에 의해서는 특정될 수 없기 때문에(Malpezzi, 2002), 회귀분석의 적용 과정에서 단순화되거나 포착되지 않은 변수 또는 다중공선성 등에 의해 모형의 설명력이 왜곡될 가능성이 항상 존재한다. 특히, 헤도닉 가격 모형은 일반적으로 변수 간의 선형 관계와 독립성을 가정하는데, 다양한 이유로 실제 시장에서는 이러한 가정이 위배될 수 있음을 고려해야 한다.

이러한 헤도닉 모형의 한계를 극복하려는 방향은 다양하게 이루어져 왔는데, 크게 분류하면 다음의 두 가지를 생각할 수 있다. 첫 번째는 전통적 선형 헤도닉 모형의 가정을 완화하거나 입지효과를 통제하는 방식으로 모형 해석성은 유지하면서도 설명력을 높이는 것이다. 예를 들어, 변수관계를 Box-Cox 함수로 표현하거나(Goodman, 1978; Halvorsen and Pollakowski, 1981; Rasmussen and Zuehlke, 1990 등), k-근접 이웃 모형(Meese and Wallace, 1991; McCluskey and Anand, 1999 등)을 사용하는 방식은 단순히 선형 함수로 표현된 헤도닉 모형에 비해 변수 관계에 내재된 비선형성을 더 효과적으로 포착할 수 있는 것으로 알려져 있다. 또한, 측정 가능한 입지 변수들만으로 포착될 수 없는 입지 효과를 모형에 반영하기 위한 공간 상관 모형(대표적으로, Dubin, 1992; Anselin and Bera, 1998; Basu and Thibodeau, 1998; Anselin, 2001; Conway et al., 2010 등) 등도 이에 해당할 수 있다.

두 번째는 실제 시장의 복잡성과 비선형성을 보다 유연하게 포착할 수 있는 새로운 모형을 사용하는 것이다. 특히 최근 데이터 분석 기법, 특히 기계학습 기법의 폭발적인 성장과 함께 다양한 기계학습 모형이 주택 가격 감정 모형 분야에 적용되기 시작했는데, 이런 기법은 헤도닉 모형에 비해 직관성과 해석성이 떨어지지만, 예측력 자체는 유의미하게 상승하는 것으로 알려져 있다. 최근 데이터 과학 분야의 폭발적 성장과 더불어 다양한 기계학습 기법이 소개되었는데, 그 중 주택가격감정에 사용되는 모형은 분류 문제뿐 아니라, 회귀 문제에 있어서도 안정적으로 높은 예측력을 갖는 모형들이다. 대표적인 알고리즘들을 나열하면, 의사결정트리 (Fan, et al., 2006), 랜덤 포레스트 (Antipov and Pokryshevskaya, 2012; Hong, et al., 2020), 서포트 벡터 머신 (Gu, et al., 2011; Zurada, et al., 2011; Mu, et al., 2014) 그라디언트 부스트 (Park, B., and Bae, J. K., 2015), 인공신경망 (McCluskey and Anand, 1999; Limsombunchai, 2004; Selim, 2009) 등이 있다.

이러한 연구들은 주로 기계학습 등을 이용한 모형이 주택 감정 시스템으로 활용되는 경우, 예측력이 얼마나 상승하는지를 보여준다. 대부분의 연구에서 기계학습 기반의 모형은 전통적 모형에 비해 월등한 예측력을 보이는 것으로 조사된다. 예를 들어, Hong, et al.(2020)은 의사결정나무의 앙상블 기법인 랜덤 포레스트 알고리즘을 통해 우리나라 서울 강남구 아파트 가격에 대한 예측 모형을 추정하였는데, 이들은 기계학습 기반 모형의 예측력이 헤도닉 모형에 비해 상대적으로 뛰어날 뿐만

아니라, 평균 백분율 오차가 5~6% 정도에 불과할 정도로 정확성이 상승할 수 있음을 보여주었다.

최근 그래디언트 부스트를 기반으로 그 효율성과 예측력을 증가시키는 알고리즘들이 소개되었는데, 본 연구에서 주목하고 있는 XGBoost, LightGBM, CatBoost가 그것이다. 이러한 알고리즘들은 수행시간이 느리고 과적합 규제가 없는 그래디언트 부스트의 한계를 극복한 것으로 평가되며, 특히 분류 문제에 있어서 다른 기계 학습 기법에 비해 예측 성능이 높은 것으로 알려져 있다. 이러한 기계 학습 기법은 비교적 최근에 소개되었으며, 주택 가격 감정 문제에 이 알고리즘들이 얼마나 효율적으로 사용될 수 있는지 기존의 알고리즘 (랜덤 포레스트나 인공 신경망 등)에 비해 상대적으로 아직 덜 알려진 것으로 보인다.

III. 분석 기법의 소개

1. 그래디언트 부스트 알고리즘

본문에서 사용된 XGBoost, LightGBM, CatBoost는 모두 그래디언트 부스트 방식에 기반하고 있으므로, 이 장에서는 먼저 이에 대해 간략히 서술한다. 그래디언트 부스트는 대표적인 기계학습 알고리즘 중 하나로, 주로 분류 문제 또는 값 예측 문제 (회귀)에 사용된다. 기계학습에서 부스팅(boosting)이란 단순하고 약한 학습자(weak learner)를 오차를 최소화하는 방향으로 결합시켜 보다 정확한 학습자(strong learner)를 만드는 앙상블 형식의 알고리즘을 의미한다. 개별 예측기로는 설명력이 낮지만 가벼운 예측기를 먼저 만들고, 이를 통해 나타난 오류는 그 다음 예측기가 보완한다. 이러한 방식으로 각 예측기의 예측력을 합치면 기존보다는 정확한 모델이 만들어지고, 다시 남아 있는 예측 오류를 다음 예측기를 통해 보완하여 기존 모형에 더하는 과정을 반복하는 방식이다.

그래디언트 부스트 방식은 손실함수(loss function)을 정의하여, 경사하강법을 통해 이를 최소화하는 상태를 찾는 방식을 택한다. 이때 손실함수는 일반적으로 실제 값과 예측 값 간 차이의 제곱의 합으로 정의되나, 그래디언트 부스트로 강점은 미분 가능한 함수라면 다른 형식의 손실함수를 신축적으로 사용할 수 있다는 것이다. 경사하강법은 이러한 손실함수의 크기를 줄이는 방향으로 모형을 지속적으로 변화시켜 나가는 방식을 의미한다. 손실함수의 값을 고도가 있는 언덕에 비유한다면, 기계학습의 목표는 가능한 한 저지대로 이동하는 것일 것이다. 이때 어느 방향으로 얼마만큼 이동해야 아래로 내려가는 지는 계수 변화 또는 모형 추가에 따른 손실함수 값의 변화 방향 및 크기를 추적하면 될 것이다. 만약 손실 함수가 예측 오류의 제곱합인 경우, 손실 함수의 경사값 (미분값)은 실제에 대한 예측의 잔차항이 된다. 이런 경우, 알고리즘은 이 잔차를 목표값으로 하는

예측자를 다시 학습하여 기존 모형에 더하는 방식으로 예측력을 높이게 된다. 그래디언트 부스트는 이처럼 예측오차를 감소시키는 방향을 따라 계수를 수정하거나 모형을 누적시키는 과정을 더이상 손실함수의 값을 하락시킬 수 없을 때까지 (즉, 가장 높은 예측력에 달할 때까지) 반복하는 것이다.

이를 기술적으로 표현하면 <표 1>과 같다.

2. XGBoost 알고리즘

XGBoost는 Extreme Gradient Boost의 약자로, 기본적으로 그래디언트 부스트의 방식을

<표 1> 그래디언트 부스트 알고리즘

투입요소 : 학습 표본 $(x_i, y_i)_{i=1}^n$, 미분 가능한 손실함수 $L(y, F(x))$, 최대반복횟수 M

알고리즘 :

1. 다음을 만족시키는 상수 γ 를 찾아 $F(x)$ 의 초기값으로 부여한다.

$$F_0(x) = \arg \min \sum_{i=1}^n L(y_i, \gamma).$$

2. m 에 대하여 1부터 M 까지 아래를 반복한다.

- 2-1. 다음과 같은 유사잔차(pseudo-residuals)를 계산한다.

$$r_{i,m} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x) = F_{m-1}(x)} \quad \text{for } i = 1, \dots, n.$$

- 2-2. 약한 학습자 $h_m(x)$ 를 위의 유사잔차에 대하여 학습한다.

- 2-3. 아래의 극대화 문제의 해를 통해 승수 γ_m 로 계산한다.

$$\gamma_m = \arg \min \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)).$$

- 2-4. 모형 $F(x)$ 를 아래와 같이 업데이트한다.

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x).$$

3. $F_M(x)$ 를 최종 반환한다.

따르지만 그래디언트 부스트의 단점 중 하나인 느린 수행 시간과 과적합 문제를 해결한 알고리즘이다. 그래디언트 부스트의 경우, 손실함수를 감소시키는 최적의 함수를 찾기 위해 가능한 경우의 수를 모두 탐색한다. 이때 만약 고려되는 변수의 수가 많은 경우 연산 효율성은 급격히 떨어질 수 있다. 특히, 각각의 범주형 변수를 알고리즘을 통해 연산하기 위해 각 범주형 변수에 포함된 범주값을 더미변수화하는 경우가 많으므로, 결과적으로 많은 범주를 포함한 범주형 변수가 소수만 포함되더라도 극단적인 비효율성으로 인한 연산력의 저하에 노출될 수 있다. XGBoost는 변수의 분포를 고려하여 이런 비효율적 탐색 과정을 간략화하여 결과적으로 모형의 연산 효율성과 추정력을 상승시키는 알고리즘이다. 일반적인 그래디언트 부스트의 경우, 과적합에 대응하는 기능이 별도로 존재하지 않지만, XGBoost에서는 과적합에 대한 규제를 통해 보다 안정적인 예측이 가능한 것으로 알려져 있다. XGBoost는 다른 기계 학습자에 비해 예측 성능이 뛰어나며, 병렬 CPU를 통한 학습이 가능해 그래디언트 부스트에 비해 빠른 수행시간을 갖는 것으로도 평가받는다.

3. LightGBM 알고리즘

그래디언트 부스트 계열의 알고리즘은 반복 누적하는 약한 학습자로 의사결정트리를 사용한다. 이때 LightGBM은 일반적인 트리 분할 방법과 다르게 리프 중심 트리 분할 방식을 사용한다. 보통의 의사결정트리 기반 알고리즘은 개별 트리가 과적합되지 않도록 균형 트리 분할 방식을 사용하여 트리의 깊이를 줄인다. 즉, 최대한 트리의 균형을 맞추면서 개별 트리의 깊이를 최소화하는 것이다. 하지만 LightGBM은 리프 중심 트리 분할 방식을 사용하는데, 이는 개별 의사결정트리가 비대칭적이 되더라도 (즉, 과적합 가능성이 있더라도) 최대한의 효율적 학습을 수행하도록 유도하는 것이다. LightGBM 알고리즘은 이러한 약한 학습자를 쌓아갈수록 결국은 최종적으로 균형 트리 분할 방식에 비해 예측 오류를 줄여줄 수 있다는 것에 착안하고 있다. 이러한 방식의 대표적인 강점은 예측 성능을 유지하거나, 오히려 강화시키면서도 수행시간이 더 빠르다는 것이다. 일반적으로 XGBoost가 그래디언트 부스트에 비해 빠르다고 알려져 있지만, LightGBM은 한층 더 가벼운 연산을 수행할 수 있다. LightGBM은 XGBoost와 마찬가지로 병렬 연산 기능을 제공함으로써 큰 데이터에 대한 효율적인 연산을 가능하게 한다.

4. CatBoost 알고리즘

그래디언트 부스트를 비롯한 여타 많은 기계 학습 알고리즘(랜덤 포레스트 등)에서는 범주형

변수를 처리하기 위해 정수화하거나 더미변수화한다. 이러한 방식은 약점을 가지고 있는데, 먼저 범주형 변수를 정수화하는 경우 실제로는 카디널리티(cardinality)가 없는 변수를 카디널리티가 있는 변수로 변환하는 것으로 의도치 않은 정보의 왜곡을 낳을 수 있다. 더미변수화하는 방식은 범주형 변수의 범주 개수 만큼의 변수를 추가하는 것이 되므로, 범주의 개수가 많으면 많을수록 모형의 복잡성을 높이고, 연산 효율성을 훼손시키게 된다. 특히 주택 감정 모형에서와 같이 행정동 또는 각 시점 등과 같이 다수의 범주값을 갖는 범주형 변수를 포함하는 연산 과정에서는 알고리즘의 복잡성이 너무 높아져 더미변수화 된 변수들(즉, 범주값의 구분자들) 중 상당수가 연산에 반영되지 못할 수 있다.

CatBoost 알고리즘은 정보 획득(information gain)량이 동일한 여러 속성을 하나의 속성으로 묶어 버림으로써, 결과적으로 범주형 변수들을 효과적으로 알고리즘 연산에서 반영할 수 있게 설계되었다. 즉, 범주형 변수의 성질(즉, 다른 변수값 간에 대소비교는 불가능하지만, 하나의 변수로 묶이는 성질)을 유지하면서 모형을 학습할 수 있다는 것을 의미한다. 결과적으로 CatBoost는 변수 숫자의 확장으로 인한 정보의 탈락을 최소화할 수 있으며, 그로 인한 예측력과 연산속도의 상승을 기대할 수 있다.

IV. 분석 데이터

본 연구의 분석 대상은 2009년 7월부터 2019년 12월까지 서울시에서 일어난 아파트 거래로, 총 표본 개수는 620,617개이다. 아파트는 우리나라의 대표적인 주거 방식으로 많은 표본을 수집할 수 있을 뿐 아니라, 단독주택 등에 비해 비관측 요소가 상대적으로 적고 데이터로 집계되는 주택 속성이 균일하기 때문에, 관찰 가능한 주택 속성을 통해 가격을 예측해야 하는 본문의 분석의 특성상 더 적합한 분석 대상일 수 있다(Hong et al., 2020). 본 연구에서 사용한 표본 크기가 일반적인 헤도닉 모형 분석의 사례에 비해 상당히 크기 때문에, 복잡성이 높은 기계학습 알고리즘을 학습하기에도 충분할 뿐더러 연구 결과의 일반성을 주장하는 데에도 강점이 있을 것으로 보인다.

아래 <표 2>는 수집된 자료의 정의를 소개하고 있으며 <표 3>은 그에 대한 기초통계를 제공한다. 수집 자료에 대한 개략은 다음과 같다. 먼저, 아파트 매매가격은 국토교통부에서 공개하는 아파트 매매 실거래가를 활용하였다. 주택 속성은 크게 구조적 속성과 입지적 속성이 있는데, 본 연구에서 포함하고 있는 구조적 속성으로는 아파트의 크기(전용 면적), 방 개수, 화장실 개수, 층 수, 난방 방식(개별난방, 중앙난방, 지역난방), 복도 형태(복도식, 계단식, 혼합식), 경과 년 수, 아파트 단지 내 세대수, 동수, 가구 당 평균 주차 대수, 단지 내 최고층의 높이, 최저층의 높이, 건폐율, 용적률이 있다. 입지적 속성으로는 가장 가까운 지하철역로부터의 거리, 주요 편의시설(공원, 박물관)

〈표 2〉 분석 자료의 정의와 단위

분류	변수	설명	단위
종속변수	가격	거래시점의 매매가격	원
구조적 속성	크기	전용면적	m ²
	방 수	방의 개수	개
	화장실 수	화장실의 개수	개
	층	위치한 층의 높이	층
	난방	개별/중앙/지역난방의 여부	범주형
	경과년수	준공년도와 거래년도의 차이	년
	세대수	아파트 단지에 존재하는 세대의 수	세대수
	동수	아파트 단지에 존재하는 건물(동)의 수	동수
	용적률	(건물 연면적 / 대지 면적)×100	%
	건폐율	(건물 면적 / 대지 면적)×100	%
	주차대수	아파트 단지 주차가능 대수 / 세대 수	대수
	최고층	아파트 단지 내 건물 중 최고층의 높이	층
	최저층	아파트 단지 내 건물 중 최저층의 높이	층
입지적 속성	행정구역	행정동의 구분	범주형/더미 변수
	위치	위도, 경도	좌표값
	지하철로부터의 거리	가장 가까운 지하철역로부터의 거리	미터
	공원으로부터의 거리	가장 가까운 공원으로부터의 거리	미터
	초등학교로부터의 거리	가장 가까운 초등학교로부터의 거리	미터
	중학교로부터의 거리	가장 가까운 중학교로부터의 거리	미터
	고등학교로부터의 거리	가장 가까운 고등학교로부터의 거리	미터
	대학교로부터의 거리	가장 가까운 대학교로부터의 거리	미터
	박물관으로부터의 거리	가장 가까운 박물관으로부터의 거리	미터
	주민센터로부터의 거리	가장 가까운 주민센터로부터의 거리	미터

으로부터의 거리, 교육기관(초등, 중등, 고등, 대)으로부터의 거리, 행정시설(주민센터로부터의 거리)이 있다. 이때 거리는 GIS 자료를 통해 유클리드 거리(m)를 계산하였다. 또한, 입지적 속성값

〈표 3〉 기초통계

변수	평균	표준편차	최소값	최대값
가격	54,043.88	36,949.47	700	700,000
크기	79.49762	28.52365	11.95	325.39
방개수	2.969445	0.667722	1	8
화장실수	1.659708	0.492461	1	5
세대수	1,032.104	1,153.402	5	9,510
동수	11.88267	14.04946	1	122
주차대수	1.142896	0.479476	0.02	11.95
용적률	287.9124	126.9176	2	1,477
건폐율	24.90165	27.63889	2	2,457
최고층수	19.18334	6.811636	4	69
최저층수	12.16432	5.424588	1	54
층수	9.445466	6.190082	-4	68
경과년수	13.78384	7.813038	0	49
지하철로부터의 거리	770.5905	618.0864	2.645571	5,583.588
공원으로부터의 거리	1,036.23	526.1945	55.73893	3,268.238
초등학교로부터의 거리	337.1795	169.516	10.59662	1,810.038
중학교로부터의 거리	471.2093	252.5819	2.587437	2,130.155
고등학교로부터의 거리	579.8585	333.7149	24.61597	2,837.36
대학교로부터의 거리	1,867.998	1,190.588	50.28124	7,111.538
박물관으로부터의 거리	1,829.797	1,068.645	35.34001	6,839.077
주민센터로부터의 거리	1,938.611	992.8313	16.82066	6,521.695

자체는 아니지만 입지적 차이를 구분할 수 있는 변수로써 행정구역(동)과 위치값(위도, 경도)도 수집되었다.

V. 정량분석결과

1. 예측력 비교 분석

1) 비교대상: 헤도닉 모형의 설정

본 연구는 앞서 소개한 XGBoost, LightGBM, CatBoost이 얼마나 실제 주택 시장의 복잡성과 비선형성을 효율적으로 포착할 수 있는지 알아보기 위해, 이러한 기계학습 모형들과 선형 헤도닉 가격 모형의 예측력을 비교하였다. 본문에서 사용한 헤도닉 모형은 다음과 같다.

$$\ln P_i = \beta_0 + X_{s,i}\beta_s + X_{l,i}\beta_l + T\beta_t + \epsilon_i \quad \langle \text{식 1} \rangle$$

위에서 $\ln P_i$ 는 주택매매가격의 자연로그이며(하첨자 i 는 i 번째 주택을 의미한다), β_0 는 상수항이다. $X_{s,i}$ 와 $X_{l,i}$ 는 각각 i 번째 주택의 구조적 속성들과 입지적 속성들을 담은 벡터를 의미한다. β_s 와 β_l 는 각각 구조적 속성들과 입지적 속성들의 계수값을 담은 벡터이다. 본문의 모형에서 구조적 속성으로는 크기, 방 개수, 화장실 개수, 난방 방식, 층수, 경과년수, 세대 당 평균 주차가능 대수, 용적률, 건폐율, 단지 내 세대수, 단지 내 동 수, 단지 내 최고층 높이, 단지 내 최저층 높이가 포함되었다. 아파트가 분석대상인 경우, 개별 주택 자체의 특성 이상으로 해당 아파트 단지의 특성이 유의미한 영향을 미칠 것으로 보고 다양한 변수를 포함하였다. 입지적 속성으로는 가까운 지하철역으로부터의 거리, 주요 편의시설(공원, 박물관)로부터의 거리, 교육기관(초등, 중등, 고등, 대)으로부터의 거리, 행정시설(주민센터로부터의 거리)에 더해 행정동을 구분하는 더미변수가 포함되었다. 거리는 GIS 자료를 통해 유클리드 거리를 계산하였다. 행정동 더미가 포함된 이유는 주요 입지로부터의 거리뿐 아니라, 다양한 요소(예를 들어, 입지 주변 거주자들의 평균소득 등)들도 입지 가치에 큰 영향을 미칠 수 있기 때문이다. 이러한 입지별 차이의 상당 부분은 행정동의 구분과 연관되어 있기 때문에, 이를 포착하는 더미 변수를 모형에서 입지적 속성에 추가로 고려하였다. 즉, 행정동 더미의 추정 계수값에는 이처럼 직접 관찰할 수 없는 동 단위의 효과가 집약되어있는 것으로 이해할 수 있을 것이다. T 는 거래 시점에 대한 더미변수이며, β_t 는 그에 대한 계수이다. 거래 시점에 따라 이자율, 거시경제적 환경 등 다양한 요소가 복합적으로 가격에 영향을 미치므로, 다양한 개별 요소를 직접 통제하는 것보다는 이를 시점에 대한 더미변수의 형태로 한꺼번에 포착하는 것이 효과적일 것이다. 마지막으로 ϵ_i 는 오차항으로 평균이 0인 정규분포를 따른다고 가정하였다.

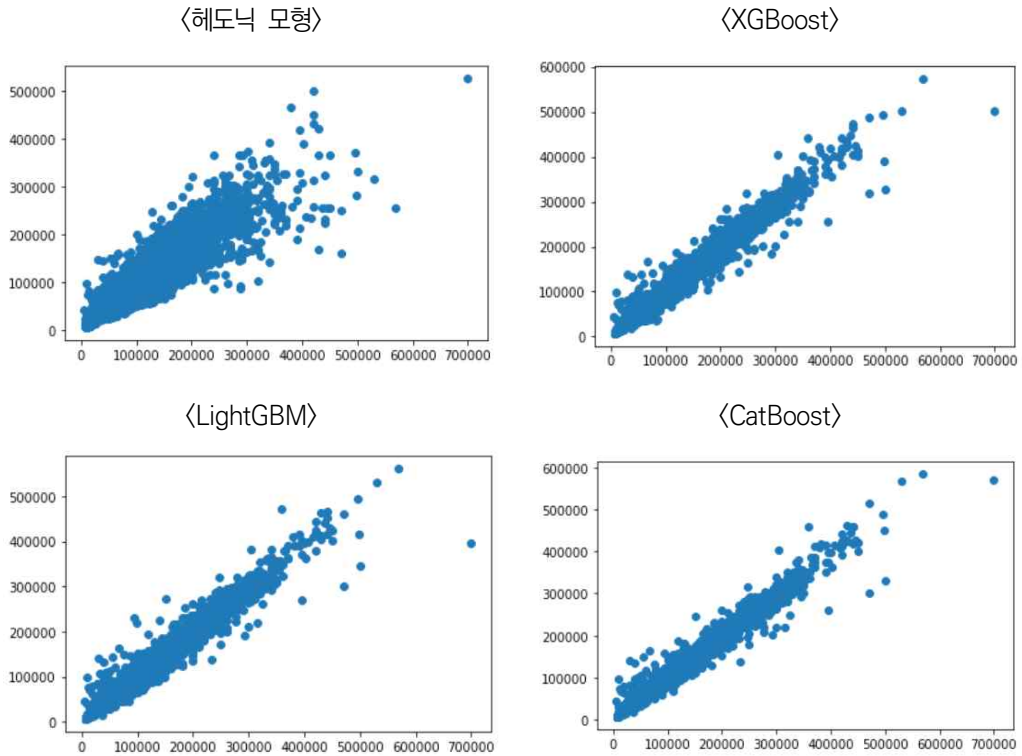
본문에서는 헤도닉 모형의 독립변수로 상당히 다양한 변수를 포함하고 있는데, 한 가지 고려해야 할 점은 본문의 맥락에서 헤도닉 모형을 사용하는 이유가 종속변수 값의 예측 자체에 있다는 것이다. 일반적으로 변수를 많이 포함하게 되는 경우, 다중공선성 등의 문제로 인해 개별 변수의 영향력이 왜곡되는 경우가 나타날 수 있다. 그러나, 중요한 점은 종속변수의 예측에 대해서는 설사 다중공선성 문제가 있더라도 예측력이 저하되지 않는다는 것이다 (Greene, W. H., 2003). 이는 개별 변수의 계수가 변수간 공선성의 존재로 인해 왜곡되더라도, 그러한 관계를 포함한 최종결과가 종속변수에 대한 예측값으로 반환되기 때문이다. 이런 경우 모형의 예측력을 상승시키기 위해서는 가능한 다양한 변수를 포함하도록 설계할 수 있다. 다시 말해, 본문에서는 다소간 공선성을 일으킬 수 있는 변수들이 포함되어 있더라도 회귀분석 기반 모형의 예측력을 극대화하는 데에 집중하였다는 것이다. 이처럼 예측력이 극대화 된 헤도닉 모형을 기계학습 모형과 비교하는 것이 모형 간 예측력의 차이가 (변수 선정의 차이가 아닌) 각 모형의 비선형성과 복잡성의 포착력에 있다는 것을 보여줄 수 있을 것이다.

2) 모형 간 예측력 비교

이 장에서는 헤도닉 가격 모형, XGBoost, LightGBM, CatBoost 알고리즘을 통해 주택가격감정 모형을 측정하고, 그 예측력을 비교하였다. 모형 간 예측력을 객관적으로 비교하기 위해 수집된 전체 표본 620,617개는 학습 표본과 평가 표본으로 분할되었다. 이때 학습 표본은 전체의 90%인 558,555개, 평가 표본은 10%인 62,062개로 임의 할당되었다. 본문은 먼저 학습 표본을 통해 각 모형을 추정하였다. 그리고 이렇게 측정된 모형들에 대해 평가 표본들의 속성값만을 관찰하고, 매매가격을 예측하도록 하였다.²⁾

〈그림 1〉은 평가 표본의 속성값을 통해 각각의 모형이 예측한 주택 가치와 실제 시장에서 거래된 주택 가격의 산점도를 나타낸 것이다. 가로축은 실제 주택 매매가격을 의미하며, 그에 대한 예측치는 세로축에 나타나 있다. 만약 완벽하게 추정값과 실제값이 일치한다면 모든 점이 45도 선에 위치하게 될 것이며, 상관계수는 1이 될 것이다. 반면 추정치와 실제값이 아무런 관계가 없다면, 산점도에는 일정한 규칙이 나타나지 않을 것이며, 상관계수 역시 0에 가까울 것이다. 그림에 나타난 모든 모형에서 뚜렷한 정의 관계가 관찰된다. 한 가지 확인할 수 있는 것은 헤도닉 모형의 산점도가 다른 기계 학습 알고리즘 기반의 모형의 경우에 비해 분산이 크게 나타난다는 것이다. 특히 이러한 분산의 정도는 저가 주택에서보다 고가 주택에서 더 심하게 나타난다. 반면 XGBoost, LightGBM, CatBoost

2) 본 연구에서는 1. 전체 표본의 수가 충분히 크고, 2. 각 기계 학습자들이 앙상블 기법에 기반하기 때문에 이처럼 임의분할을 하더라도 그로부터 얻은 분석 결과의 질적인 해석은 거의 동일하다.



주: 가로축은 실제 주택매매가격, 세로축은 그에 대한 예측값.

〈그림 1〉 주택감정=모형의 산점도 비교

모형의 산점도에서는 이러한 현상이 나타나지 않는 것을 확인할 수 있다.

본문을 각 모형의 정확성을 다음의 두 가지 기준을 통해 평가하였다. 첫째는 평균 절대 백분율 오차(Mean Absolute Percentage Error, 이하 MAPE)이다. 이는 개별 추정치가 실제값에서 몇 퍼센트만큼 평균적으로 이탈하고 있는지를 보여주는 가장 직접적인 평가 기준으로, 다음과 같이 정의된다.

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{\hat{P}_i - P_i}{P_i} \right| \quad \langle \text{식 2} \rangle$$

이때 P_i 와 $\hat{P}_i$ 는 각각 i 번째 실제 주택매매가격과 예측값을 의미한다.

두 번째는 R-squared로, 이는 모형을 통해 예측할 수 있는 표본 간 차이(분산)과 실제 차이의 비율을 의미한다. R-squared는 아래와 같이 정의된다.

$$R^2 = 1 - \frac{\sum_{i=1}^n (P_i - \hat{P}_i)^2}{\sum_{i=1}^n (P_i - \bar{P})^2}, \quad \langle \text{식 3} \rangle$$

이때 $\bar{P}$ 는 평가 표본의 평균을 의미한다.

아래 <표 4>는 각 모형들의 MAPE, <표 5>는 R-squared를 비교하여 보여준다. 먼저, MAPE와 R-squared 모두에서 헤도닉 모형과 기계 학습 기반 알고리즘의 예측력 차이가 상당하다는 것을 확인할 수 있다. 헤도닉 모형의 경우 평균 백분율 오차는 11.5% 가량으로, 평가 표본의 개수가 62,206개에 달한다는 것을 고려할 때 낮은 수치라고 평가할 수는 없을 것이다. 그러나 기계학습 기반의 예측력은 모두 4% 이내로, 데이터 자체의 관측 오류나 변수로 관찰 불가능한 특이 표본(예를 들어, 판매자의 사정으로 실제 가치에 비해 크게 저평가된 거래 등)으로 인한 불가피한 설명력 손상이 있음을 감안할 때, 본문에서 사용된 알고리즘들은 시장 자체의 비선형성이나 입지 가치로 인한 주택 가치 결정구조의 복잡성 등 다양한 요소로 인한 가격 차이를 상당 부분 포착할 수 있다는 것을 의미한다. 또한, 기존의 다른 기계 학습(랜덤 포레스트나 서포트 벡터 머신) 또는 인공신경망 기반

<표 4> Mean average percentage error (MAPE)의 비교

	MAPE
헤도닉 모형	11.508
XGBoost	3.728
LightGBM	3.854
CatBoost	3.609

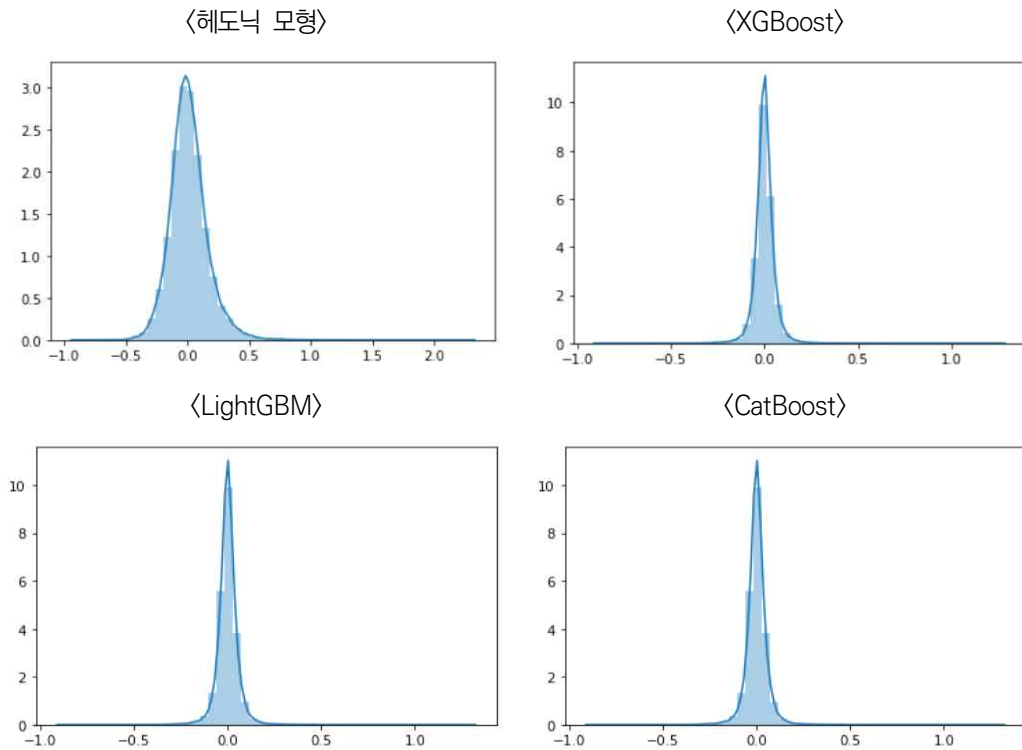
<표 5> R-squared의 비교

	R-squared
헤도닉 모형	0.8962
XGBoost	0.9856
LightGBM	0.9836
CatBoost	0.9878

주택 감정모형의 평균 백분율 오차가 낮은 경우에도 5% 내외였다는 것을 생각해 보면, 그래디언트 부스트에 기반한 세 알고리즘의 활용성이 상당히 높다는 것을 알 수 있다. 본문에서 추정된 모형 중에서도 CatBoost의 예측력이 다소 더 높게 나타나는데, 이 모형은 평균 오차에서 뿐만 아니라, 다음 장에서 설명될 이상치의 발생빈도에서도 더 우수한 성능을 나타내는 것으로 확인된다.

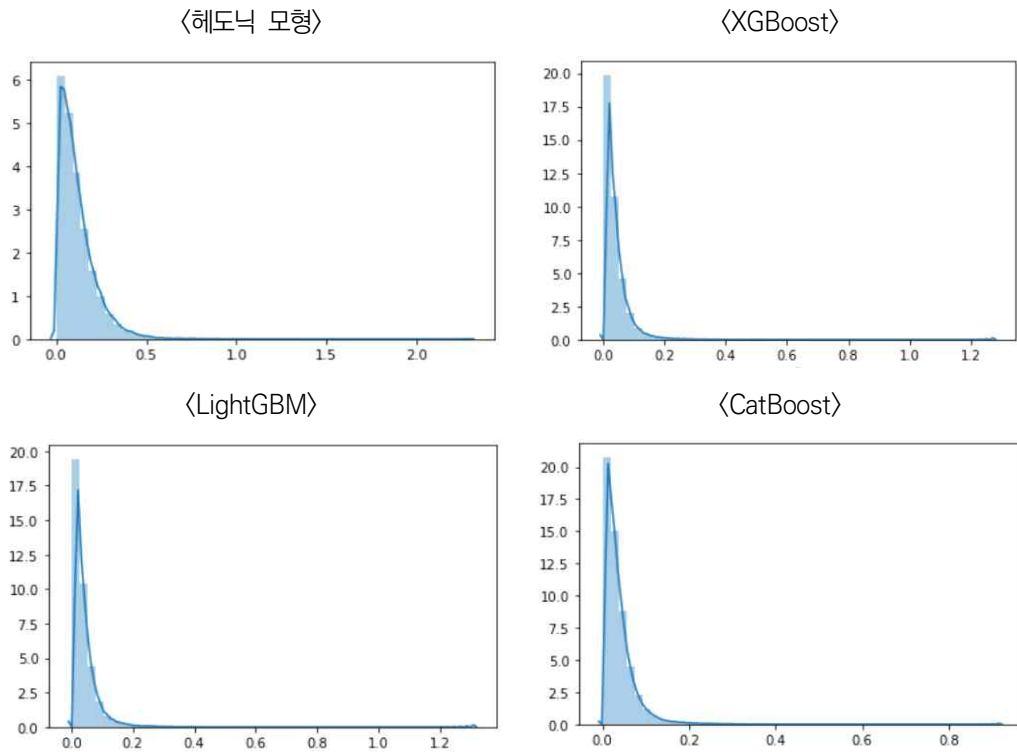
3) 오차의 분포 및 성질

이 장에서 우리는 각 모형들을 통한 예측의 분포 및 이상치의 발생빈도를 비교분석하였다. 먼저 예측 오차(백분율 기준)의 분포를 통해 각 모형을 통한 예측의 특성을 간단히 비교할 수 있다. 다음 <그림 2>와 <그림 3>은 각 모형의 예측 오차와 그 절대값의 히스토그램을, <표 6>은 표준편차와 분위값을 보여주고 있다. 그래프 형태와 표준편차 값의 차이에서 확인할 수 있듯이 헤도닉 모형에



주: 가로축은 실제값과 예측값 간의 % 격차 ($= \frac{P_i - \hat{P}_i}{\hat{P}_i}$), 세로축은 그에 대한 빈도.

<그림 2> 예측 오차의 히스토그램



주: 가로축은 실제값과 예측값 간 % 격차의 절대값 ($= \left| \frac{P_i - \hat{P}_i}{\hat{P}_i} \right|$), 세로축은 그에 대한 빈도.

〈그림 3〉 예측 오차(절대값)의 히스토그램

〈표 6〉 예측 오차의 분포

	헤도닉 모형	XGBoost	LightGBM	CatBoost
표준편차	0.156275	0.056608	0.05787	0.052316
1사분위 값	-0.0837	-0.02357	-0.02508	-0.02443
2사분위 값 (중위값)	-0.00032	0.001036	0.000459	0.000626
3사분위 값	0.092038	0.026486	0.026019	0.026064
표본 수	62,062			

비해 세 기계학습 알고리즘을 통한 예측이 더 조밀하게 형성되는 것을 확인할 수 있다. 표준편차 값은 헤도닉 모형, XGBoost, LightGBM, CatBoost 순으로 각각 0.156, 0.056, 0.057, 0.052로,

앞에서 확인한 R-squared 및 MAPE의 순과 동일한 것을 발견할 수 있다. <표 6>은 1,2,3 사분위 값도 제공하고 있는데, 이를 통해 전체의 상위 50%가 1사분위(25%) 값과 3사분위(75%) 값 사이에 존재한다는 것을 확인할 수 있다. 즉, 예측력이 가장 높은 CatBoost의 경우, 절반 정도의 예측이 -2.4%와 2.6%의 예측 오차 범위에 들어간다는 것이다.

다음으로, 각 모형들을 통한 예측에서 나타나는 오류 중 이상치의 발생 빈도에 대해 알아본다. 실제 값과 예측값 사이의 차이는 모형의 불완전성이나 관측 불가능한 속성들(매매 또는 구매자의 특성 등) 또는 자료 자체의 결함으로 인해서 불가피하게 발생할 수밖에 없으나, 종종 모형을 통한 예측에서는 추정값이 실제값에 비해 지나치게 크거나 작은 경우가 발견될 수 있다. 평균적 예측 오차 자체는 낮더라도, 이러한 이상치가 빈번하게 발생하는 경우, 모형의 신뢰성이 크게 떨어지기 때문에 가능한 한 이를 낮추는 것이 중요하다.

예측 이상치의 발생은 예측 모형의 복잡성 및 비선형성과 관련이 있을 수 있다. 선형 회귀분석 기반의 예측 모형의 경우, 종속변수에 대한 추정치는 속성값의 선형투사(linear projection)에 의해 이루어진다. 그러므로 계수가 과대 평가되거나 과소 평가된 속성의 속성값이 특정 표본에서 매우 큰 경우에만 큰 편차가 발생할 것이다. 반면, 그래디언트 부스트 기반의 세 알고리즘의 경우, 이상치의 발생 메커니즘을 정형화하기 어렵다. 이는 모형의 과적합 문제와도 연관되어 있을 수 있다. 세 알고리즘 모두 단순하고 약한 예측 모형을 오차를 수정하는 방향으로 반복 누적시키는 방식을 택하는데, 이때 약한 예측 모형으로는 의사결정트리나 사용된다. 이때 만약 의사결정트리들의 예측방향이 학습표본 자체에 포함된 특이성에 과도하게 적응하는 경우 평가표본의 예측에 대해서는 이상치를 발생시킬 가능성이 있다. 하지만 만약 기계학습 알고리즘 기반 모형들이 실제 시장의 비선형적 구조를 더 효과적으로 포착할 수 있다면, 모형에 내재된 비선형성이 이상치의 발생으로 연동되지는 않을 것이다.

다음의 <표 7>은 이러한 이상치의 발생빈도를 비교하고 있다. 본문 맥락에서 예측 이상치는 추정값이 실제값에 비해 지나치게 크거나 작은 경우로, 여기서는 다음의 세 가지를 기준으로 정의하였다; 1. 실제값이 예측값의 50%이상 떨어진 경우, 2. 실제값이 예측값의 75%이상 떨어진 경우, 3. 실제값이 예측값의 100%이상 떨어진 경우. 표의 결과를 요약하면 다음과 같다. 첫째, 기계 학습 알고리즘의 이상치 발생 빈도가 헤도닉 모형에 비해 현저히 낮다. 헤도닉 모형의 경우, 620,617개의 평가 표본 중 예측값이 실제의 50% 이상 벗어난 사례가 545건으로 전체의 약 0.87% 정도가 이에 해당한다. 반면, XGBoost, LightGBM, CatBoost의 경우 각각 58, 43, 31건으로 선형회귀 기반 모형에 비해 10~15분의 1 이하로 감소한 것을 확인할 수 있다. 모형뿐 아니라 데이터 자체의 결함도 존재할 가능성이 있음을 감안하면, 기계 학습 알고리즘의 복잡성으로 인한 불안정성

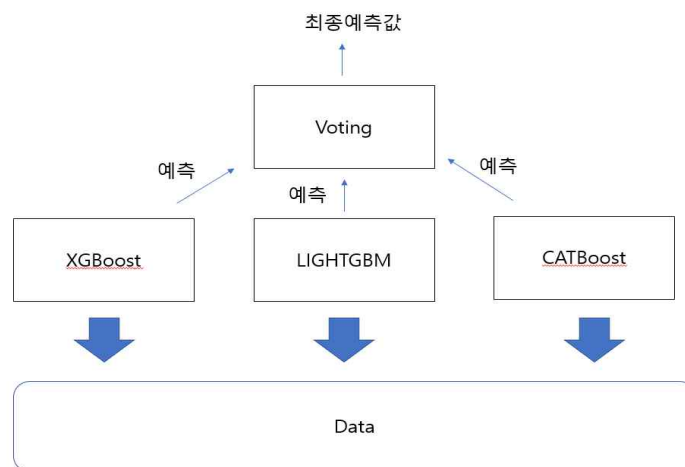
〈표 7〉 이상치 발생 비율

	오차 50% 이상	오차 75% 이상	오차 100% 이상
헤도닉 모형	0.00878	0.00209	0.00058
XGBoost	0.00085	0.00021	0.00005
LightGBM	0.00077	0.00016	0.00002
CatBoost	0.0005	0.00011	0

문제는 크지 않음을 알 수 있다. 둘째, 기계 학습 알고리즘 중에서도 CatBoost 알고리즘 기반 모형의 예측 이상치 발생율이 다른 두 알고리즘에 비해 더 낮은 것으로 확인된다. 특히 CatBoost에서는 예측치가 실제의 100% 이상 벗어나는 큰폭의 예측 오류가 620,617건 중 단 한 건도 발생하지 않았다. 또한, 평균적 예측력은 XGBoost가 LightGBM보다 다소 높지만(〈표 4, 5〉), 이상치의 발생빈도는 LightGBM가 더 낮은 것으로 나타났다.

2. 기계학습 예측기의 결합

이 장에서는 위에서 제시된 기계 학습 알고리즘을 얻은 모형의 결합을 통해 더 높은 예측력을 가진 모형을 구현할 수 있는지에 대해 간략히 알아보았다. 모형의 결합은 보팅(voting)이라고도 하는데, 말 그대로 서로 다른 알고리즘을 통해 얻은 예측값들을 투표(다수결) 또는 평균하여 개별 알고리즘에 포함될 수 있는 오류의 가능성을 중화하는 앙상블 기법이다(〈그림 4〉). 이때 투표의 경우 하드보팅,

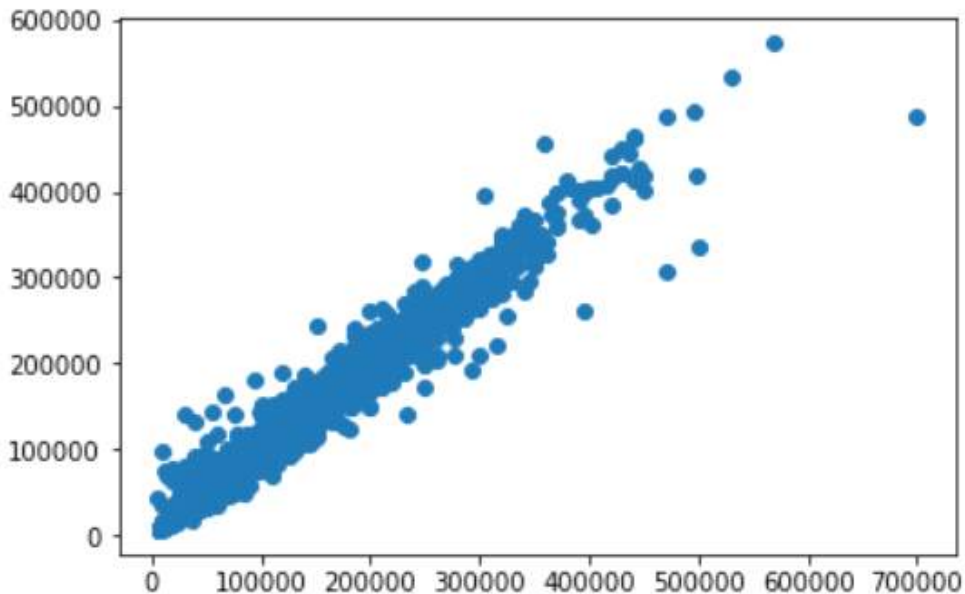


〈그림 4〉 서로 다른 알고리즘의 결합

평균의 경우 소프트 보팅이라고 부른다. 하드 보팅의 경우, 최종 예측값이 범주변수의 형태인 분류 문제에 적용될 수 있기 때문에, 본 연구에서는 위의 세 알고리즘(XGBoost, LightGBM, CatBoost)을 통한 소프트 보팅의 예측 성능을 측정해 보았다. 본문에서는 가장 직관적인 형태의 소프트 보팅인 세 알고리즘의 예측값 평균을 사용하였다.

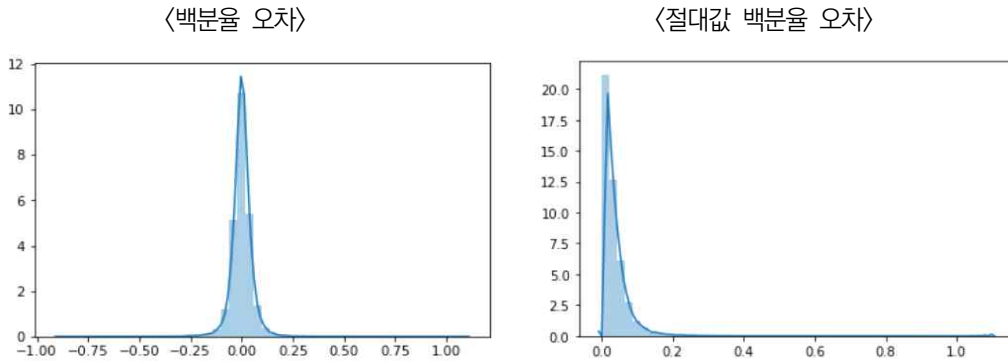
〈그림 5〉는 이러한 앙상블 모형의 예측치와 실제 값 간 산점도이며, 〈그림 6〉은 백분율 오차의 히스토그램이다. 앙상블 모형의 산점도와 히스토그램은 본질적으로 세 알고리즘의 산점도와 히스토그램과 유사할 수밖에 없을 것이다. 그러나, 세 알고리즘의 산점도와 앙상블 모형의 산점도를 자세히 비교해 보면, 각각의 세 알고리즘의 산점도에서 상대적으로 예측오차가 벌어졌던 구간(45도 선에서 분산된 구간)들이 약간이나마 조밀해진 것을 확인할 수 있다.

〈표 8〉은 예측 성능을 평가했던 기준(MAPE, R-squared)을 앙상블 모형에 적용한 것으로, 이러한 예측 성능의 상승 효과를 보다 직접적으로 드러내고 있다. 흥미로운 것은 R-squared와 MAPE 기준 모두에서 약간이나마 앞서 소개된 알고리즘들보다 예측 성능이 상승하는 것으로 나타난다는 것이다. 〈표 8〉과 〈표 4〉를 비교해 보면, XGBoost, LightGBM, CatBoost의 MAPE는 각각 3.72, 3.85, 3.60인 데에 반해, 앙상블 모형의 경우 동일한 표본에 대한 MAPE가 3.54로 더 낮아지는 것으로 확인된다. 이는 각각의 알고리즘 고유의 특성으로 인해 생겼던 소음들이 다른 알고리즘에는 나타나지



주: 가로축은 실제 주택매매가격, 세로축은 그에 대한 예측값.

〈그림 5〉 결합된 모형을 통한 예측의 산점도



주1: 첫 번째 그래프의 가로축은 실제값과 예측값 간 % 격차 ($= \frac{P_i - \hat{P}_i}{\hat{P}_i}$), 세로축은 그에 대한 빈도.

주2: 첫 번째 그래프의 가로축은 실제값과 예측값 간 % 격차의 절대값 ($= \left| \frac{P_i - \hat{P}_i}{\hat{P}_i} \right|$), 세로축은 그에 대한 빈도.

〈그림 6〉 결합 모형의 예측 오차 히스토그램

〈표 8〉 결합된 모형의 예측 성능

R-squared	MAPE
0.9871	3.549

않기 때문에, 이를 평균하는 과정에서 예측에 포함된 소음의 크기가 작아진 효과로 해석할 수 있다. 즉, 세 알고리즘 모두 기존의 모형에 비해 뚜렷한 성능 향상이 나타나지만, 앙상블 기법을 통해 이를 결합하는 경우 예측 성능을 더 상승시킬 수 있다는 것이다. 이러한 예측 성능의 향상은 〈표 9〉와 〈표 6〉의 비교에서도 발견할 수 있다.

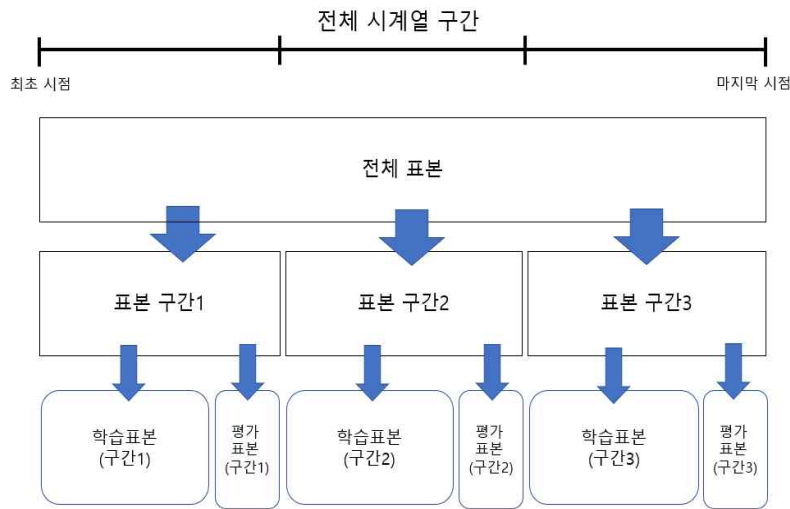
〈표 9〉 결합된 모형의 백분율 예측 오차 분포

표준편차	0.052749
1사분위 값	-0.02319
2사분위 값(중위값)	0.000745
3사분위 값	0.024807
표본 수	62,062

3. 시계열 순서에 의한 표본 분할 검증

앞의 기본 정량분석에서는 전체 표본을 9:1의 비율로 임의분할하여 학습 데이터와 평가 데이터를 구분하였다. 이는 본문의 맥락에서 감정 모형의 활용 가능성이 현시점만의 자산가치 감정에 국한되지 않기 때문에, 가능한 한 많은 표본을 학습함으로써 예측력을 극대화하려는 시도로 볼 수 있다. 그러나 이처럼 시계열 순서를 무시하는 경우, 현 시점에서 유사한 다른 자산의 가치들이 아직 관찰되지 않은 자산에 대한 예측력은 제대로 평가될 수 없다.³⁾ 따라서 이 장에서는 시계열 순서를 고려하여 학습 표본과 평가 표본을 분할하고, 각 모형의 예측력을 비교한 결과를 수록하였다.

분할 방식은 아래와 같다. 먼저, 분석 대상의 전체 시계열 구간(본 연구에서는 월 단위까지 관측)을 균등하게 분할한다.⁴⁾ 다음으로는 분할된 각각의 구간 표본을 시계열 순서로 나열하고, 순서가 빠른 90%는 학습 표본으로, 나머지 10%는 평가 표본으로 사용한다(<그림 7>). 이처럼 시계열 구간 분할을 이용하는 경우, 모든 학습 표본은 평가 표본에 비해 과거값이므로 결과적으로 과거 관측값을 토대로



<그림 7> 시계열 구간에 기반한 표본 분할

- 3) 시계열 순서를 무시하고 학습하는 경우(아직 관찰되지 않은), 해당 시점 또는 그 이후의 변화를 학습할 수 있기 때문이다.
- 4) 그런데, 주택 거래 빈도는 매년 일정하지 않으므로 이 경우 시계열을 균등분할하는 경우, 각 구간에 포함된 표본의 개수는 달라질 수 있다. 반면, 표본 개수를 동일하게 분할하는 경우, 시계열 구간의 길이가 균등하지 않을 것이다. 본 연구에서는 맥락상 분할 된 시계열의 길이를 균일하게 맞추는 쪽에 초점을 맞추었다.

미래 자산가치를 예측하는 모형이 된다.

한 가지 고려해야 할 점은 표본의 분할이 곧 표본의 크기가 줄어든다는 것을 의미하므로, 자연스럽게 모형의 예측력 저하를 피할 수 없다는 것이다. 즉, 앞 장에서 제시된 예측력과의 차이가 시계열 순서를 고려함으로써 발생할 뿐 아니라, 표본 수 감소에 의해서도 나타날 수 있다는 것이다. 이로 인한 영향을 줄이기 위해서는 분할된 각 구간에 충분한 수의 표본이 포함될 수 있을 정도로 분할이 이루어져야 할 것이다. 반면, 구간 분할 없이 시계열 순서로 전체 표본을 학습 표본과 평가 표본으로 할당하는 경우, 평가 표본과 학습 표본의 시간 격차가 지나치게 벌어지게 된다. 본문에서는 이 두 가지 측면을 고려하여, 전체 시계열 구간을 3개의 하위 구간으로 분할한 경우의 결과를 수록하였다.⁵⁾

아래 <표 10>은 각 모형들의 MAPE, <표 11>은 R-squared를 비교하여 보여준다. 먼저 확인할 수 있는 것은, 표본수가 약 3분의 1로 감소하였음에도 불구하고, 예측력 저하가 심하지 않다는 것이다. 예를 들어 CatBoost의 경우, MAPE 기준 정확성이 4.2~5.9로 3배의 표본 수를 사용한 앞 장의 분석 결과에 비해 약 1~2% 정도의 평균 오차 증가만이 있는 것으로 나타났다. 이는 시계열 배열 순서를 예측에 고려하더라도 모형의 정확성이 상당한 수준으로 유지될 수 있음을 의미한다. 둘째, 앞 장에서와 마찬가지로 XGBoost와 LightGBM에 비해 근소하지만 CatBoost의 예측능력이 더 우수한

<표 10> 각 시계열 구간 별 MAPE의 비교

	구간 1	구간 2	구간 3
XGBoost	6.18	4.62	5.66
LightGBM	6.93	4.61	5.53
CatBoost	5.93	4.2	4.82

<표 11> 각 시계열 구간 별 R-squared의 비교

	구간 1	구간 2	구간 3
XGBoost	0.943	0.977	0.97
LightGBM	0.938	0.973	0.973
CatBoost	0.946	0.978	0.981

5) 본 연구의 자료는 2009년 7월부터 2019년 12월까지를 포함하므로, 3개로 분할하는 경우 각 구간이 약 3년 반 정도의 기간을 포함한다고 볼 수 있다.

것으로 나타났다. MAPE 기준 XGBoost는 4.62~6.18, LightBoost는 4.61~6.93의 정확성을 보이고 있는데, CatBoost는 4.2~5.93으로 약 평균 1% 가량 예측력이 더 높은 것으로 나타났다.

VI. 결론 및 정책적 시사점

본 연구는 2009년부터 2019년까지 서울에서 거래된 아파트 매매 데이터를 통해 헤도닉 모형과 세 알고리즘(XGBoost·LightGBM·CatBoost) 기반의 주택 가치 감정 모형을 비교·분석하였다. 분석 결과, 전통적인 선형 헤도닉 모형에 비해 기계 학습 알고리즘의 예측력이 월등히 높게 나타났다. 회귀분석 기반 모형의 경우, 예측의 평균 백분율 오차는 약 11.5% 내외인 반면, XGBoost·LightGBM·CatBoost는 각각 3.7%, 3.8%, 3.6%에 불과했다. 이는 기계 학습 기반의 예측력이 다른 모형과의 상대적인 비교뿐 아니라, 절대적 수치 측면에서도 유의미하게 낮아질 수 있다는 것을 의미한다. 본 연구에서 사용한 세 알고리즘 중 미세하나마 CatBoost 알고리즘의 예측력이 다른 두 알고리즘에 비해 더 높게 나타나는 것도 확인할 수 있었다. 특히, 이상치의 발생빈도가 CatBoost에서 더 낮았으며, 이는 모형의 평균적 정확성뿐 아니라, 안정성이 다른 알고리즘에 비해 우수하다는 것을 의미한다.

흥미로운 것은, 세 알고리즘의 앙상블 모형이 개별 알고리즘보다 더 높은 예측 정확성을 가지는 것으로 나타났다는 것이다. 이는 본 연구가 기계학습을 통한 주택감정모형의 구축 과정에서, 예측력이 높은 특정 알고리즘을 선별해 활용하는 것뿐만 아니라, 다양한 알고리즘의 조합을 통해 정확도를 더 상승시킬 여지가 있음을 보여주고 있음을 의미한다.

이처럼 기계학습 알고리즘을 통해 감정 모형의 정확도가 상승한다는 것은 무엇을 의미할까? 현재도 다양한 분야에서 주택 가격 감정 시스템 (또는 모형)을 활용하고 있다. 예를 들어, 주택가격지수나 조세액의 추정 시 직접 관찰되지 않은 (시장에서 거래되지 않은) 주택의 가격은 모형을 통해 추정되고 있다. 그러나 본 연구에서도 알아본 바와 같이 선형 헤도닉 모형 기반의 가격 추정은 약 10% 내외의 상당한 예측오차를 동반할 수 있으므로, 보다 정교한 추정이 요구되는 실용적 활용 가능성은 제약되어 있다고 볼 수 있다. 기계학습을 이용해 모형의 정확도를 실용적으로 활용할 수 있는 일정 수준 이하로 끌어올리는 경우, 기존의 활용되던 분야에서의 생산성과 정확성이 상승할 뿐 아니라, 주택감정모형을 통한 다양한 활동(사업성 분석이나 자산 관리 등)이 본격적으로 가능해질 것으로 예상된다.

그럼에도 불구하고 이와 같은 기계학습 기반의 모형들은 모형의 해석 가능성이 매우 낮다는 단점을 가지고 있다. 전통적인 헤도닉 모형은 쉽게 해석 가능한 특정 형태의 수식으로 변수 간 관계를 표현하는 반면, 본 연구에서 분석한 기계학습 모형들은 변수와 가격 간의 관계를 직접적으로 표현하는

수익을 갖지 않기 때문에 종합적인 예측력은 상승하지만 그 가치 결정의 구조를 직관적으로 분석하는데에는 한계가 있다. 후속 연구들에서는 기계학습을 통해 얻은 예측자가 변수 변화의 구조에 어떤 식으로 예측값을 변경하는지를 2차적으로 분석하는 등의 시도들도 가능할 것으로 보인다. 이후에는 기계학습 모형의 높은 예측 정확성을 통해 수행할 수 있는 다양한 활용법뿐만 아니라, 기계학습을 통해 실제 부동산 시장의 가치 결정구조에 어떤 비선형성이 존재하는지를 파악하는 모형들도 더 제시될 것으로 기대한다.

참고문헌

- Anselin, L., 2001, "Spatial econometrics". A Companion to Theoretical Econometrics, 310330.
- Antipov, E. A. and Pokryshevskaya, E. B., 2012, "Mass appraisal of residential apartments: An application of random forest for valuation and a CART-based approach for model diagnostics," *Expert Systems with Applications*, 39(2), 1772-1778.
- Basu, S. and Thibodeau, T. G., 1998, "Analysis of spatial autocorrelation in house prices," *The Journal of Real Estate Finance and Economics*, 17(1), 61-85.
- Conway, D., Li, C. Q., Wolch, J., Kahle, C. and Jerrett, M., 2010, "A spatial autocorrelation approach for examining the effects of urban greenspace on residential property values," *The Journal of Real Estate Finance and Economics*, 41(2), 150-169.
- Dubin, R. A., 1992, "Spatial autocorrelation and neighborhood quality," *Regional Science and Urban Economics*, 22(3), 433-452.
- Fan, G. Z., Ong, S. E., and Koh, H. C., 2006, "Determinants of house price: A decision tree approach," *Urban Studies*, 43(12), 2301-2315.
- Goodman, A. C., 1978, "Hedonic prices, price indices and housing markets," *Journal of Urban Economics*, 5(4), 471-484.
- Greene, W. H., 2003, *Econometric analysis*. Pearson Education India.
- Gu, J., Zhu, M. and Jiang, L., 2011, "Housing price forecasting based on genetic algorithm and support vector machine," *Expert Systems with Applications*, 38(4), 3383-3386.
- Halvorsen, R. and Pollakowski, H. O., 1981, "Choice of functional form for hedonic price equations," *Journal of Urban Economics*, 10(1), 37-49.
- Hong, J., Choi, H. and Kim, W. S., 2020, "A house price valuation based on the random forest approach: The mass appraisal of residential property in South Korea," *International Journal of Strategic Property Management*, 24(3), 140-152.
- IAAO. Standard on Mass Appraisal of Real Prop.: IAAO: Kansas City, MO, USA, 2017.
- Lancaster, K. J., 1966, "A new approach to consumer theory," *Journal of Political Economy*, 74(2), 132-157.

- Limsombunchai, V., 2004, "House price prediction: Hedonic price model vs. artificial neural network," In New Zealand Agricultural and Resource Economics Society Conference (pp. 25-26).
- Malpezzi, S., 2002, "Hedonic pricing models: A selective and applied review," *Housing Economics and Public Policy*, 67-89.
- McCluskey, W., and Anand, S., 1999, "The application of intelligent hybrid techniques for the mass appraisal of residential properties," *Journal of Property Investment & Finance*.
- Meese, R. and Wallace, N., 1991, "Nonparametric estimation of dynamic hedonic price models and the construction of residential housing price indices," *Real Estate Economics*, 19(3), 308-332.
- Mu, J., Wu, F. and Zhang, A., 2014, "Housing value forecasting based on machine learning methods," In *Abstract and Applied Analysis* (Vol. 2014). Hindawi.
- Park, B., and Bae, J. K., 2015, "Using machine learning algorithms for housing price prediction: The case of fairfax county, virginia housing data," *Expert Systems with Applications*, 42(6), 2928-2934.
- Rasmussen, D. W. and Zuehlke, T. W., 1990, "On the choice of functional form for hedonic price functions," *Applied Economics*, 22(4), 431-438.
- Rosen, S., 1974, "Hedonic prices and implicit markets: Product differentiation in pure competition," *Journal of Political Economy*, 82(1), 34-55.
- Selim, H., 2009, "Determinants of house prices in Turkey: Hedonic regression versus artificial neural network," *Expert Systems with Applications*, 36(2), 2843-2852.
- Sheppard, S., 1999, "Hedonic analysis of housing markets," *Handbook of Regional and Urban Economics*, 3(1), 595-1635.
- Wang, D. and Li, V. J., 2019, "Mass appraisal models of real estate in the 21st century: A systematic literature review," *Sustainability*, 11(24), 7006.
- Zhou, G., Ji, Y., Chen, X. and Zhang, F., 2018, "Artificial neural networks and the mass appraisal of real estate," *International Journal of Online and Biomedical Engineering (IJOE)*, 14(03), 180-187.
- Zurada, J., Levitan, A., and Guan, J., 2011, "A comparison of regression and artificial

intelligence methods in a mass appraisal context”, Journal of Real Estate Research, 33(3), 349-387.

(논문 접수일: 2020.10.17. 수정논문 접수일: 2020.11.30. 논문 채택일: 2020.12.11.)

부록. 주요 하이퍼파라미터 설정 내용

〈부록 표 1〉 XGBoost의 하이퍼파라미터

하이퍼파라미터명	설정내용
base_score	0.5
gamma	10
learning_rate	0.1
max_depth	20
min_child_weight	1
n_estimators	100
n_jobs	1
nthread	None
objective	reg:linear
random_state	0
reg_alpha	0
reg_lambda	2
scale_pos_weight	1
colsample_bytree	1
subsample	1
verbosity	1

〈부록 표 2〉 LightGBM의 하이퍼파라미터

하이퍼파라미터명	설정내용
boosting_type	gbdt
class_weight	None
colsample_bytree	1.0
importance_type	split
learning_rate	0.1
max_depth	200
min_child_samples	20
min_child_weight	0.001
min_split_gain	0.0
n_estimators	3000
n_jobs	-1
num_leaves	20000
random_state	None
reg_alpha	0.0
reg_lambda	0.0
subsample	1.0
n_estimators	3000
min_data_in_leaf	5

〈부록 표 3〉 CatBoost의 하이퍼파라미터

하이퍼파라미터명	설정내용
nan_mode	Min
eval_metric	RMSE
iterations	4000
sampling_frequency	PerTree
fold_permutation_block	0
leaf_estimation_method	Newton
grow_policy	SymmetricTree
penalties_coefficient	1
boosting_type	Plain
model_shrink_mode	Constant
feature_border_type	GreedyLogSum
bayesian_matrix_reg	0.10
l2_leaf_reg	1
random_strength	1
rsm	1
boost_from_average	True
max_ctr_complexity	4
model_size_reg	0.5
subsample	0.80
use_best_model	False
depth	12
sparse_features_conflict_fraction	0
leaf_estimation_backtracking	Anyimprovement
best_model_min_trees	1
model_shrink_rate	0
min_data_in_leaf	1
loss_function	RMSE
learning_rate	0.1
score_function	Cosine
leaf_estimation_terations	1
bootstrap_type	MVS
max_leaves	64

An Application of XGBoost, LightGBM, CatBoost Algorithms on House Price Appraisal System

Jengei Hong^{*}

Abstract

This paper compares the predictive power of a conventional hedonic pricing model and three machine learning algorithm (XGBoost, LightGBM, CatBoost) based models by using 620,617 apartment data in Seoul from 2009 to 2019. The results are summarised as follows; First, the predictive power of the machine learning models are significantly high not only in the comparison to the conventional model but also in the absolute accuracy related to its practical usefulness. The mean percentage error of XGBoost, LightGBM, and CatBoost were only, respectively, 3.7%, 3.8%, and 3.6% while those of the hedonic model was around 11%. Second, we found that CatBoost algorithm is slightly more performative to the other two algorithms in terms of overall predictive power and frequency of outlier occurrences. Third, this paper show that an ensemble model of the three algorithms can raise the predictive power further.

Keywords : Machine Learning, XGBoost, LightGBM, CatBoost, Mass Appraisal

^{*} Jengei Hong, Corresponding author, School of Management & Economics, Handong Global University, Assistant Professor, hwgh024@handong.edu

© Copyright 2020 Housing Finance Research Institute. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

자산심사제도 도입 전후 내집마련 디딤돌 대출의 특성 변화 분석

김성진*

요 약

2019년 9월 말부터 주택도시기금의 주택 구입 및 전세자금대출을 받으려면 소득 외에 일정 수준 이하의 순자산 기준을 만족해야 한다. 이는 서민 주거안정을 지원하는 기금 대출에서 소득이 적으나 자산이 많은 부적격자를 제외하여 재원의 낭비를 막으려는 목적을 가진다. 자산심사 제도가 도입된 지 1년이 넘은 지금, 이 제도의 영향에 대한 점검이 필요한 시점이다. 본 연구는 분위 회귀분석을 통해 자산심사 제도 도입 전후 디딤돌 대출 특성 변화를 분석하고, 이를 통해 자산심사제도의 적정성과 효과를 파악하였다. 우선 기존 대출자의 특성은 크게 변화하지 않았다. 이는 기존 제도를 유지하면서 일부 부적격자들만 제한하는 효과가 있었던 것을 의미한다. 특성요인 중에는 대출자 특성과 금리 우대조건에 해당하는 변수의 영향력이 자산심사 도입 이후 커졌다. 또한, 상대적으로 대출금액이 하위 분위에서 영향력이 확대된 것을 확인하였다. 다만 이러한 효과는 자산심사 제도 도입에 따른 간접적인 영향에 그친다는 한계를 가진다. 따라서 향후 자산심사 자료 등을 확보하여 보다 개선된 분석이 수행되어야 할 것으로 판단된다.

핵심어 : 주택도시기금, 디딤돌대출, 자산심사제도, 대출 특성요인, 분위회귀분석

I. 서론

우리나라는 1981년부터 정책적으로 국민주택기금을 조성하여 근로자·서민의 주거 안정을 도모하는 여러 가지 사업을 진행하고 있다. 2015년부터는 국민주택기금이 주택도시기금으로

* 김성진, 주저자, 우리은행 부동산연구실 연구위원, sjkim2210@wooribank.com

© Copyright 2020 Housing Finance Research Institute. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

명칭이 개편되면서 주거복지 외에 도시재생으로 역할을 확대하였으며, 국민의 주거 문제를 좀 더 폭넓게 다루어 종합적인 주택·도시 문제를 해결하고자 하는 노력을 기울이고 있다. 주택도시기금이 도시재생으로 역할이 확대되었지만, 기존의 주거 안정을 위한 노력을 소홀히 하는 것은 아니다. 2013년 공유형 모기지, 2014년 내집마련 디딤돌 대출, 2015년 버팀목 전세자금 대출, 주거안정 월세 대출, 2018년 중소기업 취업청년 버팀목 전세대출 등 새로운 대출상품을 출시하여 서민의 주거 안정을 다각도로 지원하고 있으며, 기존의 대출상품도 시장 상황에 맞게 개선하여 취약계층, 신혼부부, 생애최초 주택구입자, 다자녀 가구, 청년층 등 정책적으로 지원이 필요한 계층에는 보다 우대된 조건을 통해 지원을 확대하고 있다.

하지만 이러한 대출에서 자격기준을 대출가구의 소득수준에 한정하여 판단하다 보니, 대출자격 자체에 대한 논란이 꾸준히 제기되어 왔다. 특히 근로소득자는 소득의 파악이 용이하지만, 자영업자, 무직 등 소득이 불명확한 경우에는 근로소득자보다 상대적으로 기금 대출을 받기 수월한 경우가 많았다. 이는 비단 주택도시기금의 문제만은 아니어서 임대주택 등 정책적으로 대상을 한정하고 지원하는 제도에 있어서는 종종 제기되는 문제였다. 이에 주택도시기금에서는 기존의 대출 제도에 자산심사 제도를 추가 도입하여 기존의 소득 외에 기타 부동산, 자동차, 일반자산 및 부채, 금융자산 및 부채 등을 종합적으로 심사하여 설정한 기준¹⁾을 넘지 않은 경우에만 기금대출의 자격을 부여하기 시작했다. 이러한 자산심사 제도는 2019년 9월 30일부터 도입되었는데, 1년 남짓 남은 시점인 지금, 아직 도입기간이 길지 않아 그 영향과 효과에 대한 검증이 제대로 이루어지지 않은 것으로 판단된다. 만약 현재의 자산심사 제도가 지나치게 까다로운 조건이라면 필요 이상으로 대상자를 제한하여 혼란을 가져올 것이며, 반대로 너무 완화된 조건이라면 도입에 따른 대출의 번거로움만 증가하였을 뿐 자산심사의 실제적인 효과는 없을 것이기 때문이다.

본 연구에서는 2019년 이후 실제 디딤돌 대출로 취급된 대출자료를 이용하여 자산심사 제도 도입 전후의 대출자 특성 변화를 비교하여 살펴보고, 회귀분석과 분위 회귀분석의 통계적 방법을 이용하여 대출자의 대출 결정요인을 분석하고자 한다. 회귀분석의 방법은 개별특성이 종속변수에 미치는 영향을 파악하는 가장 효과적인 방법이며, 분위 회귀분석은 회귀분석의 단점을 보완하여 설명변수의 영향이 종속변수의 수준에 따라 어떻게 달라지는지도 살펴볼 수 있는 장점이 있다. 이를 통해 자산심사 제도 도입으로 인해 기금 대출 대상자의 특성이 어떻게 변화하였으며, 자산심사 제도의 도입이 기금대출에 어떠한 영향을 미쳤는지 살펴보는 것을 목적으로 한다.

1) 2020년 기준 버팀목대출 대상자의 순자산 기준은 3.91억원 이하임.

II. 선행연구

본 연구와 같이 주택담보대출의 대출 이용에 관련한 특성을 분석하는 논문은 크게 두 가지로 구분된다. 하나는 대출 특성요인의 차이에 따라 어떤 대출상품을 선택하는지에 대한 것이고, 다른 하나는 특성요인에 따라 대출의 이용 정도가 어떻게 달라지는지에 대한 논문이다. <표 1>에서는 이 두 가지 분류의 국내·외 논문을 정리하였다.

먼저 대출 선택과 관련된 논문은 주로 이항 로짓이나 프로빗 등의 분석방법을 활용한다. 정의철, 권혜진(2006)은 주택담보대출 상환방식에 영향을 주는 요인에 대한 이원 프로빗 분석을 실시하였고,

<표 1> 주택담보대출 관련 주요 연구

구분	저자	분석모형	종속변수	설명변수
대출조건 선택	정의철, 권혜진 (2006)	이원 프로빗	대출상환방식	소득, 연령, 맞벌이, 자녀수, 순자산, 금리차, 지역 등
	강성덕 (2014)	이항로짓	수익-손익 공유형 선택	나이, 성별, 가구원수, 무주택기간, 신혼유무, 아파트가격, 면적, 단지 규모, 경과년수
	김정협, 이현석 (2016)	2단계 선택모형	만기-분할 원금-원리금	개별특성:성별, 나이, 직업, 주택유형, 규모, 소득 등 금융특성:신용등급, DTI, LTV, 대출액, 상환기간 등
	Coulibaly et al. (2009)	이항로짓	고정-변동금리 선택 여부	고정-변동 금리차, 대출만기, 리파이낸싱 여부, 서브프라임 모기지 여부, 가구의 수입 초과 소비여부, 소득대비 대출잔액비율, 리스크 태도 등
대출이용 특성	이대식 외 (2011)	다중회귀	대출금액	나이, 소득, 부채, 면적, 주택보유수, 담보평가금액, LTV, CB등급, 대출기간
	강성진, 이주석 (2012)	Type-II 토빗모형	대출 유무 및 대출금액	대출만기, 거치기간, 소득, 맞벌이, 연령, 결혼, 주택시장전망, 아파트, 수도권, 주택보유채수, 면적 등
	김기태, 오동훈 (2015)	다중회귀	대출신청금액	연령, 세대원수, 주택연수, 단지규모, 청약저축가입기간, 주택매매가격, 부채금액, 소득대비부채비율 등
	김성진(2016)	분위회귀	대출금액	나이, 성별, 직업, 가구원수, 연소득, 생애최초, 다자녀, 다문화, 장애인, 수도권, 아파트, 전용면적, 주택가격, 대출기간 등
	Boehm et al. (2006)	다중회귀	모기지 이자율	성별, 학력, 가구원수, 소득, 연령, 주택입지, 가격, 대출기간, LTV, 개인보험 여부 등
	Al-Bahrani et al. (2015)	분위회귀	모기지 이자율	신용도, 대출기간, 대출금액, LTV, 대출목적 등

선택에 영향을 미치는 특성요인으로 소득, 연령, 맞벌이, 자녀수, 순자산, 금리차 등을 사용하였다. 김정협, 이현석(2016)도 주택담보대출의 상환방식을 만기일시 및 분할 상환방식으로 구분하고, 분할 상환방식을 다시 원금균등상환방식과 원리금균등상환방식으로 구분하는 2단계 선택모형을 통해 특성요인에 따른 상환방식 선택의 차이를 연구하였다. 성별, 나이 등 개별특성 외에, 신용등급, DTI비율, LTV비율 등의 금융특성을 사용한 것이 다른 분석과의 차이점으로 만기일시 및 분할 상환방식의 선택에서는 투자 여부와 관련된 특성이, 분할상환방식 간의 선택에 있어서는 소득수준, 대출금리 등 대출상환과 관련된 특성이 중요한 요인이라고 설명하고 있다. 강성덕(2014)은 국민주택기금의 공유형 모기지 시범사업의 두 가지 상품인 수익공유형과 손익공유형의 선택에 영향을 주는 요인을 로짓분석을 활용하여 분석하였다. 설명변수로는 대출자의 연령, 성별, 가구원수 등의 대출자 특성요인과 아파트가격, 전용면적, 단지규모, 아파트 경과년수 등의 주택 특성요인을 활용하였다. 외국 논문으로는 Coulibaly et al.(2009)의 연구가 있는데 고정금리와 변동금리 대출 선택의 결정요인을 로짓모형을 활용하여 살펴보았다. 고정-변동금리차, 리파이낸싱 여부, 서브프라임 모기지 여부, 가구의 수입 초과 소비 여부 등 우리나라와는 다른 미국의 상황을 반영한 변수들을 분석에 포함하였다. 살펴본 연구 모두 자료의 한계로 제한된 변수를 차별화해 사용하였으며, 특히 서베이와 같이 표본수가 충분하지 않을 때에는 설명력이 그리 높지 않는 등 한계를 갖고 있었다.

본 연구와 유사한 주택담보대출 이용의 특성에 관한 연구에는 주로 회귀분석에 기반한 연구들이 주를 이루었다. 이대식 외(2011), 김기태, 오동훈(2015), Boehm(2006) 등이 다중회귀모형으로 분석하였으며, 강성진, 이주석(2012), 김성진(2016), Al-Bahrani(2015) 등은 분위회귀모형을 분석에 활용하였다. 이대식 외(2011)는 보금자리론의 대출자료를 이용하여 차입자의 특성에 따른 대출 이용을 살펴보았다. 나이, 소득, 부채, 면적, 담보평가금액, LTV, CB등급, 대출기간 등의 설명변수를 사용하였으며, 대부분 요인이 주택담보대출에 영향을 주는 것으로 분석되었다. 김기태, 오동훈(2015)은 국민주택기금의 공유형 모기지 대출 특성자료를 활용하여 대출신청금액에 미치는 영향을 살펴보았는데, 수익형과 손익형의 두 가지 대출 간에는 서로 다른 요인이 영향을 미치는 것으로 확인되었다. 수익공유형 모기지는 세대원수, 부채비율, 청약저축 가입 기간, 주택가격, 단지 규모의 특성요인이 양(+)의 영향을 미치는 것으로 나타났고, 손익공유형 모기지에는 특성요인 중 청약저축가입기간이 음(-)의 부호로 차이를 보였다. 김성진(2016)은 주택도시기금의 구입 및 전세대출의 대출금액이 각각의 특성요인에 따라 어떤 영향을 주는지를 토빗 회귀모형과 분위 회귀모형 등을 활용하여 분석하였다. 대출자 특성과 금리 우대조건 등 대부분에서 양(+)의 영향을 미치는 것으로 분석되었으며, 기금 지원이 효과를 나타내고 있음을 증명하였다. Al-Bahrani(2015)는 분위

회귀모형을 이용하여 소비자의 신용도, 대출기간, 대출금액, LTV 비율 등이 주택담보대출의 부대비용에 미치는 영향을 분석하였으며, 높은 신용등급을 보유한 대출자가 상대적으로 낮은 대출관련 비용을 부담하는 것을 증명하였다.

이와 같이 주택담보대출의 대출특성과 관련된 연구들은 서베이나 실제 대출자료를 이용하여 분석하였으며, 자료의 한계로 인해 설명변수는 연구마다 다소 차이를 보였다. 다만 대출자 특성, 대출상품 특성, 담보주택 특성, 금융변수, 시장 등 외적요인을 설명변수로 주로 활용하였으며, 정도의 차이는 있지만 대부분 유의미한 값을 도출하였다. 본 연구는 주택담보대출 이용 특성과 관련된 문헌들에서 주로 사용된 회귀분석을 기반으로 분위 회귀분석도 추가하였다. 설명변수로는 기존 문헌들에서 고려한 요인 중 대출자 특성, 대출상품 특성, 담보주택 특성 관련 요인은 설명변수에 포함하되, 짧은 분석기간과 자료의 한계로 금융변수, 시장요인은 분석에서 제외하였다.

III. 분석 방법 및 자료

1. 분위 회귀분석

회귀분석은 종속변수와 설명변수 간의 관계, 즉 설명변수의 변화에 따른 종속변수의 변화를 살펴볼 때 가장 일반적이고 유용한 모형으로 그 활용도가 매우 높다. 본 연구에서는 다중 회귀분석 외에 종속변수의 수준에 따라 설명변수의 영향력을 달리 측정하는 분위 회귀분석을 함께 활용하여 디딤돌 대출의 특성을 보다 세부적으로 살펴보았다. Koenker and Bassett(1978)에 의해 제시된 분위 회귀모형은 다중 선형 회귀모형의 한 종류이지만 OLS(Ordinary Least Squares)처럼 종속변수가 아닌 중위수나 특정 분위수를 예측할 때 주로 활용된다. 분위 회귀모형은 조건이 다른 r-분위(conditional r-quantiles)의 분포를 설명하여, 종속변수의 평균을 중심으로 하는 선형모형을 나타내는 OLS와 다른 차이를 보인다. 각 수준별로 설명변수가 미치는 종속변수의 결과가 다르다면 평균을 기준으로 하는 OLS로는 정확한 파악을 할 수 없게 된다. 하지만 분위 회귀분석은 조건부 분위함수에 기반하여 종속변수의 조건부 분포상의 여러 지점에서 설명변수 사이의 관계에 대한 정보를 제공하는 장점을 가진다.

$$y_i = \alpha_i + \beta_i * D_i + \gamma * x_i + \epsilon_i \quad \langle \text{식 1} \rangle$$

일반적인 회귀분석이 위와 같은 함수의 모형을 갖는다면, 분위 회귀분석의 아래의 함수 형태를 가진다.

$$y_i = \beta_q * x_i + \epsilon * q_i$$

$$\Omega_q(y_i|x_i) = \beta_q * x_i \quad \langle \text{식 2} \rangle$$

x_i = 처리변수 및 공변량 포함 벡터
 $\Omega_q(y_i|x_i)$ = 공변량에 의한 q 분위 결과 변수

OLS와 비교하여 분위 회귀분석의 장점을 살펴보면, OLS는 분위별 분석을 위해 구간을 구분할 경우, 많은 정보를 잃게 되어 편기(biased)된 추정결과를 가져올 수 있는 단점이 있지만, 분위 회귀분석은 모든 관측치를 사용하여 자료를 나누는데 따른 정보손실의 단점을 보완하여 효율적인 값을 추정할 수 있다. 둘째, 분위 회귀분석은 종속변수의 분위수별로 서로 다른 헤도닉적인 특성을 고려할 수 있게 됨으로써 OLS에서 가질 수 있는 이분산(heterogeneity)의 문제를 해결할 수 있다. 마지막으로 분위 회귀모형은 이상치(outliers)나 오차항 분포에 민감하지 않은 로버스트성(robustness)의 특징을 가진다. 이러한 특징들로 분위 회귀분석은 종속변수의 정도에 따라 각 변수의 중요성이 상대적으로 달라지는 점을 유의하게 확인할 수 있어 실증분석에 높은 활용도를 가진다.

2. 분석 자료

본 연구에 사용된 데이터는 2019년 1월부터 2020년 6월까지(1년 6개월간) A은행에서 취급된 내집마련 디딤돌 대출의 개별 특성자료이다. 주택도시보증금의 주택구입자금인 내집마련 디딤돌 대출은 5개 은행 및 주택금융공사에서 취급하고 있으나, 자료의 구득의 한계로 본 연구에서는 A은행에서 취급된 대출자료만 분석에 사용하였다. 다만 은행별 대출조건이 같은 기금 대출이므로 대출특성이 유의미하게 다르지 않을 것으로 가정하였다. 모형에서 사용된 변수는 <표 2>에 설명되어 있으며, 종속변수는 차주별 대출금액으로 한정하였다.

종속변수인 대출금액에 영향을 미치는 설명변수를 크게 대출자 특성, 금리 우대조건, 담보물 특성, 대출특성의 4개의 범주로 구분하였다. 먼저 대출자 특성은 나이, 성별, 직업유무, 연소득으로 대출 차주에 따라 대출금액에 차이를 줄 수 있는 요인들로 구성되었다. 금리 우대조건은 디딤돌 대출에서 금리 우대 또는 대출한도 완화 등을 통해 기금에서 별도로 지원하는 가구에 대한 변수이며, 신혼부부,

〈표 2〉 모형의 변수 정의

변수명			단위	내용
종속변수	대출금액(loan)		백만원	총 대출금액
대출자 특성	나이(age)		세	가구주 나이
	성별(sex)		-	남=1, 여=0
	직업유무(job)		-	직업보유=1, 무직=0
	연소득(income)		백만원	부부합산 연소득(기금취급기준)
금리 우대조건	신혼부부(marriage)		-	생애최초 신혼부부=1, 비신혼부부=0
	다자녀(child)		-	1자녀=1, 2자녀=2, 3자녀이상=3
	기타(etc)		-	생애최초 신혼부부, 다자녀를 제외한 금리 우대조건=1, 해당없음=0
담보물 특성	수도권(sma)		-	수도권=1, 비수도권=0
	아파트(apt)		-	아파트=1, 아파트 외=0
	전용면적(lotsize)			주택 전용면적
	주택가격(price)		백만원	주택 담보인정가액
대출특성	대출기간 (term)	15년(year15)	-	대출기간 15년=1, 해당없음=0
		20년(year20)	-	대출기간 20년=1, 해당없음=0
		30년(year30)	-	대출기간 30년=1, 해당없음=0

다자녀, 기타로 구분하였다. 신혼부부는 일반 디딤돌 대출보다 우대가 강화된 생애최초 신혼부부로 한정하였으며, 다자녀는 자녀 수에 따라 우대금리가 확대되는 점을 반영하여, 1자녀, 2자녀, 3자녀 이상의 세 가지로 구분하였다. 기타 우대조건은 생애최초가 아닌 신혼부부, 다문화, 노인우대, 한부모 가구 등 나머지 금리우대 사항을 포함하였으며, 동일하게 0.2%의 금리우대를 적용받는다. 담보물 특성은 대출의 담보가 되는 주택의 특성으로 수도권 여부, 아파트 여부, 전용면적, 주택가격으로 구분하였다. 대출특성은 대출상품 자체의 특성으로 대출기간만을 선택하였다. 디딤돌 대출의 대출기간은 10년, 15년, 20년, 30년의 4가지 만기상품으로 운영되는데, 15년, 20년, 30년의 3가지 더미변수로 설정하여 3가지 값이 0인 경우에는 대출기간이 10년인 것으로 설정하였다.

종속변수 및 설명변수의 기초적인 특성을 나타내는 기술통계는 〈표 3〉에 정리되어 있으며, 개별 데이터 중 이상치가 발견되거나 특성이 누락된 자료는 제외²⁾하였다. 변수의 특성을 본 연구의 핵심인

2) 이상치는 변수들이 가질 수 있는 범위를 벗어나는 값이 있을 경우, 분석에서 제외하였으며, 내부 시스템을 통해 오류를 확인할 수 있는 경우에는 값을 수정하여 활용하였음.

〈표 3〉 변수의 기술통계

구분		자산심사 도입 전				자산심사 도입 후			
		최소값	최대값	평균	표준편차	최소값	최대값	평균	표준편차
종속변수	대출금액	5	240	119.81	52.235	5	260	116.02	52.528
대출자 특성	나이	20	87	41.82	9.818	20	87	42.22	10.055
	성별	0	1	0.67	0.471	0	1	0.64	0.479
	직업유무	0	1	0.92	0.277	0	1	0.93	0.260
	연소득	1.0	70.0	37.80	16.455	1.2	70.0	35.33	15.579
금리 우대조건	신혼부부	0	1	0.31	0.464	0	1	0.35	0.478
	다자녀	0	3	0.78	0.904	0	3	0.77	0.899
	기타	0	1	0.50	0.500	0	1	0.48	0.500
담보물 특성	수도권	0	1	0.57	0.496	0	1	0.50	0.500
	아파트	0	1	0.79	0.405	0	1	0.75	0.434
	전용면적	12.54	99.00	66.81	16.342	12.88	99.51	64.90	16.475
	주택가격	11.0	600.0	243.05	93.619	10.0	528.0	218.90	86.171
대출특성	대출기간	10	30	24.84	7.270	10	30	25.37	7.008
	15년	0	1	0.10	0.298	0	1	0.08	0.272
	20년	0	1	0.16	0.370	0	1	0.16	0.367
	30년	0	1	0.64	0.481	0	1	0.67	0.471
자료수		N=7,380				N=5,807			

자산심사 도입 전과 자산심사 도입 후로 구분하여 살펴보았는데, 먼저 분석에 사용된 자료의 수는 자산심사 도입 전이 7,380건, 도입 이후가 5,807건으로 도입 전이 다소 많았다. 종속변수인 대출금액의 평균은 도입 전이 1억 1,981만원이었던 데 비해, 자산심사 도입 이후 1억 1,602만원으로 다소 줄어들었다. 시장에서 주택가격이 상승하였음에도 불구하고 대출금액이 줄어든 것은 자산심사 제도 도입에 따라 자산은 많으나, 소득이 적은 일부 부적격자들이 제외되면서 수도권, 아파트 비중이 감소하고, 평균 전용면적과 주택가격이 줄어들었기 때문으로 판단된다. 대출자 특성에서는 자산심사 도입 이후 평균 나이가 다소 증가하였으나, 남성비율과 연소득은 다소 줄어들었다. 금리 우대 조건에서는 생애최초 신혼부부 비중이 다소 증가하였으나, 기타 우대조건은 소폭 감소하였다. 신혼부부는 현 정부에서 가장 중점적으로 정책지원을 하는 계층으로 자산심사 도입 외에도 지속적인

지원 강화가 비중 증가를 가져온 것으로 판단된다. 담보물 특성에서는 자산심사가 도입된 이후 수도권과 아파트 비중이 줄고, 전용면적과 주택가격도 줄어드는 것으로 나타났는데, 이 시기 수도권의 가격 상승과 맞물려 디딤돌 대출의 지방 비율이 늘면서 나타나는 결과로 판단된다. 또한, 고가의 주택보다는 상대적으로 저가의 주택에 대한 대출이 늘어, 실질적으로 수혜계층이 저소득 위주로 일부 늘어난 것을 의미한다. 마지막으로 대출 기간은 30년 만기의 선택비율이 64%에서 67%로 더 늘어난 것을 알 수 있었다.

IV. 실증 분석

1. 회귀분석 결과

앞에서 설명한 변수를 사용하여 대출금액을 종속변수로 하고, 4개 범주의 특성요인을 설명변수로 하는 회귀분석 모형을 설정하였다. 모형은 자산심사가 도입된 2019년 9월 30일 전후의 2개의 모형을 각각 설정하였으며, 모형식³⁾은 아래와 같다.

$$\text{대출금액} = f(\text{대출자특성, 금리우대조건, 담보물특성, 대출특성}) \quad \langle \text{식 3} \rangle$$

$$\begin{aligned} \text{loan}_i = & b_0 + b_1 * \text{age}_i + b_2 * \text{sex}_i + b_3 * \text{job}_i + b_4 * \text{income}_i \\ & + b_5 * \text{marriage}_i + b_6 * \text{chld}_i + b_7 * \text{etc}_i \\ & + b_8 * \text{sma}_i + b_9 * \text{apt}_i + b_{10} * \text{lotsize}_i + b_{11} * \text{price}_i \\ & + b_{12} * \text{year15}_i + b_{13} * \text{year20}_i + b_{14} * \text{year30}_i \end{aligned} \quad \langle \text{식 4} \rangle$$

디딤돌 대출과 관련된 4가지 범주의 특성요인들이 대출금액에 미치는 영향을 살펴보기 위해 다중 회귀분석을 실시하였다. 자산심사 도입 전후의 변화를 살펴보기 위해 도입 전후 모형을 별도로 설정하여 분석하였으며, 그 결과는 <표 4>에 정리되어 있다. 우선 대부분의 설명변수가 유의수준 1% 내지 5%에서 통계적으로 유의한 것으로 나타났으며, 상수항을 제외하면 계수값의 부호가 바뀌지 않아 영향의 방향이 달라지지는 않은 것으로 분석되었다. 상수항의 부호는 (+)에서 (-)로 바뀌었으나, 자산심사 도입 후에는 통계적으로 유의한 값을 갖지 않았다. 모형의 설명력을 나타내는 수정된 R^2 은

3) 모형식에 사용된 영문 변수명은 ‘<표 2> 모형의 변수 정의’에 표기된 것과 동일함.

〈표 4〉 회귀분석 결과

구분		자산심사 도입 전				자산심사 도입 후			
		계수값	표준오차	t값	P값	계수값	표준오차	t값	P값
상수항		14.4694**	4.2789	3.38	0.0010	-2.8872	4.6790	-0.62	0.5370
대출자 특성	나이	-0.5915**	0.0565	-10.47	0.0000	-0.4335**	0.0584	-7.43	0.0000
	성별	-3.2633**	0.9972	-3.27	0.0010	-2.4024*	1.0531	-2.28	0.0230
	직업유무	8.5040**	1.6245	5.23	0.0000	11.1579**	1.8200	6.13	0.0000
	연소득	0.3654**	0.0305	11.98	0.0000	0.2715**	0.0332	8.18	0.0000
금리 우대조건	신혼부부	6.2504**	1.5269	4.09	0.0000	14.1143**	1.6823	8.39	0.0000
	다자녀	-2.0412**	0.5559	-3.67	0.0000	-1.6917**	0.5879	-2.88	0.0040
	기타	2.1645	1.2170	1.78	0.0750	4.9452**	1.3724	3.60	0.0000
담보물 특성	수도권	4.1501**	1.1015	3.77	0.0000	3.0178*	1.1798	2.56	0.0110
	아파트	-7.5109**	1.3302	-5.65	0.0000	-10.7733**	1.3245	-8.13	0.0000
	전용면적	0.1274**	0.0371	3.43	0.0010	0.1525**	0.0368	4.14	0.0000
	주택가격	0.2835**	0.0057	49.43	0.0000	0.3528**	0.0063	56.43	0.0000
대출특성 (대출기간)	15년	19.8166**	1.9392	10.22	0.0000	18.2603**	2.2231	8.21	0.0000
	20년	28.5528**	1.7377	16.43	0.0000	23.1828**	1.9141	12.11	0.0000
	30년	45.2639**	1.5000	30.18	0.0000	40.3204**	1.6615	24.27	0.0000
모형설명력		Adj R^2 =0.4922				Adj R^2 =0.5589			

주: **유의수준 1%에서 통계적으로 유의($p < 0.01$), *유의수준 5%에서 통계적으로 유의($p < 0.05$).

자산심사 도입 전 모형은 0.4922인 반면, 도입 이후의 모형은 0.5589로 도입 이후의 설명력이 보다 높아진 것으로 나타났다.

대출자 특성에서는 나이와 성별이 음(-)의 계수값을 가져 나이가 많을수록, 세대주가 여자보다 남자일 때 대출금액이 줄어들었다. 반대로 직업을 갖고 있고, 연소득이 높을수록 대출금액이 증가하는 것으로 분석되었다. 금리 우대조건은 신혼부부와 기타 우대조건일 경우 양(+)의 계수값으로 대출금액이 증가하는 것을 알 수 있으나, 다자녀일 경우에는 음(-)의 값을 가져 다자녀 우대를 받을수록 평균적으로 대출금액이 줄어드는 것으로 분석되었다. 이는 디딤돌 대출 대상자의 평균 자녀수가 0.7~0.8명 수준에 불과하고, 실제 3자녀의 평균 주택가격과 대출액이 1~2자녀보다 적었기 때문으로 다자녀 우대는 금리 우대는 받지만, 실제 대출 증가에 영향을 주지 못했음을

의미한다.

담보물 특성은 수도권에 있고, 전용면적이 크고, 주택가격이 높을수록 대출금액이 증가하는 것으로 나타났다. 반면, 아파트는 다른 주택유형에 비해 가격이 높다는 일반적인 인식과는 달리 동일한 조건 하에 다른 주택유형에 비해 대출금액이 적은 것으로 나타났다. 이는 아파트를 구입하는 대출자는 다른 주택유형 대출자에 비해 경제적 부담이 상대적으로 낮아 대출금액이 적을 가능성이 있으며, 이는 분위 회귀분석 결과에서 다시 한번 확인할 필요가 있겠다. 대출기간은 만기가 길어질수록 대출금액이 증가한 것으로 나타나, 대출금액이 많이 필요할 때 대출기간을 늘려 월상환부담액을 줄이는 일반적인 대출 특성이 반영되었다.

자산심사 도입 전후로 계수값의 변화를 살펴보면, 대출자 특성에서는 자산심사 도입 이후 직업 유무의 계수값은 커졌으나, 나이, 성별, 연소득의 계수값은 절대값이 작아진 것으로 나타났다. 과거에는 소득만 가지고 대출자격을 살펴봤다면, 자산심사 도입 이후에는 자산을 함께 고려하기 때문에 소득은 없지만 자산이 많은 부적격자들이 대출에 제한을 받으면서 직업 유무의 영향이 증가한 것으로 판단된다. 금리 우대조건에서는 다자녀의 영향이 줄고, 기타 우대조건과 생애최초 신혼부부의 계수값이 증가하였다. 특히 신혼부부의 계수값 증가가 두드러졌는데, 자산심사 도입과 함께 신혼부부 지원 확대가 지속적으로 이루어지고 있기 때문으로 판단된다. 신혼부부는 상대적으로 자산축적 기간이 짧아 소득기준을 완화하는 등 지원을 확대해준다면 그 효과가 확실하게 나타나는 계층으로 확인되었다. 담보물 특성에서는 아파트의 음(-)의 계수값이 커졌으며, 전용면적과 주택가격의 양(+)의 계수값이 증가하였다. 반대로 수도권의 계수값은 다소 줄어들었다. 자산심사 도입 이후 수도권 입지를 제외한 담보물이 갖는 특성에 대한 영향이 증가한 것을 알 수 있다. 이는 주택가격 상승기로 담보물의 특성이 대출 증가에 미치는 영향이 확대된 것으로 볼 수 있다. 마지막으로 대출기간이 늘어날수록 대출금액이 양(+)의 상관관계를 보였으나, 자산심사 도입 이후 그 영향은 다소 줄어들었다.

종합하면, 자산심사 도입 이후 설명변수의 계수 부호가 바뀔 만큼의 변화는 발생하지 않아 기존 디딤돌 대출 가능자들이 자산심사 도입 이후에도 대부분 대출 이용이 가능했을 것으로 판단된다. 이는 자산심사가 이전 대부분의 대상자들은 대출이 가능하면서 일부 부적격 대상자들만 제한하는 효과를 거둔 것을 의미할 수 있다. 큰 변화는 발생하지 않았지만, 특성요인 중에는 담보물 특성의 중요도가 높아지고, 신혼부부, 기타 우대조건 등 금리 우대조건, 직업 유무의 대출자 특성의 중요도 역시 높아진 것을 알 수 있다. 이는 자산심사 도입으로 대출에 있어서 담보물 자체의 가치와 정책적 지원계층에 대한 지원의 중요도가 상대적으로 높아져 효과를 본 것을 의미한다고 하겠다.

2. 분위 회귀모형 분석 결과

디딤돌 대출의 종속변수인 대출금액의 크기에 따른 설명변수의 영향의 차이를 살펴보기 위해 분위 회귀분석을 추가로 실시하였다. OLS 모형은 설명변수의 평균적인 영향만을 분석할 수 있지만, 분위 회귀분석은 종속변수의 분위별 영향의 차이를 살펴볼 수 있기 때문이다. 즉, 분위 회귀분석을 통해 대출금액 수준의 차이에 따라 설명변수의 영향이 어떻게 달라지는지 파악할 수 있으며, 자산심사 제도의 도입 효과가 어떤 분위에 보다 크게 나타나는지 살펴볼 수 있다.

먼저 자산심사 도입 전의 OLS 및 분위 회귀모형의 분석 결과는 <표 5>에 정리되어 있다. 분위 회귀분석은 종속변수인 대출금액의 분위기를 10%, 25%, 50%, 75%, 90%의 5개 분위로 구분하였으며, 각각의 분위는 하위 10%, 하위 25%, 50%(중위), 상위 25%, 상위 10%를 의미한다. 각 분위의 분위 회귀분석 결과, 5% 유의수준에서 유의하지 않은 변수가 일부 나타났으나, 이를 감안하여 설명변수의 분위별 영향력의 차이를 살펴보면 다음과 같다.

대출자 특성에 대한 분위 회귀분석 결과, 나이는 가장 낮은 분위(하위 10%)에서 계수의 절대값이 가장 크게 나타난 반면, 성별, 직업유무, 연소득은 중위 수준에서 영향이 가장 크게 나타났다. 다른 변수들과 달리 나이는 대출금액이 적어질수록 나이의 증가로 인한 대출감소 효과가 크게 나타났는데, 이는 고령자의 경우 상대적으로 소득이 적거나 1~2인 가구일 확률이 높아 대출금액이 적을 때 다른 연령대에 비해 상대적으로 더 적은 대출을 받는 것을 알 수 있다. 금리 우대조건의 경우 하위 10%나

<표 5> 모형 분석 결과(자산심사 도입 전)

구분		OLS	Quantile Regression				
			10%	25%	50%	75%	90%
상수항		14.4694** (3.38)	11.1927 (1.41)	11.0065* (2.03)	5.2904 (1.03)	3.2486 (0.65)	18.2813** (4.88)
대출자 특성	나이	-0.5915** (-10.47)	-0.6106** (-4.86)	-0.5467** (-8.03)	-0.5774** (-7.90)	-0.3706** (-5.78)	-0.2929** (-5.79)
	성별	-3.2633** (-3.27)	-1.9155 (-1.19)	-2.0403 (-1.79)	-3.2351* (-2.33)	-3.2279** (-3.51)	-1.8659 (-1.79)
	직업유무	8.5040** (5.23)	8.2771** (2.85)	11.4063** (4.40)	9.3726** (3.47)	7.6733** (3.91)	7.4407** (6.72)
	연소득	0.3654** (11.98)	0.3284** (5.42)	0.3662** (7.00)	0.3913** (10.38)	0.3054** (7.10)	0.2513** (5.76)

〈표 5〉 계속

구분		OLS	Quantile regression				
			10%	25%	50%	75%	90%
금리 우대조건	신혼부부	6.2504** (4.09)	6.0949 (1.80)	8.0497** (2.99)	6.3103** (3.43)	5.4642** (4.96)	5.5282** (4.30)
	다자녀	-2.0412** (-3.67)	-4.7578** (-4.49)	-2.7388** (-3.72)	-3.0247** (-4.25)	0.2024 (0.34)	1.8371** (3.24)
	기타	2.1645 (1.78)	2.0603 (0.85)	3.0509 (1.74)	2.9429* (2.59)	0.5184 (0.49)	0.1995 (0.16)
담보물 특성	수도권	4.1501** (3.77)	4.1177 (1.69)	1.0779 (0.55)	1.7775 (1.11)	2.8011* (2.37)	6.1392** (4.90)
	아파트	-7.5109** (-5.65)	-8.8697** (-4.24)	-7.9801** (-3.97)	-7.5836** (-4.67)	-2.7182* (-2.09)	1.5146 (1.10)
	전용면적	0.1274** (3.43)	-0.0996 (-1.20)	-0.1454** (-3.04)	0.1359** (3.05)	0.1882** (6.31)	0.2000** (4.51)
	주택가격	0.2835** (49.43)	0.1837** (12.96)	0.2644** (37.45)	0.3330** (50.31)	0.4170** (59.82)	0.3921** (33.80)
대출특성 (대출기간)	15년	19.8166** (10.22)	12.0873** (4.03)	13.6819** (6.81)	19.1472** (10.05)	17.6456** (12.26)	17.5984** (7.93)
	20년	28.5528** (16.43)	24.2003** (8.80)	25.1244** (12.01)	27.7622** (14.96)	22.9988** (12.25)	20.7175** (10.43)
	30년	45.2639** (30.18)	44.5708** (17.14)	43.9720** (29.42)	45.0405** (31.42)	34.3148** (25.76)	28.9590** (15.43)
모형설명력		0.4922	0.2199	0.2842	0.3263	0.3766	0.3836

주: 1) 모형설명력은 OLS에서는 Adjusted R^2 , 분위회귀분석에서는 Pseudo R^2 를 의미.

2) *: 유의수준 5%에서 통계적으로 유의($p < 0.05$).

** : 유의수준 1%에서 통계적으로 유의($p < 0.01$).

3) ()는 t 값을 의미하며, 진하게 표시된 부분은 계수값의 절대치가 가장 큰 분위를 의미.

하위 25%의 계수의 절대값이 가장 높게 나타났다. 신혼부부나 기타 우대조건인 경우 하위 25%에서 계수값이 가장 높았는데, 이는 대출금액 분위가 25%에서 우대에 따른 대출증대 효과가 가장 큰 것을 의미한다. 다자녀는 특이한 특징을 보였는데, 대출금액이 적은 하위 분위에서는 음(-)의 값을 가졌지만, 상위 분위에서는 양(+)의 값으로 변환되어 대출금액이 많을 때에는 다자녀일수록 대출금액을 증가시키는 것을 알 수 있었다. 즉, 대출금액이 적을 때에는 자녀 수가 많을수록 대출이

줄어들지만, 대출금액이 많은 상위 분위일 때에는 자녀수가 많을수록 대출을 증가시킴을 의미한다. 담보물 특성은 아파트(하위 10%)를 제외하면 대출금액이 상위 분위에 있을 때 변수의 영향력이 큰 것으로 나타났다. 수도권, 전용면적, 주택가격 등의 담보물 특성은 대출금액이 높을 때 해당 설명변수의 변화에 따른 대출금액의 증가가 크게 나타나는 것을 의미한다. 하지만 아파트는 대출금액이 적은 하위 분위에서는 아파트일수록 대출금액이 줄어들었으나, 반대로 대출금액을 높을 때에는 양(+)의 부호값을 가져 아파트가 오히려 대출금액을 증가시키는 변수로 나타났다. 다만 변수의 유의도는 낮게 나타났다. 대출기간은 15년, 20년, 30년 모두 중위(50%) 분위일 때 영향도가 가장 큰 것으로 나타났다. 전반적으로 대출자 특성과 대출기간은 중위 분위, 금리 우대조건은 하위 분위, 담보물 특성은 상위 분위에서 설명변수 변화에 따른 대출금액 증가효과가 가장 크고 민감한 것으로 분석되었다.

〈표 6〉은 자산심사 도입 이후의 OLS와 분위 회귀모형의 분석 결과를 나타낸다. 대출자 특성에서 나이, 성별은 하위 10%에서, 직업유무, 연소득은 하위 25%에서 가장 계수의 절대값이 커서, 대출금액 하위 분위에서 각 변수의 변화에 따른 대출금액 민감도가 크게 나타났다. 금리 우대조건은 신흥부부,

〈표 6〉 모형 분석 결과(자산심사 도입 후)

구분		OLS	Quantile regression				
			10%	25%	50%	75%	90%
상수항		-2.8872 (-0.62)	-14.0574* (-2.45)	-16.4367** (-2.94)	-20.2463** (-4.76)	-8.7014** (-2.75)	2.5252 (0.85)
대출자 특성	나이	-0.4335** (-7.43)	-0.4196** (-4.68)	-0.4067** (-5.92)	-0.3055** (-5.62)	-0.2161** (-4.83)	-0.1604** (-3.78)
	성별	-2.4024* (-2.28)	-2.0945 (-1.18)	-1.4361 (-0.82)	-1.6200 (-1.21)	-1.3409 (-1.50)	-1.4372 (-1.78)
	직업유무	11.1579** (6.13)	12.2492** (5.77)	15.2005** (6.13)	11.4829** (4.70)	7.5558** (5.70)	3.7911* (2.50)
	연소득	0.2715** (8.18)	0.2409** (4.77)	0.2940** (4.92)	0.2418** (7.57)	0.1622** (9.16)	0.0809* (2.55)
금리 우대조건	신흥부부	14.1143** (8.39)	19.3920** (7.10)	17.4244** (12.11)	12.9381** (5.99)	6.3182** (4.07)	3.3385* (2.53)
	다자녀	-1.6917** (-2.88)	-3.4058** (-3.09)	-4.0715** (-3.05)	-0.2958 (-0.36)	-0.1503 (-0.41)	-0.0464 (-0.09)
	기타	4.9452** (3.60)	8.5099** (3.14)	6.6727** (4.16)	4.6353* (2.40)	1.1763 (0.86)	0.5811 (0.58)

〈표 6〉 계속

구분		OLS	Quantile regression				
			10%	25%	50%	75%	90%
담보물 특성	수도권	3.0178* (2.56)	6.2329** (3.68)	0.9533 (0.47)	1.3537 (0.90)	2.0631* (2.40)	3.2055** (4.64)
	아파트	-10.7733** (-8.13)	-14.1166** (-10.37)	-14.3093** (-7.90)	-6.0094** (-4.34)	-0.4371 (-0.40)	1.4968 (1.25)
	전용면적	0.1525** (4.14)	0.0253 (0.35)	0.0319 (0.58)	0.1473** (3.54)	0.1986** (7.12)	0.1693** (5.39)
	주택가격	0.3528** (56.43)	0.2255** (19.60)	0.3217** (36.70)	0.4276** (59.56)	0.5135** (75.55)	0.5621** (34.49)
대출특성 (대출기간)	15년	18.2603** (8.21)	12.8601** (3.24)	18.7697** (4.91)	22.3686** (7.99)	11.6865** (7.08)	6.3346** (3.11)
	20년	23.1828** (12.11)	17.2442** (8.34)	25.2953** (9.65)	24.8853** (13.59)	13.8540** (8.22)	8.5026** (5.02)
	30년	40.3204** (24.27)	39.3094** (24.12)	45.2630** (13.36)	38.5015** (20.07)	22.0199** (15.78)	13.6092** (5.85)
모형설명력		0.5589	0.2245	0.3104	0.4017	0.4645	0.4634

주: 1) 모형설명력은 OLS에서는 Adjusted R^2 , 분위회귀분석에서는 Pseudo R^2 를 의미.

2) *: 유의수준 5%에서 통계적으로 유의($p < 0.05$).

** : 유의수준 1%에서 통계적으로 유의($p < 0.01$).

3) ()는 t 값을 의미하며, 진하게 표시된 부분은 계수값의 절대치가 가장 큰 분위를 의미.

기타 우대조건 하위 10%에서, 다자녀는 하위 25%에서 계수의 절대값이 크게 나타나, 대출자 특성과 마찬가지로 대출금액이 적을 때 변수의 영향력이 큰 것으로 분석되었다. 특히 신혼부부는 대출금액 하위 분위의 계수값이 가장 크고, 상위 분위와의 차이도 크게 두드러져 하위 분위에서의 정책 지원 효과가 크게 나타나는 것을 알 수 있다.

담보물 특성은 수도권은 하위 10%, 아파트는 하위 25%, 전용면적은 상위 25%, 주택가격은 상위 10%에서 계수의 절대값이 가장 높아 설명변수에 따른 영향력이 가장 큰 것으로 분석되었다. 담보물 특성은 다른 설명변수 범주와 다르게 종속변수인 대출금액의 상위 분위에서 영향력이 큰 것을 알 수 있었다. 대출기간 범주 중 15년은 중위(50%) 분위에서, 20년과 30년은 하위 25% 분위에서 가장 계수값이 크게 나타났는데, 자산심사 도입 전의 모두 중위(50%) 분위에서 계수값이 가장 컸던 것과 다소 차이를 보였다.

분위 회귀분석의 결과를 자산심사 도입 전과 후를 비교하여 살펴볼 때, 공통적인 특징은 대출자

특성, 금리 우대조건 등 대출 대상자와 관련되는 설명변수는 상대적으로 중·하위 분위에서 계수의 절대값이 커서 영향력이 크게 나타났으나, 담보물 특성은 상위 분위의 영향력이 크게 나타났다는 점이다. 즉, 대출금액이 적을 때 대출 대상자의 특성이 대출 증가에 영향을 크게 주고, 대출금액이 많을 때에는 담보물 특성이 대출의 증가에 상대적으로 큰 영향을 주게 된다. 자산심사 도입 전과 후의 차이는 전반적으로 도입 전보다 도입한 이후에 계수의 절대값이 상대적으로 대출금액의 낮은 분위로 옮겨간 경우가 늘었다는 점이다. 즉, 대출금액이 많을 때보다 대출금액이 적을 때 각 특성 요인에 따른 대출 증대효과가 크게 증가하였다는 사실이다. 이는 자산심사 도입으로 대출 대상자에 대한 심사가 강화되어, 대출금액을 적게 받는 하위 분위 계층일수록 제도의 효과가 보다 큰 영향을 미치는 것으로 이해할 수 있다.

V. 결론

본 연구는 2019년 이후의 대출자료를 이용하여 주택도시기금의 주택 구입자금인 내집마련 디딤돌 대출의 자산심사제도 도입 전후의 대출특성 변화를 살펴보았다. 종전에는 기금대출에서 대출자의 자격 기준을 소득에 한정하여 가능 여부를 따졌다면, 자산심사 제도를 도입한 이후에는 소득 기준 외에 순자산 기준을 추가로 살펴봄으로써, 보다 지원이 필요한 대상에게 대출을 실행하도록 개선하였다. 하지만 이러한 자산심사 제도의 도입이 기존 대출 대상자에게는 피해를 주지 않았는지, 실제로 기대한 효과를 보았는지에 대한 사후검증은 아직 제대로 이루어지지 않았다. 이에 본 연구에서는 분위 회귀분석을 사용하여 자산심사 제도 도입 전후의 모형을 구분하여 설정하고, 각각의 설명변수의 영향을 비교함으로써 대출특성의 변화 및 자산심사 제도의 효과를 살펴보았다.

먼저 첫 번째 결과는, 자산심사 제도 전후의 대출특성이 크게 바뀌지는 않았다는 점이다. 이는 기존의 제도 하에 디딤돌 대출을 받을 수 있는 대상자는 대부분 자산심사제도 도입 이후에도 대출이 가능하였음을 의미한다. 일부 소득은 적지만 순자산이 많은 부적격 차주만이 자산심사 도입으로 대출 대상에서 제외되었을 뿐 대부분의 대상자는 여전히 디딤돌 대출을 받을 수 있는 것으로 해석할 수 있을 듯 하다. 만약 자산심사 도입 이후에 대상이 크게 바뀌었다면 제도의 적절성 논란이 발생할 여지가 있었음을 간과할 수 없다. 둘째는, 그럼에도 불구하고 회귀분석에서 알 수 있는 결과는 세부적으로는 대출특성의 변화가 발생하였다는 점이다. 제도 이후에 생애최초 신혼부부, 기타 우대조건 및 직업유무 등의 계수 영향력이 증대하였다. 이는 정부에서 정책적으로 지원하고자 하는 계층에서 영향력이 증가하여 자산심사 제도가 효과를 일부 보이고 있는 것으로 유추된다. 또한 주택가격 상승 등 주택시장의 강세로 담보물 특성이 대출 증가에 미치는 영향 역시 늘어난 것을 알

수 있었다.

마지막으로, 분위 회귀분석에서는 자산심사 도입 이후 종속변수인 대출금액이 하위 분위에서 설명변수의 영향력이 크게 나타나는 경우가 늘었다는 점이다. 특히 대출자 특성 및 금리 우대조건 등 대출자 관련 설명변수에서 이러한 현상이 나타나 상대적으로 대출금액이 적은 하위 분위에서 민감도가 확대된 것으로 판단된다. 즉, 자산심사 제도의 도입으로 대출금액이 상대적으로 적은, 상대적으로 기금 지원이 필요할 수 있는 하위 계층에 효과가 늘어난 것일 수 있겠다.

하지만 이러한 효과를 단정하기에는 모형이 갖는 한계가 있다. 자료의 한계상 대출액이 증가할 때 지원효과가 확대되었다고 가정하였으나, 정책적 지원효과가 반드시 대출액 증가만을 가져오지 않을 수 있다. 또한 자산심사 제도 도입 전후의 상황을 완전하게 통제하지 못했으며, 모형을 별도로 설정하는 데 따른 모형 설명력의 차이로 단순히 계수값 비교만으로는 효과를 정확히 파악하기 어렵다는 점이다. 도입 전후의 변수를 통일시켜 객관적인 비교를 하고자 하였으나, 향후 두 시기의 차이를 하나의 모형으로 파악하는 분석 방법론 등을 활용한 추가적인 연구가 필요할 것으로 판단된다. 두 번째의 한계는 대출자료 자체가 가지는 한계이다. 누락된 특성이 많아 상당수의 자료를 제외하고 일부 이상치도 제외하였으나 여전히 오류값들을 내포하고 있을 것으로 예상되며, 일부 내부 시스템 확인을 통해 수정하였으나, 이에 따라 모형의 설명력이 낮아졌을 가능성도 있다. 마지막으로 자산심사 제도 전후의 효과를 간접적으로 살펴보았으나, 자산심사에 대한 통과 유무에 대한 자료가 있다면, 자산심사 제도의 직접적인 효과를 보다 자세히 살펴볼 수 있을 것이란 아쉬움이 남는다. 향후 분석 방법의 개선으로 보다 의미있고 정밀한 연구가 지속 수행될 수 있기를 기대한다.

참고문헌

- 강성덕, 2014, “생애 최초 주택구입자의 주택구매 특성 분석: 공유형 모기지 시범사업을 중심으로”, 연세대학교 공학대학원 석사논문.
- 강성진, 이주석, 2012, “주택금융의 수요자 특성에 관한 연구”, 부동산학연구 18권 3호, 37-50.
- 고성수, 주민균, 2011, “국내 주택금융시장의 모기지 선택에 관한 연구”, 부동산학연구 17권 2호, 59-75.
- 구만수, 서정렬, 최열, 2013, “가구 특성을 고려한 주택구입자금 대출에 관한 분석”, 부동산학보 55권, 5-16.
- 김경훈, 2007, “토빗 회귀모형 추정량들에 관한 비교연구”, 고려대학교 석사논문.
- 김기태, 오동훈, 2015, “공유형 모기지제도의 특성 및 대출수요 영향요인 분석”, 부동산학연구 21권 4호, 51-63.
- 김성진, 2016, 주택도시기금 수요자대출 차주의 대출 결정요인 분석, 연세대학교 박사논문.
- 김정엽, 이현석, 2016, “주택담보대출 상환방식 선택에 따른 차입자 특성분석”, 부동산학연구 22권 1호, 77-88.
- 김희호, 박세운, 2012, “서울 주택가격의 결정요인: 분위수 회귀분석”, 주택연구 21권 2호, 141-168.
- 민인식, 최필선, 2015a, STATA 기초적 이해와 활용, 지필미디어.
- 민인식, 최필선, 2015b, STATA 기초통계와 회귀분석, 지필미디어.
- 박범조, 2003, “분위수 회귀접근법”, 계량경제학보 14권 4호, 93-122.
- 심기우, 민규식, 김행조, 2015, “공유형모기지 상품에 대한 이론적 고찰과 선택 결정요인 분석에 관한 연구”, 대한부동산학회지 33권 2호, 23-42.
- 윤수민, 유선종, 2016, “아파트 담보대출자의 금리선택에 관한 실증분석”, 주택연구 24권 2호, 155-175.
- 이대식, 이찬호, 신형수, 2011, “우리나라 주택담보대출 행태에 관한 실증연구: A광역시 보금자리론 차입자를 중심으로”, 경제연구 29권 4호, 219-245.
- 이수진, 조주현, 2017, “주택담보대출의 선택 요인 분석: 수(손)의 공유형 모기지 및 내집마련디딤돌대출을 중심으로”, 주택연구 25권 2호, 71-93.
- 이정영, 진창하, 2015, “주택담보대출 보유가구의 특성이 CLTV에 미치는 영향 분석”, 주택연구 23권 3호, 63-87.

정의철, 권혜진, 2006, “도시가구의 주택자금대출 상환방식 선택에 관한 실증 연구”, 국토계획 41권 5호, 105-116.

주택도시보증공사 홈페이지. <http://nhuf.molit.go.kr/>

Al-Bahrani, A. and Su, Q., 2015, “Determinants of mortgage pricing: A quantile regression analysis”, Journal of Housing Economics, 30, 77-85.

Black, H., Boehm, T. and DeGennaro, R., 2003, “Is there discrimination in mortgage pricing? The case of overages”, Journal of Banking and Finance, 27, 1139-1165.

Boehm, T., Thistle, P. and Schlottmann, A., 2006, “Rates and race: An analysis of racial disparities in mortgage rates”, Housing Policy Debate, 17(1), 109-149.

Coulibaly, B. and Li, G., 2009, “Choice of mortgage contracts: Evidence from the survey of consumer finances”, Real Estate Economics, 37(4), 659-673.

Courchane, M., 2007, “The pricing of home mortgage loans to minority borrowers: How much of the APR differential can we explain?”, Journal of Real Estate Research, 29(4), 399-439.

Koenker, R. and Hallock, K., 2001, “Quantile regression”, Journal of Economic Perspectives, 15(4), 143-156.

(논문 접수일: 2020.10.31. 수정논문 접수일: 2020.12.01. 논문 채택일: 2020.12.11.)

An Analysis of the Characteristics of Didimdol Loans with the Introduction of the Asset Review System

Sungjin Kim*

Abstract

The loan borrowers must meet the net asset standard below a certain level from September 2019, in order to get housing purchases and lease loans from the National Housing and Urban Fund(NHUF). This is aimed at preventing the waste of funds by excluding unqualified individuals with large assets but small incomes. It has been more than a year since the asset review system was introduced, so it is now necessary to check the impact of the system. In this study, the quantile regression analysis was conducted to analyze the changing characteristics of 'didimdol loans' of NHUF before and after the introduction of the asset review system. Through this analysis, we looked at the adequacy and effectiveness of the asset review system. First of all, the characteristics of borrowers have not changed much. This means that while maintaining the existing system, it had the effect of limiting only a few unqualified people. Since the introduction of asset review system, the influence of characteristics of borrowers and preferential interest rate conditions has increased. It was also noted that the impact of the lower divisions was relatively greater. However, this effect is limited to examining the indirect effects of the asset review system. Therefore, it is necessary to obtain borrowers' asset data in the future for direct and better analysis.

Keywords : National Housing and Urban Fund(NHUF), Didimdol Loan, Asset Review System, Charateristic Factors, Quantile Regression

* Sungjin Kim, Corresponding author, Wooribank, Research Fellow, sjkim2210@wooribank.com

© Copyright 2020 Housing Finance Research Institute. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

주택특성 및 가구특성이 주택구입자금 부채비율에 미치는 영향¹⁾

이성원*

요 약

본 연구에서는 주택의 투자수요라는 관점에서 적응적 기대가설에 근거하여 과거의 가격 경험을 통하여 앞으로 가격의 상승 또는 하락이 예상되는 지역에 주택을 구입함에 있어 사람들의 주택구입 자본구조가 어떻게 다를 것인가에 대해서 밝혀보고자 하였다. 주거실태조사와 한국감정원 평균주택가격상승률, CD금리 데이터를 통합하여 주택구입 자금(PLTV)을 종속변수로 설정하고, 소득, 순자산, 가구원수 등의 가구특성과 주택규모 등 주택특성을 설명변수로 하는 토빗모형을 설계하여 투자수요에 따른 주택구입자금 부채비율에 미치는 영향을 분석하였다. 연구의 결과, 기대 가격 상승률이 양수인 지역과 음수인 지역에서의 결과가 비대칭적으로 나타났는데, 기대 가격 상승률이 양수인 지역에서는 과거 평균 가격상승률이 높을수록 주택 구입자금에서 타인자본비율이 높은 것으로 나타났으며, 기대 가격 상승률이 음수인 지역에서는 둘 사이에 통계적 유의성을 찾기 어려웠다. 연구는 주택이 입지하는 지역의 지역특성이 대출 조건의 고려 요소로서 추가될 수 있음을 시사한다. 이는 가격 상승 기대감이 특정 지역의 주택 소비에 있어서 타인자본 수요를 불러오고, 그 결과 지역에 타인자본 비율이 집중되어 사회적 자본 건전성을 위협할 수 있기 때문이다.

핵심어 : 주택구입자금, 자본구조, 부채비율, 주택투자, 가구특성, 주택특성

1. 서론

주택은 일반 소비재와는 다르게 투자재적 특성을 함께 지니고 있다. 즉, 주택의 가치를 구성하는

1) 본 논문은 저자의 2016년 석사학위 논문 “주택구입자금 부채비율의 영향요인에 대한 실증분석”을 바탕으로 작성되었음.

* 이성원, 주저자, 한국주택금융공사 주택금융연구원 대리, 1944@hf.go.kr

© Copyright 2020 Housing Finance Research Institute. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

요소에는 주택에서 거주하면서 얻을 수 있는 소비적 가치(usage value)와 함께 주택의 가격 상승을 통한 자본이득(capital gain)을 취하려는 투자 가치(investment value) 또한 존재한다. 이러한 주택의 투자재적 속성은 위치적 고정성으로 인한 외부효과에 일편 기인하는데, 주거 지역을 설정하는 것은 직장이나 학교의 근접성 등 삶의 직접적 필요성에 의하기도 하지만, 주택 입지 지역의 긍정적 외부효과에 따른 주택의 가격변화를 기대·반영하기도 한다. 정리하면, 주택을 구입하는데 있어서 소비자는 그 주택의 실제적 필요성(소비 가치)과 함께 잠재적 투자 가치 또한 동시에 고려하여 의사 결정을 하게 되며, 그 결정의 주요 요인이 되는 것은 해당 주택이 위치하는 지역에 따른 영향이라고 할 수 있다.

한편, 주택을 이러한 투자재적 관점에서 바라본다면, 주택구입자금의 자본구조에 대해서 고려해 볼 필요가 존재한다. 주택소비자는 주택의 구입 의사결정과 함께 주택구입자금의 자본구조를 어떻게 설정할지에 대한 의사결정을 하게 된다. 자본구조는 자기자본과 타인자본으로서 나누어질 수 있는데, 총 자본규모에서 타인자본이 차지하는 비율을 부채비율로서 그 정도를 파악할 수 있다. 자본은 각각 그에 따른 비용(이자비용)을 수반하게 되는데, 타인자본을 조달하는 경우의 타인자본비용, 자기자본의 경우의 자기자본비용이 그것이다. 만약, 주택 소비자가 앞으로 가격이 오를 것으로 예상되는 지역의 주택 구입 의사결정을 앞둔 상황에 있다고 가정해 보자. 이 경우, 타인자본비용이 예상되는 가격상승률보다 낮아 자본이득이 발생할 것으로 판단된다면, 투자적 관점에서 소비자는 주택 구입자금 자본구조에 있어서 타인자본의 비율을 높여 기대 자본이득을 최대화할 것을 예상할 수 있다.

이에 본 연구에서는 주택의 투자수요라는 관점에서 적응적 기대가설에 근거하여 과거의 가격 경험을 통하여 앞으로 가격의 상승 또는 하락이 예상되는 지역에 주택을 구입함에 있어 사람들의 주택구입자금에서 부채비율이 차지하는 비중이 어떻게 다를 것인가에 대해서 밝혀보고자 한다. 이를 통해 연구는 주택 특성과 개인 특성 요소 외에도, 그 주택이 입지하는 지역 요소가 부채비율 설정에 영향을 주는지를 밝힘으로써 금융수요와 주택가격 간의 관계에 미시적·지역적 영향 고려를 촉구하고자 하였다.

연구를 위해 쓰이는 데이터는 주거실태조사 마이크로데이터와 한국감정원의 전국주택가격 동향조사, 해당 시점의 CD금리의 세 가지 자료를 통합하여 주거실태조사 시점으로부터 4년 이내에 아파트를 구입한 자가 거주가구를 연구의 대상으로 하였다. 주택 구입자금의 부채비율을 종속변수로, 과거 n 년 간 지역 가중 평균 주택 상승률을 가장 중요한 독립변수로 설정하여 토빗 모형을 통한 통계적 연구 방법을 사용하였다.

II. 이론적 고찰 및 선행연구 검토

1. 자본구조

자산을 구성하는 자본구조는 크게 스스로 조달한 자기자본과 타인을 통해 조달한 타인자본으로 나눌 수 있는데, 각각 스스로의 자본에 요구되는 수익률인 자기자본 비용과 타인자본에 대해서 요구하는 타인자본비용의 비용을 수반하게 된다. 이때, 비용을 각각의 자본비율로서 가중평균하게 되면 자산에 대한 가중평균자본비용(WACC, Weighted Average of Cost of Capital)을 구할 수 있으며, 이는 같은 시점, 동일한 자산에 대한 것이라 하더라도 대출 주체에 따라 달라질 수 있다. 가격과 그 자본구조와의 관계에 대해 재무 관리적 관점에서 Modigliani and Miller(1958)는 자산의 가격은 그 자본 구조에 영향을 받지 않는다고 하였으며, 그 이후로 절세 효과에 의해서 차입이 많을수록 유리하다는 수정 MM이론, 한계 비용과 한계 효용이 일치하는 지점까지 차입을 진행하게 된다는 정태이론까지 이어져 온 바 있다. 본 절에서는 가계·주택시장·주택구입자본구조의 순서대로 자본 구조가 그 가격과 어떤 관계를 가지고 있는지와 함께 각각의 논점에 대해서 이론적 배경과 선행연구를 정리·연구하고자 한다.

1) 가계 부채

자산 가격과 가계 부채의 관계는 주로 거시적 측면에서 관찰되어져 왔다. Hofmann, B.(2004)은 산업화 국가 16개국을 대상으로 VAR 모형을 사용, 가계대출과 국내총생산, 부동산가격, 실질이자율의 관계를 관찰하였다. 연구결과, Hofmann, B.은 장기대출에 있어서 자산 가격이 중요한 변수로 작용함을 밝혔으며, 실질이자율이 가계부채를 장기적으로 증가시키는 요인임을 실증분석하였다. 또한, Gerlach and Peng(2005)은 홍콩의 데이터를 사용하여 주거용 부동산 가격과 은행대출 규모와의 관계를 관찰하였으며, 대출 규모와 자산 가격은 서로 관계를 주고받지만, 자산 가격이 대출 규모에 미치는 영향이 더욱 강력한 것을 밝혔다. Oikarinen E.(2009)은 1980년대 후반 이후 핀란드 장기 시계열 자료를 이용하여 주택가격과 대출수요 사이에 관련성이 존재함을 밝혔으며, 실질이자율의 하락과 주택가격의 상승이 가계부채를 증가시키는 요인으로 작용함을 확인하였다. 동시에, 주식자산의 경우에는 주택자산과는 달리 이러한 관계가 나타나지는 않음을 밝혔다. 또한, 장영숙(2006)은 1998년부터 2006년의 시계열자료를 벡터오차수정모형으로 분석하여 한국에서

실질이자율의 하락이 가계부채를 증가시키는 원인임을 밝혔다.

2) 주택시장에서의 자본구조

주택시장 자본구조에 관한 연구는 크게 자본구조와 지역주택가격에 대한 연구, 그리고 주택금융수요에 관한 연구로 나누어 볼 수 있다. 전자는 주택시장에서의 자본구조가 그 지역의 주택 가격에 어떠한 영향을 미치는지에 대한 연구이며, 후자의 주택금융수요에 대한 연구는 거시적 또는 미시적으로 어떠한 특성 또는 변수가 주택금융수요와 연관 관계를 가지는지에 대한 연구이다.

(1) 자본구조와 지역 주택가격

Stein, J. C.(1995)은 자산 제약 하에 있는 가구의 비중이 지역 주택가격에 영향이 미칠 수 있다고 하면서 주택을 소비자의 자본구조에 따라 지역 주택가격이 영향을 받을 수 있음을 밝혔다. Mian and Sufi(2011)는 차입이 많은 지역의 주택가격 변동성이 그렇지 않은 지역보다 상대적으로 더 높은 것을 밝혔다. 또한, Mian et al.(2015)은 그러한 가격 변동성이 현실화되는 기제로서 담보권(Foreclosure) 행사로 인한 경매(Forced Sale)를 꼽았는데, 2004년부터 2006년의 미국의 데이터를 사용하여 경매에 재판을 요구하는 주(State)와 그렇지 않은 주 사이에서, 재판을 요구하지 않아 담보권 행사가 쉬운 주의 지역 주택가격이 더욱 크게 하락하는 것을 확인하였다. Cambell et al.(2009)과 Anenburg and Kung(2013) 또한 이러한 담보권 행사로 인한 경매가 지역 주택가격에 미치는 영향에 대해서 실증적으로 분석한 바 있다. 유승동, 신승우(2012)는 미국의 데이터를 사용하여 주택대출 선택이 주택가격 변동에 미치는 영향에 대해서 회귀분석을 통해 실증분석하였다.

(2) 주택금융수요

주택금융수요에 관한 연구는 크게 실질이자율 등 금융적 요인으로 인한 주택금융 수요 변화에 관한 연구와, 가구 특성 및 주택 특성으로 인한 자본구조 변화에 관한 연구로 나눌 수 있다.

Brueckner, K.(1994)는 주택금융수요는 예금규모와 주택가격 사이에서 결정되는데, 대출 금리와 투자 수익률은 이에 중요한 변수로 작용한다고 하였다. 이에, 주택금융 대출 금리와 기타 자산 수익률의 차이가 주택금융수요에 영향을 준다는 가정하에 가계의 주택소비량, 주택금융수요 결정의 효용극대화 모형을 실증 분석하였다. Follain, J. R. (1990)은 LTV는 나이, 거주기간, 소득과

관계가 있으며, 가구주의 나이가 어릴수록, 거주기간이 짧을수록, 소득이 높을수록 타인자본의 비율의 증가함을 밝혔다. Gimeno and Carrascal(2010)은 스페인의 데이터를 사용하여 주택 구입자금 대출과 주택가격, 명목대출이자율, 실질소득 간에는 거시적 상관성이 있음을 VECM (Vector Error-Correction Model)를 통하여 밝혔으며, 특히 주택구입자금대출과 주택가격 간에는 정의 상관관계가, 명목금리와는 부의 상관관계가 있음을 실증 분석하였다.

한편, 국내 주택시장을 대상으로 최막중 외(2001)는 주택 금융제약이 주택 소비규모와 점유형태 선택에 미치는 영향에 대해서 금융제약, 주택가격, 항상소득, 사용자비용, 가구원 수, 직업, 지역을 변수로 분석하였다. 연구에서는 주택자금의 최적 용자비율과 이에 따른 최적 주택구입가격 결정 모형에 기초하여, 주택금융제약은 주택수요를 제한하는 요인으로 작용하고 있음을 밝혔다. 정의철(2005)은 구조모형을 통해 주택금융수요를 추정하였는데, 주택금융수요는 금리차이, 주택수요량, 보유자산, 주거이동성의 영향을 받는다고 하였다. 김승욱, 남영우(2012)는 가구특성에 따른 LTV변화를 회귀분석을 통하여 분석하였는데, 가구주 연령이 어릴수록, 가구 총자산이 낮을수록, 주택 면적이 작을수록 타인자본비율이 증가함을 실증 분석하였다.

김도훈, 김진수(2012)는 중국 베이징의 도시가구를 대상으로 가구특성에 따라 주택담보대출 여부를 로짓회귀모형을 통해 분석하였는데, 30대, 소득 월 1만 위안 이상, 주택 구매가격 2백만 위안 이상일 경우 주택담보대출을 가지고 있을 확률이 높음을 확인하였다.

3) 주택 구입자금 자본구조

앞서의 연구들을 살펴볼 때, LTV와 같은 주택의 자본구조에 대한 연구는 국내, 국외에서 상당 부분 진행된 것을 확인할 수 있다. 특히, 금융위기 이후로 주택자본구조가 지역 주택가격에 미치는 영향에 대한 연구가 활발히 진행되었다. 또한, 자본구조 연구에 있어서 앞서의 연구들에서 주요한 변수로 사용된 것들은 실질이자율과 같은 거시적 변수와 함께 가구특성, 주택특성임을 확인할 수 있었다. 하지만, 주택구입자금의 자본구조에 대해서 면밀히 분석한 자료는 부재한 것으로 판단된다. 따라서 본 연구에서는 주택구입자금의 자본구조에서 타인자본의 비율을 PLTV(Purchase Loan to Value)로서 정의하고 연구를 진행하고자 한다.

주택구입시점에서의 자본구조 선택(PLTV)과 주택을 통한 대출비율(LTV; Loan to Value)은 그 시점에 있어서 차이가 존재한다. 연구에서 사용된 PLTV는 주택 구입 당시, 주택 구입의 목적을 위해서 소비자가 선택한 자본구조를 나타낸다. 한편, LTV는 이에 더하여 주택구입 이후에 변화·형성된 자본구조 또한 포함하는 것으로, 주택 구입이라는 목적 외의 다른 목적을 위해서 선택한 타인자본비율

설정 또한 그 범위에 포함되고 있다. 따라서 주택 구입 당시에 소비자의 의사결정과 그에 영향을 미치는 관계를 규명하기 위해서는 일정 시점의 주택가격에서 그 타인자본비율이 차지하는 비중(LTV)이 아닌 주택구입시점에서의 주택가격에서 그 당시 설정한 타인자본비율(PLTV)을 통해 연구하는 것이 더욱 적확하다고 할 수 있다.

자기자본과 타인자본의 비중과 그 비용을 통해 전체 자본비용을 계산하는 WACC(Weighted Average of Cost of Capital)는 다음과 같이 나타낼 수 있는데, 이를 주택구입자금 자본구조에 반영하면 다음과 같이 해석할 수 있다.

$$\begin{aligned}
 W_E \times R_E + PLTV \times R_D &= WACC \sim E(R_H) \\
 PLTV \times R_D + (1 - PLTV) \times R_E &\sim E(R_H) \\
 PLTV &\sim \frac{E(R_H) - R_E}{R_D - R_E}
 \end{aligned}
 \tag{식 1}$$

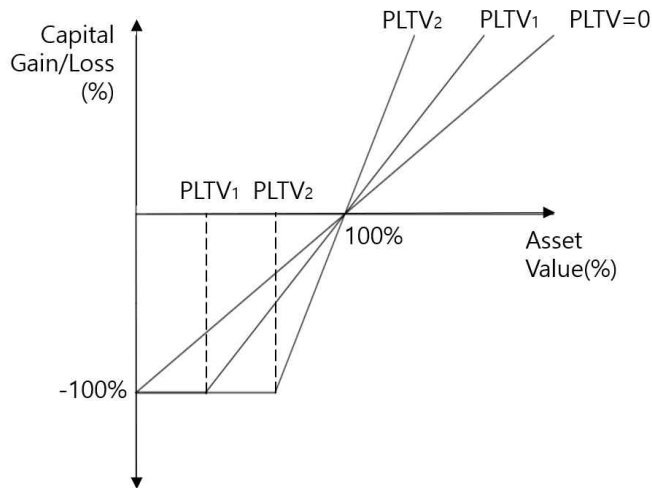
W_E : 자기자본비중
 $PLTV$: 타인자본비중
 R_E : 자기자본비용
 R_D : 타인자본비용
 $WACC$: 가중평균 자본비용
 $E(R_H)$: 기대되는 수익률

만약 주택구입을 통한 기대 자본이득(기대수익률, $E(R_H)$)이 존재하며, 이것이 타인자본비용보다 높아 타인자본 조달을 통해 자본이득을 증가시킬 수 있다면 타인자본비중을 늘리는 방향으로 주택구입자금 자본구조가 변화할 것이다. 즉, 자본이득이 기대될 경우, 주택구입자금에 있어서 타인자본의 비중(PLTV)은 주택으로부터의 기대 수익률에 비례하고, 타인자본비용에 반비례하는 관계를 가질 것이라고 예상할 수 있다.

하지만, 반대로 타인자본비용이 예상되는 수익률보다 높아 자본이득이 발생하지 않을 것이라고 예상된다고 할 때는, 투자적 관점에서 소비자는 주택 구입자금 자본구조에 있어서 타인자본의 비율을 줄일 것이라고 생각할 수는 없다. 이는 주택 가격이 하락할 것이라고 예상되는 경우, 주택구입을 통한 투자적 수요는 존재하기 힘들기 때문이다.

$$\frac{\Delta Equity}{Equity} = \frac{1}{1 - PLTV} \times \frac{\Delta Asset}{Asset}
 \tag{식 2}$$

타인자본을 설정의 효과는 <그림 1>과 같이 나타낼 수 있다. 타인자본비율이 0인 경우에는, 자산가격의 변화분과 자본의 변화분은 동일하다. 즉, 자산 가격이 $x\%$ 변화하면, 순자본 또한 $x\%$ 만큼 변동하게 된다. 하지만, 타인자본비율이 존재하는 경우에는 순자본의 변화량은 자본가격의 변화량에 ' $1/(1-PLTV)$ '만큼 변화하게 되는데, 여기에서 ' $1/(1-PLTV)$ '는 레버리지 효과를 나타내며, 레버리지도 (leverage multiplier)로도 불린다. 예를 들면, 타인자본비율이 50%인 경우에는 자산의 가격이 $x\%$ 만큼 변화하면 그에 따른 순자본은 그 두 배인 $2x\%$ 만큼 변화하게 된다. 이렇듯 타인자본의 설정은 자기자본수익률의 변동성을 증가시키는 방향으로 작용하는데, 따라서 주택가격의 상승을 예상하여 주택구입 자본구조에 있어서 타인자본비율을 증가시키는 것은 그 수익률을 단시간에 작은 자산 가격 변화를 통해 더욱 큰 자본 차익을 발생시키거나, 혹은 하락 상황의 손실을 더욱 확대하는 결과를 가져오게 된다.



<그림 1> 주택 자산의 가격 변화에 따른 순자본 이익/손실을

2. 기대가설

경제주체의 의사 결정은 어떠한 변수의 미래 값에 대한 기대(expectation)에 따라 이루어지게 된다. 본 소절에서는 20년대 후반 거시경제학의 기대형성이론(theory of expectation)을 살펴보고, 연구에서 살펴보고자 하는 주택시장에서는 이러한 기대가설 이론이 어떻게 적용·연구되어 왔는지와 함께 각각의 논점에 대해서 이론적 배경과 선행연구를 정리·연구하고자 한다.

1) 기대가설 이론

기대가설에 대한 이론은 거시경제학에서 완전예견가설, 정태적 기대가설, 적응적 기대가설, 합리적 기대가설로 발전해왔다. 완전예견가설(perfect foresight)은 고전학파의 가설로서 경제는 완전히 예측 가능하여 기대치와 실제치가 항상 동일한 것을 의미하며, 고전학파가 사용하였으나 지나치게 비현실적인 가정이다.

정태적 기대가설(fixed expectation)은 미래에 나타날 기댓값이 그것의 현재 값과 같을 것으로 예상하는 가설로서 고정된 기대라고 불리기도 한다. 정태적 기대가설에 따르면 현재의 상황이 미래에도 지속될 것으로 예상하므로 예측의 정확성이 낮고, 경제 상황의 변화를 상당 기간이 지난 후에도 이를 인지하지 못하거나 인지하더라도 그를 예측에 반영하지 못하므로 비합리적이다. 정태적 기대가설은 기대를 경제적 행위에 영향을 미치는 주요한 변수로서 도입했다는 것에 의의를 가지지만, 많은 변수들의 상호관계를 가지고 동태적으로 변화하는 현실 경제에서 현실적으로 그 설명력은 떨어진다고 할 수 있다.

Cagan(1956)에 의해서 체계화, 케인즈 학파와 통화주의 학파가 주로 사용해 온 적응적 기대가설(adaptive expectation)은 일종의 내생적 기대이론으로서 각각의 경제 주체는 현재 시점의 기대를 수정함에 있어서, 과거의 예상오차 정보를 반영한다는 가설이다. 이는 기대를 합에 있어서 $(t-1)$ 기까지의 정보를 가중 평균하여 t 기의 값을 예상함을 의미한다고 할 수 있는데, 이를 수식으로 표현하면 다음과 같다.

$$P_t^e = P_{t-1}^e + \alpha(P_{t-1} - P_{t-1}^e), (0 < \alpha < 1) \quad \langle \text{식 3} \rangle$$

적응적 기대이론에 따르면, 경제 주체들은 상황을 정확하게 파악하지 못하는 단기적인 오류를 범하지만, 장기적으로는 정확히 예측하게 된다. 하지만 적응적 기대는 기대 형성 방식이 고정되어 있다고 가정함으로써 체계적 오류를 범하게 되고, 조정계수 α 의 크기가 어떻게 결정되어야 하는지를 설명하지 못하는 한계가 존재한다. 또한, 급속한 인플레이션 등의 상황에서는 실제적으로 경제주체들의 물가예상이 비교적 정확·신속하게 조정되나, 적응적 기대가설은 그러한 상황을 설명할 수 없는 단점이 있다.

마지막으로, 합리적 기대가설(rational expectation)은 개별 경제주체들이 모든 정보를 효율적으로 이용하여 가능한 최선의 기대를 형성한다는 것으로 Muth(1961)에 의해서 주창되어 새고전학파 및 새케인즈학파가 주로 사용한다.

$$P_t^e = E(P_t | I_{t-1}), (I_{t-1} : (t-1)\text{기까지 이용가능한 모든 정보}) \quad \langle \text{식 4} \rangle$$

$$= P_t + \epsilon_t$$

합리적 기대가설은 경제학의 기본가정인 합리성의 공준을 불확실성하의 의사결정에 적용한 것으로, 이 학설에 의하면 사람들은 의사 결정 당시 수집할 수 있는 최선의 정보에 의거하여 미래에 대한 예측을 행하고, 그러한 예측을 바탕으로 실제 행동한다고 가정한다. 이때, 예측은 정확할 것이 요구되지는 않는 것이며, 따라서 예측오차가 불가피하여 예측 가능하지 않다는 성질을 가정한다. 합리적 기대를 사용해서 예측할 경우, 예측오류는 발생하게 되나, 이는 평균적으로는 0이 되고 체계적 오류가 발생하지 않는다.

한편, Bikhchandani and Sharma(2000)에 따르면 본 연구에서 가정하고 있는 적응적 기대가설, 과거 가격 경험을 바탕으로 투자 전략을 구성하는 순행투자(momentum investment)는 이성적 군집행동(rational herd behaviour)의 한 종류라고 하였으며, 이는 가격과 그것의 변동성 움직임을 크게 만든다고 하였다.²⁾ 또한, 이러한 이성적 군집행동이 나타나는 원인으로는 정보(information-based), 평판(reputation-based), 그리고 보상체계(compensation-based)의 세 가지를 꼽았다.

2) 주택 시장과 기대가설이론

Hamilton and Schwab(1985)은 합리적 기대가설에 근거하여 주택가격에 대한 기대를 측정하고자 하였다. 연구는 1970년대 중반의 49개의 도심지역 자료를 이용하여 헤도닉 가격함수를 설정, 실증분석을 진행하였는데 주택가격에 대한 기대는 구조적으로 틀려 합리적 기대가설을 뒷받침하지 못하는 결과를 나타내면서 주택시장에서 합리적 기대가설에 대한 의구심을 품게 만들었다. 한편, 과거 자본이득 또한 기대 주택가격과 부의 상관관계를 보이기도 하였다.

Mankiw and Weil(1989)은 인구 구성과 미국 주택가격의 시뮬레이션을 바탕으로 적응적 기대가설을 통해 과거의 주택가격으로 단기 미래 주택가격을 예측하는 것이 합리적 기대가설보다 현실을 더욱 잘 설명해 주는 것을 밝혔다. 한편, Mankiw and Weil의 연구에서는 적응적 기대가설을 통한 기대가격은 과거의 가격 추세선을 연장하여 사용하는 방법을 통해 구성하였다.

Clayton(1996)은 합리적 기대가설을 통한 주택 기대가격 모델을 만들고, 이를 통해 단기 주택가격변동성을 실증적으로 설명하려고 하였으나, 연구의 대상이었던 1979년부터 1991년까지의

2) Bikhchandani, S. and Sharma, S.(2000), "Herd behaviour in financial markets", IMF Staff Papers, 47(3), p. 282.

벤쿠버 지역 분기별 단독주택 가격을 완전히 설명하지는 못하였다. 모형은 비교적 가격 변동성이 작은 시기에는 가격을 잘 예측하였으나, 가격 변동성이 큰 시기에는 가격을 잘 설명하지 못하는 결과를 보여주었다. 연구는 자가 거주주택을 월세로서 대용(proxy)변수화하고, 사용가능한 주택시장 정보와 함께 실제 주택 가격과의 차이 관계를 모형으로서 사용하였다.

Levin and Wright(1997)는 영국의 주택시장에서 주택 가격의 변화는 과거 주택가격의 변화에 구조적으로 영향을 받는다는 것을 밝혔는데, 이는 순행적 투자(momentum investment)로서도 해석될 수 있다. Levin and Wright는 주택 가격이 상승할 것이라고 예상되는 경우에 이사를 계획한 사람은 새로운 주택에 대한 매수 가격(계약 가격)과 기존 주택에 대한 매도 가격 사이의 이익을 가지게 된다고 하면서 이러한 상황이 전반적으로 발생할 경우 주택 시장에 투기(speculation)가 발생할 수 있다고 하였다.

Case et al.(2003)은 2002년 주택구입자를 대상으로 한 설문조사를 바탕으로, 주택구입자의 기대는 최근 경험에 큰 영향을 받는다는 것을 밝혔으며, Piazzesi and Schneider(2009) 또한 미국 주택가격 폭등 시기에 미시건 소비자 설문조사를 바탕으로 군집분석을 실행하여 순행적 가격예상(momentum forecast)이 시장가격에 영향을 줄 수 있음을 밝혔다.

Glaeser et al.(2014)은 합리적 기대가설 모형을 이용해 도시 간 지역 균형 조건 모형을 제안하여 주택 시장 기대의 높은 변동성과 자기상관성 그리고 평균 회귀성(mean reversion)을 설명하려고 하였다. 하지만, 연구는 평균으로의 회귀성에 대해서는 설명하였으나, 1년 이상의 해변 지역 가격 변동성과 자기상관관계를 해결하지는 못하였다.

Dua(2008)는 미시건대 소비자 태도 설문조사를 통해 주택 소비자를 관찰하였다. 연구에서는 단기에서는 주택가격의 상승이 주택 소비자의 소비심리에 긍정적인 영향을 나타내었으나, 장기적 관점에서는 주택가격의 상승이 소비심리에 부정적 영향을 나타내는 것을 발견하였다. 연구에서는 이러한 결과에 대해서 주택 가격 상승이 단기적으로 영향을 미치는 것으로 판단되나, 시간이 지날수록 그 효과가 작아진다고 하였다.

최영걸 외(2004)는 적응적 기대가설 모형을 통한 가격예측모형과 합리적 기대가설 모형을 통한 가격예측모형에 의한 기대가격을 주택가격결정모형에 대입하여 서울시 주택시장에서 작동되는 가격 기대심리에 어떤 기대가설이 우세할 것인가를 검정하였는데, 검증 결과, 서울시 주택시장에서는 적응적 기대가설이 유의하게 작동하고 있음을 밝혔다. 이러한 적응적 기대가설 하에서는 과거 가격의 상승이 다시 투자수요를 불러오고, 그 투자수요에 의하여 가격이 상승하게 되는 순환적 구조를 가지게 된다.

김재원, 최막중(2011)은 서울시 주택시장에서 기대되는 자본이득과 과거 가격변화의 관계를

그랜저 인과관계 분석과 부분조정모델을 통해 각각 그 방향과 영향 정도를 분석하였다. 논문에서는 기대자본이익을 추정하기 위해 차익거래조건(arbitrage transaction condition)을 사용하였는데, 이는 CAPM(Capital Asset Pricing Model)을 통한 자산시장에서의 주택가격과 주택임대시장을 통해 산출되는 주택가격이 일치하여야 한다는 점에 착안하였다.

Ⅲ. 분석 자료와 분석 틀

본 연구는 주거실태조사 마이크로 데이터 주택구입자금 응답사항으로부터 계산한 주택구입자금 자본구조에서 부채가 차지하는 비중(PLTV; Purchase Loan to Value)을 종속변수로, 한국감정원의 평균주택가격상승률의 과거 지역 시간가중평균 주택 상승률을 가장 중요한 독립변수로 설정하고, 소득, 순자산, 가구원수, 가구주연령의 가구특성과 주택규모의 주택특성, 거시통제 변수로서 주택구입시점의 CD금리를 통제하여 토빗 모형을 통한 통계적 연구 방법을 사용한다. 지역은 주거실태조사에서 나타나는 주거지의 시군구를 단위로 하여 분석한다. 본 절에서는 해당 데이터들을 통해 연구를 설계한 내용과 그 범위, 그리고 분석 틀에 대해서 살펴보고자 한다.

1. 분석 자료의 설계

본 연구에서 쓰이는 분석 자료는 2010년과 2014년도의 주거실태조사와 한국감정원의 평균주택가격상승률 그리고 한국은행 통화금융통계의 CD유통수익률(91일)의 세 가지 자료를 통합하여 사용하였다. 자료의 통합 구성을 그림으로 표시하면 <그림 2>와 같다.

주거실태조사는 2006년부터 2년 단위로 짝수년도에 일반조사가 실시되어 가장 최근의 2014 주거실태조사에 이르고 있다. 주거실태조사 일반조사는 주택 구입자금의 자본구조 정보와 함께 거주하고 있는 주택의 소재지를 시군구까지 나타내고 있으며, 연구를 위해 통제해야 할 가구특성



<그림 2> 자료의 통합 구성 구조도

(순자산, 소득, 가구원수 등)과 주택특성(주택유형, 주택규모 등)까지 포함하고 있으며, 현재의 투기과열지구, 조정대상지역 등의 LTV 규제 영향이 없거나 비교적 적어 본 연구를 수행함에 있어서 적절한 데이터로 판단되었다.

소득, 순자산, 가구원수, 가구주연령의 가구특성과 주택규모의 주택특성 변수를 도입, 통제하는 이유는 소비재로서 주택을 소비할 경우를 통제하기 위해서이다. 연구는 주택의 투자재적 수요에 집중하기 때문에, 만약 정말 그 가구가 그 주택을 필요로 하는 경우, 즉, 소비적 필요에 따라서 어쩔 수 없이 차입을 통해 주택을 구입한 경우, 그러한 차입은 주택의 투자적 특성에 기인한 것은 아니라고 할 수 있으므로 가구 특성을 연구에서 통제할 수 있어야 할 것이다. 이를 바꾸어 이야기하면, 예산 제약 하에 처한 가구는 투자 목적의 수요가 없을 때에도 대출을 할 수밖에 없을 것이며, 이는 투자 수요는 아닌 것이다.

평균주택상승률은 한국감정원에서 2003년부터 조사된 전국주택가격동향조사를 사용하였다. 전국주택가격조사 자료는 매월 지역별, 주택 유형별, 규모별 평균 매매가격 지수를 전국주택 가격동향조사로서 포함하고 있어 연구에 적합하다. 주택유형은 단독, 연립, 아파트로 구분되어 있으며, 단독에는 다가구주택이, 연립주택에는 다세대주택이 각각 포함되어 있다. 한편, 주택유형이 아파트인 경우에는 시군구 단위까지 데이터가 구축되어 있으며, 단독주택, 연립주택의 경우에는 시도 단위까지 데이터가 구축되어 있어, 앞서 연구 대상을 아파트인 경우로 한정할 원인이 되었다. 또한, 아파트는 타 주택재화 유형과 달리 서로 비슷한 형태와 양상을 나타내므로 대상을 아파트로서 한정하는 것은 연구 결과의 해석에 더욱 용이하다고 할 수 있다.

앞서의 선행연구에서 과거평균 주택가격 상승률을 통한 기대 가격 상승률의 대용 변수(proxy variable)의 몇 가지 경우를 살펴보았다. 본 연구에서는 적응적 기대가설(adaptive expectation)을 가정, 시간가중평균을 사용하여 해당 값을 산출하였다. 적응적 기대가설이론을 가정하여 과거의 가격경험을 통해 미래 예상 자본수익률을 기대한다고 가정하면, 그 기댓값은 과거에 이미 나타난 값들을 가중 평균하여 산출되는데, 그 기본적인 수식은 아래와 같다.

$$P_t^e = P_{t-1}^e + \alpha(P_{t-1} - P_{t-1}^e), (0 < \alpha < 1) \quad \langle \text{식 5} \rangle$$

앞서의 선행연구를 통해 알아본 기대의 대용변수(proxy variable)를 구하는 방법에는 과거의 가격 추세선을 연장하는 방법, 그 평균을 사용하는 방법, 시간가중평균을 이용하는 방법, 차의 거래조건을 통해 값을 도출하는 방법 등의 여러 가지 방법이 있었으나, 본 연구에서는 시간가중평균을 사용하여 그 대용변수를 산출하였다.

본 연구에서 시간가중평균을 사용하는 데에는 두 가지 큰 이유가 존재한다. 첫 번째로, Case et al.(2003)과 Piazzesi and Schneider(2009)에서 발견할 수 있듯, 주택 소비자는 최근의 가격 변화에 더욱 큰 영향을 받기 때문에 최근의 가격 변화에 큰 가중치를 두는 것이 더욱 타당하기 때문이며, 두 번째로, 해당 조사 기간 중 서브프라임 사태가 발생하여 시장 상황, 가격 상승률에 변화가 존재할 수 있기 때문에 단순 평균을 사용하는 것은 문제가 있기 때문이다.

$$W_i = \frac{\lambda^{i-1} \times (1 - \lambda)}{1 - \lambda^n} \quad \langle \text{식 6} \rangle$$

i : 시점
 n : 자료 수
 λ : *decay factor*

시간가중평균의 방법에는 몇 가지 방법이 존재할 수 있는데, 그 방법 중 일반적으로 사용되고 있으며, 최근 값에 큰 가중치를 줄 수 있는 Boudoukh et al.(1998)에 제시된 방법을 사용하여 가중치를 계산하였다. 수식의 λ (decay factor)는 0에서 1사이의 값을 가지는데, λ 가 0과 가까워 값이 작아질수록 현재 시점에 가까운 값에 더 높은 가중치가 주어지게 되며, λ 가 1에 가까워 값이 커질수록 각 시점에 적용되는 가중치가 서로 비슷해지게 된다. 연구에서는 <표 1>의 0.25, 0.5, 0.75 중 0.25값을 사용하여 가까운 시점에 더욱 높은 가중치를 주었다.

산출된 가중치와 한국감정원의 평균주택가격상승률 데이터를 사용하여 해당 년도(주택구입년도)를 기준으로 시간가중평균값을 산출한 산식은 아래와 같다. 사례의 개별통제변수, 즉 가구특성과 주택특성, 그리고 주택구입시점 CD금리를 통제한 상태에서 해당 값이 양수 값을 가진다면, 적응적 기대가설 가정 하에서 해당 과거 평균지역주택가격 상승률이 높을수록 주택구입자금 자본구조에서 타인자본 비율을 높이는 쪽으로 작용하게 될 것이다.

<표 1> λ 값에 따른 가중치 (Sample size=4)

λ	0.25	0.5	0.75
$W(t-1)$	0.752941	0.533333	0.365714
$W(t-2)$	0.188235	0.266667	0.274286
$W(t-3)$	0.047059	0.133333	0.205714
$W(t-4)$	0.011765	0.066667	0.154286

$$E(R_{H_x}) = \sqrt[n]{\prod_{i=1}^n (1 + W_i \times R_{(x,i)})} - 1 \quad \langle \text{식 7} \rangle$$

$E(R_{H_x})$: x 지역 기대 가격 상승률

$R_{(x,i)}$: x 지역의 i 년도 가격 상승률

W_i : i 년도 가중치

n : 자료 수

본 연구의 대상은 가구 자산에 대한 조사가 이루어지지 않은 2012 주거실태조사를 제외한 2010, 2014년도의 조사를 바탕으로, 조사시점으로부터 4년 이내에 주택을 구입하여 해당 주택에 거주하는 자가 거주가구를 대상으로 한다. 자가 거주가구만을 대상으로 하는 것은 전세, 반전세, 월세와 같은 임차가구의 경우에는 실수요(소비수요)에 의한 주거 선택으로서 자본 차익을 기대하는 투자수요가 아니기 때문이며, 조사시점으로부터 4년 이내에 주택을 구입한 가구를 대상으로 하는 것은 해당 조사의 간격인 4년(2010년과 2014년)을 연속적으로 유지하기 위한 최소 년도이기 때문임과 동시에, 더 크게 조사 시점에 큰 차이가 존재하게 될수록 연구 목적상 통제해야 할 가구특성에 괴리가 발생하게 될 가능성 또한 높아지기 때문이다.³⁾

마지막으로, 현재 주택을 어떻게 마련하였는가하는 2010년 주거실태조사의 8번, 2014년 주거실태조사의 11번 설문응답에 대해서 ④경우의 증여나 상속의 경우를 연구의 대상에서 제외하였다. 이는 증여나 상속의 경우, 타인자본비율 설정이 해당 주택의 구입시점에 주택 소비자의 선택이 아닌, 증여·상속 이전의 타인자본비율을 그대로 승계하였을 경우가 많을 것으로 판단되기 때문이다.

종합하면 2010, 2014 주거실태조사 일반조사 가구 중 조사시점으로부터 4년 이내에 아파트를 구입한 자가 거주가구가 연구의 대상이 된다고 할 수 있다.

2. 주요 변수의 설정

본 연구에서 중점적으로 살펴보고자 하는 변수로는 가장 먼저 주택구입자금 부채비율(PLTV; Purchase Loan to Value)로서 연구의 종속변수로서 설정되었다. 해당 값은 위의 분석 자료의 설계에서 언급하였듯, 주거실태조사의 주택 구입자금 경로에 대한 설문 문항을 통해 산출하였다.

3) 한국감정원에서 발표되는 평균주택가격상승률 데이터(아파트, 시군구)의 경우 2003년 11월부터 그 값이 발표되어, 2004년의 값부터 연평균 값을 사용할 수 있는 점도 해당 기준을 4년으로 설정한 이유가 되었음.

2010년 주거실태조사의 경우에는 31번 문항으로서 주택 구입자금 조달 방법을 금융기관에서 융자받은 금액(여러 기관에서 융자받은 경우 합산)과 금융기관 외에 다른 곳에서 빌린 금액, 그리고 부모, 친지 등으로부터 무상으로 받은 금액 및 자기자금의 세 가지 경우로 나누고 있다. 여기에서 31번 문항의 세 가지 금액을 합산하면 매입 당시 주택가격과 동일한 값을 보여준다.

2014년 주거실태조사의 경우에는 13번 문항에서 현재 살고 있는 주택의 구입자금의 조달 경로에 대해 자기자금, 금융기관에서 융자받은 금액(여러 기관에서 융자받은 경우 합산), 금융기관 외에 다른 곳에서 빌린 금액, 그리고 부모, 친지 등으로부터 무상으로 받은 금액으로서 네 가지로 구분하고 있다. 2014년 주거실태조사의 경우에도 이 네 가지 금액을 합산하면 문항 12번의 매입, 증여, 상속 당시 주택가격과 동일한 값을 나타낸다.

주택구입자금 부채비율은 2010년과 2014년 주거실태조사 조사 자료를 동일하게 활용하는 것을 견지하는 동시에, 부모, 친지 등으로부터 무상으로 받은 금액의 경우는 CD금리 등 일반적인 타인자본 비용을 적용받기 보다는 자기자본 비용을 적용받는 것이 타당한 바, 금융기관 융자금과 금융기관 외에 다른 곳에서 빌린 금액을 타인 자본으로 보고, 이를 매입 당시 주택가격으로 나누어 산출하였다.

앞서의 선행연구에서 순자산과 대출과의 관계를 확인 한 바 있다. 이에 본 연구에서는 순자산을 주요 통제변수로서 도입, 활용하였다. 순자산은 가구 총자산에서 가구 총부채의 차이로서 산출하였으며, 2010년의 경우 37번 문항 설문 합계(총 자산)에서 38번 문항의 총 부채액을 차감, 2014년의 경우 52번 문항의 총 자산에서 53번 문항의 총 부채를 차감, 자연로그를 취하여 사용하였다. 순자산은 앞서 그 산출 방식에서 확인할 수 있듯, 가용 자산에서 부채를 뺀 것으로 조사시점에서의 잔여자본(residual equity)를 뜻한다. 따라서 순자산이 높을수록 자기자본을 조달할 수 있는 확률이 높아지고, 순자산이 낮을수록 자기자본 조달 확률이 낮아져 순자산과 주택구입자금의 타인자본비율은 반비례할 것임을 예상할 수 있다.

소득은 연평균 총 소득으로서 2010년 주거실태조사의 34번 설문 문항의 총 합계와 2014년 주거실태조사의 49번 설문 문항의 총 합계의 월평균 총소득 자료를 연평균으로 변환한 후, 자연로그를 취하여 사용하였다. 앞서의 순자산이 일정하다고 가정할 경우, 소득이 높을수록 타인자본비용의 감당 및 채무 상황이 용이할 것이라고 예상할 수 있다. 따라서 소득과 주택구입자금의 타인자본비율은 선행연구에서 관찰하였던 바와 같이 정비례하는 관계를 가져 소득이 높을수록 그 비율이 높을 것이라고 예상할 수 있다.

가구원 수, 주택면적, 가구주연령 역시 각각의 주거실태조사 문항을 통해서 도출하였으며, 선행연구에 따라 가구원 수(+)가 많을수록, 주택면적(+)이 넓을수록, 가구주연령(-)이 어릴수록 주택 구입 자본비용에서 타인자본비율이 높을 것이라고 예상하였다.

CD금리는 한국은행 통화금융통계의 CD유통수익률(91일)의 1년 평균 제시된 값을 그대로 사용하였으며, CD금리가 타인자본비용의 대용 변수(proxy variable)이므로 타인자본비율과는 부의 상관관계를 가질 것이라 예상하였다.

지역은 주거누리 주거실태조사에 등록된 시군구 자료를 시군구 코드의 다섯 자리로 변환, 여기에서 다시 시도를 구분하는 앞의 두 자리 코드를 이용하여 분류하였다. 서울특별시(11), 경기도(41), 그리고 인천광역시(28)의 수도권(=1), 부산광역시(26), 대구광역시(27), 광주광역시(29), 대전광역시(30), 울산광역시(31)의 주요 광역시(=2), 마지막으로 비수도권·비주요광역시(=0)의 세 가지 경우로 나누어 더미변수로 구성하였다.

3. 분석 방법과 모델

본 연구는 앞서 언급하였듯, 주거실태조사 마이크로 데이터 주택구입자금 응답사항으로부터 계산한 주택구입자금 자본구조에서 부채가 차지하는 비중(PLTV; Purchase Loan to Value)을 종속변수로, 한국감정원의 평균주택가격상승률의 과거 지역 평균 주택 상승률을 가장 중요한 독립변수로 설정하고, 소득, 순자산, 가구원수, 가구주연령의 가구특성과 주택규모의 주택특성, 거시통제 변수로서 주택구입시점의 CD금리를 통제하여 토빗 모형을 통한 통계적 연구 방법을 사용한다.

본 연구의 종속변수 PLTV의 경우, '0'에서 중도 절단되어 있어 일반적인 OLS(Ordinary Least Squares) 회귀분석을 사용하는 것에 한계가 존재한다. 타인자본을 사용하는 주택 소비자만을 대상으로 할 경우, 표본 선택 편의(sample selection bias)의 문제가 발생하며, 타인자본을 사용하지 않는 주택 소비자의 값을 관측치를 '0'으로 사용 할 경우에는 종속변수 분포가 제한됨으로써 대칭적 분포를 가정하고 있는 최소제곱추정방식의 문제(truncation bias)가 발생한다.

토빗 모형(tobit model)은 경제학자 Tobin이 절단된 자료의 분석을 위하여 고안한 것으로 제한종속변수모형(limited dependant variable regression model) 또는 절단 회귀모형(censored regression model)으로 불리기도 하며, 종속변수의 값은 연속이나, 정보가 절단(censored)된 자료로서 일부에 대해서만 이용 가능한 경우, 회귀분석을 진행하기 위해 이용가능한 방법이다. 본 연구의 종속변수 PLTV의 경우 그 값은 연속적인 값은 갖고 있으나, 대출을 하지 않은 집단에 대해서는 관측치가 존재하지 않아 그 일부에 대해서만 이용 가능하다고 볼 수 있다. 이러한 경우 통상적 최소제곱추정방식에 의한 경우에는 대상들을 '0원 대출'이라고 보든, 무시하고 관측치가 이용 가능한 그룹에 대해서만 추정을 하든 잘못된 결과를 불러오게 된다.

따라서 본 연구에서는 종속변수의 문제를 해결할 수 있는 토빗 모형을 사용, 그 좌측 절단선(left-censoring)을 '0'으로 지정해 분석을 진행하도록 하였다.

타인자본비율과 적응적 기대가설을 바탕으로 한 기대 수익률의 관계를 파악하기 위해서 소득, 순자산 등의 요인을 포함한 토빗 모형을 설정하면 다음과 같다.

$$\begin{aligned} Y_i^* &= \beta_1 + \beta_2 E(R_H) + \beta_3 \text{소득} + \beta_4 \text{순자산} \dots + u, u \sim N(0, \sigma^2) \\ Y_i &= Y_i^* \text{ (if } Y_i^* \geq 0) \\ Y_i &= 0 \text{ (if } Y_i^* = 0) \end{aligned} \quad \langle \text{식 8} \rangle$$

위 수식에서 Y_i 는 타인자본비율을 의미하며, $E(R_H)$ 는 기대 수익률을 의미하는데, 타인자본비율은 '0'에서 절단되는(Censored) 형태를 보이므로 잠재변수 타인자본비율 Y_i 는 '0'에서 '0'의 값을 가지며 좌측 중도절단된다.

IV. 결과 분석

〈표 2〉는 2010년도와 2014년도의 주거실태조사 데이터를 통합하여 과거 4년간의 시간가중평균 지역주택가격 상승률⁴⁾을 적용, 토빗모형 회귀분석 결과를 나타낸 것이다. 통계 모형은 로그 우도값(log likelihood)이 약 173.84, 자유도 7의 카이제곱분포를 따르며, 통계적으로 유의한 것으로 나타났다.

먼저, 소득(ln_income)의 경우, 계수 값이 0.035로 나타나 소득이 1% 상승할 경우 주택구입자금에서 부채비율이 0.035만큼 증가하는 정의 상관관계를 확인하였으며, 이는 99% 신뢰수준에서 통계적으로 유의하였다. 이는, 소득이 높아질수록 대출상환에 대한 여력이 상승하고, 주거비에 대한 지출 또한 증가하는 것으로 해석할 수 있으며, 기존의 LTV를 통한 선행연구와 일치하는 결과이다.

또한, 순자산변수(ln_netequity)에 대해서도 순자산이 1% 상승할 경우, 주택구입자금 부채비율이 -0.113만큼 감소하며, 순자산이 높을수록 주택구입자금에서 타인자본비율이 낮아지는 부의 상관관계를 보여주어 기존의 선행연구와 동일한 방향의 결과를 나타내는 것으로 확인되었으며, 이는 통계적으로 99% 신뢰수준에서 유의하였다.

가구주 연령(age)의 경우에도 Follain(1990), 김승욱, 남영우(2012)의 선행 연구와 동일한 방향성이 확인되었는데, 가구주 연령이 낮을수록 주택구입자금 자본구조에서 부채의 비중이

4) λ값은 0.25를 사용하였다.

〈표 2〉 2010, 2014 통합 데이터 분석 결과

종속변수: 주택구입자금 부채비율(pltv)		Coef.	<i>t</i>	$p> t $
# of obs	2,758			
Log likelihood	173.83566			
LR chi2(7)	476.61			
Prob>chi2	0.0000			
Pseudo R^2	3.6965			
소득 (ln_income)		0.035	3.56	0.000
순자산 (ln_netequity)		-0.113	-18.99	0.000
가구주연령 (age)		-0.002	-4.04	0.000
주택면적 (area)		0.001	4.67	0.000
가구원수 (nohm)		0.001	0.10	0.923
CD금리		-4.032	-2.98	0.003
기대가격상승률(람다:0.25)		0.073	1.40	0.162
상수항		1.282	15.65	0.000

주: 223 left-censored observations at pltv ≤ 0, 2,535 uncensored observations.

높아지는 것을 확인하였으며, 통계적으로 99% 신뢰수준에서 유의하였다.

주택면적(area)과 주택구입자금 타인자본비율의 관계는 주택면적이 넓을수록 타인자본비율이 높아지면서 정의 상관관계를 나타내었는데, 이는 김승욱, 남영우(2012)의 LTV를 통한 선행연구와는 상반되는 결과였다. 한편, 가구원수(nohm)에 대해서는 부의 관계를 나타내었지만 통계적으로 유의미하지 않았다.

앞서 거시적인 분석을 통한 선행연구들이 다수 존재하였던 금리와 대출규모의 경우에는, CD금리와 주택구입자금 부채비율을 통한 미시적 분석에서도 부의 상관관계를 갖는 것으로 확인되었으며, CD금리가 1만큼 변동할 때, 주택구입자금 부채비율은 -4.032만큼 변동하는 것으로 나타났으며, 99% 신뢰수준에서 통계적으로 유의했다.

앞서의 모든 변수에서 선행연구와 일치하는 결과가 나타났으나, 본 연구에서 가장 관심을 가진 중요한 변수로서 해당 지역의 과거 평균가격 상승률과 주택구입자금 부채비율에 대해서는 0.073의 계수 값으로 정의 관계를 나타내는 것으로 관찰되었지만, 이는 1.40의 *t* 값을 가져 통계적으로 유의미하지 않은 것으로 판단되었다.

앞서 2008년부터 2014년까지 통합 데이터를 통한 결과에서 과거 가격의 시간가중평균을 사용한 기대가격상승률과 주택구입자금 부채비율에는 통계적으로 유의한 관계가 없는 결과가 나타났지만, 이는 가격이 상승하여 기대가격상승률이 양수인 지역과 과거 가격이 하락하여 기대가격상승률이 음수인 경우가 비대칭적인 결과를 나타낼 수 있다. 이는 먼저, 자본구조의 이론적 배경에서 살펴보았듯, 기대가격상승률이 음수로 예상되어 자본이득이 발생하지 않을 것으로 예상될 경우에 주택소비자는 주택을 구입하는 의사결정 자체가 이루어지지 않을 것이기 때문이며, 두 번째로는, Case et al.(2003), Piazzesi and Schneider(2009)가 보여주었듯, 주택소비자는 최근의 경험을 통하여 미래가격을 예측(momentum forecast)하는 경향이 있으며, 이는 가격 상승시기에 더욱 유의할 수 있기 때문이다. 따라서 본 연구에서는 과거 주택가격이 상승하여 양수의 기대가격상승률이 형성된 지역과 과거 주택가격이 하락하여 음수의 기대가격상승률이 형성된 지역을 구분하여 추가적으로 토빗 회귀분석을 실행하였다.

〈표 3〉은 기대가격상승률을 기준으로 하여 기대가격상승률이 양수인 경우를 분석한 결과이며, 〈표 4〉는 기대가격 상승률이 음수의 경우를 분석한 결과이다.

분석 결과, 순자산, 가구주연령, 주택면적 변수의 경우에는 기대가격상승률이 양수인 경우와 음수인 경우를 나누지 않고 분석한 기존의 경우와 동일한 방향과 비슷한 결과를 나타내었다. 소득변수의 경우에는 기대가격 상승률이 양수인 경우, 통합표본과 동일하게 정의 상관관계와 99% 신뢰수준에서의 통계적 유의성을 나타내었으나, 기대가격 하락예상지역에서는 통계적으로 유의하지 않은 모습을 보여주었다. 이는 가격상승 예상지역의 경우 고소득자의 부채 투자가 나타나지만, 그렇지 않은 경우에는 소득정보보다는 실제 거주수요에 따라 부채가 발생하는 것으로 해석할 수 있다.

한편, 연구에서 중점적으로 살펴보고자 하는 기대가격상승률 변수와 주택구입자본구조의 타인자본비율간의 관계의 경우 〈표 3〉의 기대가격 상승예상 지역의 경우에는 기대 주택가격 상승률이 1만큼 더 높을수록 주택구입자금 부채비율이 약 0.284만큼 증가하며 정의 상관관계를 나타내었으며, 이는 99% 신뢰수준에서 통계적으로 유의한 것을 확인할 수 있었다.⁵⁾ 하지만, 〈표 4〉의 기대가격 하락 예상지역의 경우, 기대가격상승률 변수와 주택구입자본구조의 타인자본 비율간의 관계가 유의미하지 않은 결과를 보여주는 것으로 나타났다. 이에 더하여 기대가격 하락 예상지역의 경우, CD금리 변수와 소득변수도 통계적으로 유의하지 않은 모습을 보여주었다. 이는 기대가격 하락 예상지역의 경우, 주택의 수요는 투자 가치(investment value)보다는 그 사용가치(usage value)에 영향을 받게 되며, 따라서 주택구입자금의 자본구조 또한 소득이나 기대가격 예상

5) 2014년 주거실태조사만을 분석한 경우에는 계수값 0.227을 가지며, t 값은 1.52로 나타났으며, 2010년 주거실태조사만을 분석한 경우에는 계수값 0.301을 가지며, t 값은 3.89로 나타남.

〈표 3〉 2010, 2014 통합 데이터, 기대 상승률 상승지역 분석 결과

종속변수: 주택구입자금 부채비율(pltv)		Coef.	<i>t</i>	$p> t $
# of obs	1,938			
Log likelihood	150.9090			
LR chi2(7)	383.75			
Prob > chi2	0.0000			
Pseudo R^2	4.6838			
소득 (ln_income)		0.041	3.48	0.001
순자산 (ln_netequity)		-0.121	-17.06	0.000
가구주연령 (age)		-0.002	-3.38	0.001
주택면적 (area)		0.001	3.91	0.000
가구원수 (nohm)		0.002	0.37	0.708
CD금리		-7.860	-4.68	0.000
기대가격상승률(람다:0.25)		0.284	4.34	0.000
상수항		1.389	13.96	0.000

주: 156 left-censored observations at pltv<=0, 1,782 uncensored observations.

〈표 4〉 2010, 2014 통합 데이터 기대 상승률 하락지역 분석 결과

종속변수: 주택구입자금 부채비율(pltv)		Coef.	<i>t</i>	$p> t $
# of obs	820			
Log likelihood	40.633			
LR chi2(7)	127.34			
Prob > chi2	0.0000			
Pseudo R^2	2.7638			
소득 (ln_income)		0.024	1.37	0.170
순자산 (ln_netequity)		-0.103	-9.52	0.000
가구주연령 (age)		-0.002	-2.92	0.004
주택면적 (area)		0.001	3.40	0.001
가구원수 (nohm)		-0.004	-0.46	0.642
CD금리		-0.473	-0.19	0.849
기대가격상승률(람다:0.25)		0.066	0.16	0.876
상수항		1.214	8.30	0.000

주: 67 left-censored observations at pltv<=0, 753 uncensored observations.

상승/하락률을 통한 설명이 어려워지는 것으로 해석할 수 있다.

V. 결론 및 정책적 시사점

연구는 주택 소비자의 자본구조 선택에 있어서 금융/경제적 시장 상황 외에도 적응적 기대가설에 근거하여 주택을 소비하고자 하는 지역의 과거 가격 변화에 영향을 받을 것이라는 가설을 제시하고, 시간가중평균을 통해 기대가격 상승률을 산출하여 그 관계를 실증 분석하였다.

2010년 주거실태조사와 2014년 주거실태조사를 통합한 데이터를 통하여서는 과거 시간가중평균 지역주택가격 상승률과 주택구입자금의 타인자본비율의 관계에 있어서 통계적으로 유의한 결과를 찾기가 어려웠지만, 가격 상승이 예상되는 지역과 가격 하락이 예상되는 지역을 나누어 분석한 결과, 가격이 상승이 예상되는 지역에서는 지역의 과거 주택가격 경험과 주택구입자금의 타인자본비율이 서로 정의 상관관계를 가지며, 이는 통계적으로 유의한 것으로 나타나, 비대칭적인 결과를 나타내었다. 주택 가격 상승이 기대되어 자본이득이 발생할 것으로 예상되는 경우에는 주택의 투자적 가치(investment value)를 통해 주택 구입을 위한 타인자본비율이 증가하는 결과를 나타내었으며, 이는 연구 가설을 잘 뒷받침하는 것으로 해석할 수 있다.

한편, 과거 지역주택가격이 하락한 지역에서는 시간가중평균 한 과거 가격변동이 유의미하지 않은 결과를 보여주는 것으로 나타나, 하락이 예상되는 지역의 경우 주택의 수요는 투자 가치보다는 그 사용가치(usage value)에 영향을 받게 되며, 따라서 주택구입자금의 자본구조 또한 소득이나 과거 평균주택가격상승률 등의 투자적 가치를 통한 설명이 어려워지는 것으로 해석할 수 있다.

또한, 이번 연구에서는 기존 LTV 연구의 통제 변수를 도입하여, 소득, 순자산, 가구원수, 가구주나이의 가구특성, 주택규모의 주택특성과 타인자본비율 간의 관계를 비교해 볼 수 있었다. 소득과 순자산, 가구주나이의 주택면적 모두가 LTV 등을 통해서 분석된 기존 선행연구의 결과와 큰 차이를 보이지 않았으나, 가구원수의 경우에는 대체적으로 통계적인 유의성이 부족한 것으로 판단되었다. 연구 결과, 주택구입자금의 타인자본비율(PLTV)과 LTV는 대체적으로 상당히 비슷한 결정요인을 가지고 있다고 해석할 수 있다.

부동산 자산 편중도가 70% 이상을 차지하는 상황에서 주택 관련 대출을 예측·파악하는 것은 국가와 국민의 자본건전성을 위해서 매우 중요하다. 본 연구는 주택가격 상승의 기대에 따라 주택구입자금에서 타인자본비율이 증가함을 밝혔다. 이는 주택 가격 변화가 주택시장의 자본구조를 가늠할 수 있는 정책적 신호(signal)로서 사용될 수 있음을 시사한다.

실무적으로 주택자산을 통한 대출 의사결정을 진행할 때 주로 고려하는 것은 LTV와 DTI로

대변되는 부동산 담보평가 가치와 대출자의 소득 능력으로, 주택 특성과 개인 특성에 근거하여 대출을 실행하는 것으로 파악할 수 있다. 연구는 이러한 주택 특성과 개인 특성 요소 외에도, 그 주택이 입지하는 지역, 그 지역의 과거 가격 변수로서의 지역 특성이 대출 조건의 고려 요소로서 추가될 수 있음을 시사한다.⁶⁾ 연구의 결과에 따르면, 가격이 상승하여 주택 소비자가 그 지역의 가격 상승 기대감에 근거하여 타인자본비용을 높여 사용한 경우, 그에 대한 사회적 타인자본 비용은 가격이 지속 상승하지 않은 지역의 사회적 타인자본 비용보다 높은 결과를 가져온다. 이는 가격 상승 기대감이 특정 지역의 주택 소비에 있어서 타인자본 수요를 불러오고, 그 결과 지역에 타인자본 비율이 집중되어 사회적 자본건전성을 위협할 수 있기 때문이다.

현재 투기과열지구는 9억원 이하 40%, 9억원 초과 20%(15억 이상 대출 불가), 조정대상지역은 9억원 이하 50%, 9억원 초과 30% 등의 강화된 LTV가 적용되며, 투기지역과 투기과열지구 등에 대해 LTV가 강화 적용되는 현황과 선례는 있지만, 이는 타인자본의 사회적 비용을 추가하여 고려했기 보다는 가격상승을 제한하려는 움직임이라고 볼 수 있으며, 지금까지 대출 등 주택 자산에 대한 타인자본 조달은 지역적 고려 없이 진행되어 왔다. 국민 자본건전성을 제고하기 위해서는 주택가격이나 소득능력 이외에도 지역적/미시적 변수를 더욱 고려하여 그 위험을 통제, 자본건전성을 제고할 수 있어야 할 것이다.

학술적으로, 본 연구는 거시적으로 주택 담보대출 수요와 주택가격의 관계를 파악하는 것뿐만 아니라, 미시적, 지역적으로 둘은 중요한 관계를 가지고 있으며, 이러한 관계를 파악하려는 것 또한 중요하다는 것을 시사한다. 시점이나 경기변동에 따라서 사람들의 기대가 달라지며, 가격 상승기에 기대가 긍정적으로 변한다는 연구 결과나 거시적으로 이자율 등과 타인자본 수요를 파악한 논문은 많으나, 미시적·지역적 관점의 연구는 아직까지 미미하다.

마지막으로 본 연구는 소득, 순자산, 가구주연령, 가구원수, 주택면적을 통한 개별적 통제가 완벽할 수는 없다는 한계점이 존재한다. 이는 주택담보대출의 조건, 내역 등과 함께 이를 개별적으로 통제할 수 있는 데이터 즉, 거시적(수도권/비수도권) 정보는 존재하지만, 지역적, 미시적 정보가 존재하지 않는 것에 일편 기인한다.⁷⁾ 이후로의 연구로는 이러한 주택 구입자금의 자본구조의 해석에 있어서 주택 구입자금 금융수요 등 금융수요와 주택가격 간의 관계에 있어서 미시적·지역적으로 해석 할 수 있는 데이터, 연구가 더욱 필요하다고 할 수 있으며, 이를 파악하는데 초점이 맞추어지길 기대해본다.

6) 비록 담보평가 시에 지역요인이 감안되긴 하지만, 본 연구에서 시사하는 바인 기대 가격 상승률을 의미하는 바는 아니다.

7) 통계청의 가계금융·복지조사와 한국주택금융공사의 주택금융 및 보금자리론 수요실태조사가 존재하지만, 수도권/비수도권 구분 외의 지역 정보는 공개되지 않음.

참고문헌

- 구만수, 서정렬, 최열, 2013, “가구 특성을 고려한 주택구입자금 대출에 관한 분석”, 부동산학보 55권, 5-16.
- 김도훈, 김진수, 2012, “국제경제: 중국 베이징 도시가구의 주택담보대출 이용실태 분석”, 국제지역연구 16권 3호, 135-155.
- 김승욱, 남영우, 2012, “주택가격변화에 따른 가계부채의 위험증가에 대한 연구”, 부동산학보 51, 240-251.
- 김재원, 최막중, 2011, “The relationship between the expected capital gain and past price change rate in Seoul housing market”, 서울대학교 환경대학원 석사학위논문.
- 유승동, 신승우, 2012, “주택대출 선택이 주택가격 변동에 미치는 영향에 대한 실증분석”, 한국지역개발학회지 24권 제2호, 29-45.
- 정의철, 2005, “구조모형을 통한 주택금융수요 추정에 관한 연구”, 국토계획 40권 6호, 147-157.
- 최막중, 지규현, 조정래, 2001, “주택 금융제약이 주택 소비규모와 점유형태 선택에 미치는 영향에 관한 실증분석”, 주택연구 10권 1호, 33-48.
- 최영걸, 이창무, 최막중, 2004, “서울시 주택시장에서 작동되는 가격기대심리에 관한 실증연구”, 국토계획 39권 2호, 131-141.
- Anenberg, E., and Kung, E., 2013, “Estimates of the size and source of price declines due to nearby foreclosures”, The American Economic Review, 104(25), 2527-2551.
- Bikhchandani, S. and Sharma, S., 2000, “Herd behaviour in financial markets”, IMF Staff Papers, 47(3), 279-310.
- Boudoukh, J. and Richardson, M. and Whitelaw, R., 1998, “The best of both worlds: A hybrid approach to calculating value at risk”, Risk 11, 64-67.
- Brueckner, K., 1994, “The demand for mortgage debt: Some basic results”, Journal of Housing Economics, 3(4), 251-262.
- Campbell, J. Y., Giglio, S. and Pathak, P., 2011, “Forced sales and house prices”, The American Economic Review, 101(5), 2108-2131(24).
- Case, K. E. · Shiller, R. J. and Quigley, J. M., 2003, “Home-buyers, housing and the macroeconomy”, Berkeley Program on Housing and Urban Policy, 149-188.

- Case, K. E. · Shiller, R. J., 1989, "Forecasting prices and excess returns in the housing market", *Journal of the American Real Estate and Urban Economics Association*, 18, 253-273.
- Case, K. E. and Shiller, R. J., 1989, "The efficiency of the market for single-Family Homes", *The American Economic Review*, 79, 125-137.
- Clayton, J., 1996, "Rational expectations, market fundamentals and housing price volatility", *Real Estate Economics*, 24, 441-470.
- Dua, P., 2008, "Analysis of consumers' perceptions of buying conditions for houses", *Journal of Real Estate Finance and Economics*, 37, 335-350.
- Follain, J. R., 1990, "Mortgage choice", *Real Estate Economics*, 18(2), 125-144.
- Franco, M. and Miller, Merton H., 1958, "The cost of capital, corporation finance and the theory of investment", *The American Economic Review*, 48(3), 261-297.
- Gerlach, Stefan and Peng, W., 2005, "Bank lending and property prices in Hong Kong", *Journal of Banking & Finance*, 29, 461-481.
- Gimeno and Carrascal, M., 2010, "The relationship between house prices and house purchase loans: The Spanish case", *Journal of Banking & Finance*, 34(8), 1849-1855.
- Gleaser, E. L., Joseph, G., Eduardo, M., Nathanson, and Charles, G., 2014, "Housing dynamics: An urban approach", *Journal of Urban Economics*, 81, 45-56.
- Hamilton and Schwab, 1985, "Expected appreciation in urban housing markets", *Journal of Urban Economics* 18, 103-118.
- Hofmann, B., 2004, "The determinants of bank credit in industrialized countries: Do property prices matter?", *International Finance*, 7(2), 203-234.
- Hott, C., 2011, "Lending behaviour and real estate prices", *Journal of Banking & Finance* 35(9), 2429-2442.
- Levin, E. J. and Wright, R. E., 1997, "The impact of speculation on house prices in United Kingdom", *Economic Modelling*, 14, 567-585.
- Mankiw, N. and Weil, D. N., 1989, "The baby boom, the baby bust and the housing market", *Regional Science and Urban Economics*, 19, 235-258.
- McKinsey Global Institute, 2013, "Beyond Korean style", <http://www.mckinsey>.

com/insights/asia-pacific/beyond_korean_style

- Mian, A. and Sufi, A., 2008, "The consequences of mortgage credit expansion: Evidence from the 2007 mortgage default crisis", NBER no.13936, Federal Reserve Bank of Chicago, issue May, pp. 129-132.
- Mian, A. and Sufi, A., 2011, "House prices, home equity-based borrowing, and the US household leverage crisis", The American Economic Review, 101(5), 2132-2156(25).
- Mian, A., Sufi, A. and Trebbi, F., 2015, "Foreclosures, house prices, and the real economy", The Journal of Finance, 70(6), 2587-2634.
- Monika, P. and Schneider, M., 2009, "Momentum traders in the housing market: Survey evidence and a search model", The American Economic Review, 99(2), 406-11.
- Muth, J. F., 1961, "Rational expectations and the theory of price movements", Econometrica, 29(3), 315-335.
- Oikarinen, E., 2009, "Interaction between housing prices and household borrowing : The finnish case", Journal of Banking & Finance, 33, 747-756.
- Phillip, C., 1956, "The monetary dynamics of hyperinflation, in Milton Friedman (editor)", Studies in the Quantity Theory of Money, Chicago: University of Chicago Press.
- Ross, Westerfield and Jordan, 2007, "Corporate Finance Essentials", Fifth Edition.
- Stein, J. C., 1993, "Prices and trading volume in the housing market: A model with downpayment effects", Quarterly Journal of Economics, May 1995, 379-406.

(논문 접수일: 2020.10.15. 수정논문 접수일: 2020.11.23. 논문 채택일: 2020.12.11.)

The Effect of Housing and Household Characteristics on the Purchase Loan to Value

Seongwon Lee^{*}

Abstract

The aim of the study is to find out how the home purchasing capital structures of people vary as expected rate of return of house rise or fall in prices through past experience of price change. The result from the area where its expected price change have positive number appeared to have high debt capital ratios in house purchase financing. It was, in the area of expected price change is negative, it was difficult to find statistical significance between the two. In addition, net assets(-) and income(+), the age of head of household(-), housing size(+) showed similar results with a previous study. The study suggesting that local and regional characteristics like past price changes and expectations derived from past price changes can be considered as an additional element in the loan terms. This is because, household can bring up the loan financing demand in certain area where price expectations rise.

Keywords : Purchase Loan to Value, Home Purchase Financing, Capital Structure, Debt Financing, Housing Investment, Household Characteristics, Housing Characteristics

^{*} Seongwon Lee, Corresponding author, Korea Housing Finance Corporation(HF), Assistant Manager, 1944@hf.go.kr

© Copyright 2020 Housing Finance Research Institute. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

인구구조 변화와 주택가격전망을 고려한 주택연금의 중장기 수요추정

이종훈*

요 약

주택연금 제도의 안정적인 운영을 위해서는 주택연금의 주요 특성이 잘 반영된 수요추정을 전제로 정책수립 및 리스크관리가 이루어져야 한다. 본 연구에서는 주택연금 가입에 영향을 주는 주요 요인으로 인구구조의 변화와 주택가격전망을 선정하여 Bass 확산모형을 통한 수요추정을 수행하였다. 인구구조의 변화를 확산모형에 반영하기 위해 주택연금 누적 생존가입자의 확산추세를 분석하였고, 주택가격전망이 주택연금 신규 가입에 미치는 영향을 살피기 위해 확산모형을 통한 비교분석을 시행하였다. 분석 결과, 주택연금 신규 가입자 수는 2030년대 중반부터 2040년까지 정점에 이르고, 이후에는 사망자 수와 장기적인 균형을 이루어 전체 가입자는 안정적인 수준을 유지할 것으로 나타났다.

핵심어 : 주택연금, Bass 확산모형, 인구구조 변화, 주택가격전망, 수요추정

I. 서론

한국은 2000년 65세 인구 비중이 7.3%가 되어 고령화사회에 진입한 뒤 17년만인 2017년 고령사회¹⁾에 들어서 고령화 속도가 세계에서 가장 빠른 수준으로, 2067년에는 전체 인구 중 65세 이상 인구 비중이 46.5%로 증가하여 고령인구 비중이 세계에서 가장 높아질 것으로 추정된다(통계청, 2019). 또한, OECD 자료에 따르면 한국의 노인빈곤율은 43.8%로 OECD 국가 중 가장

1) UN은 65세 이상 인구가 전체 인구에서 차지하는 비중이 7% 이상이면 고령화사회, 14% 이상이면 고령사회, 20% 이상이면 초고령사회로 구분함.

* 이종훈, 주저자, 한국주택금융공사 주택연금부 대리, 1864@hf.go.kr

© Copyright 2020 Housing Finance Research Institute. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

높은 수준으로서, 이는 빠른 속도로 늘어나고 있는 고령층의 상당수가 적절한 생활 수준을 유지하는데 필요한 소득을 확보하지 못하였음을 말해준다.

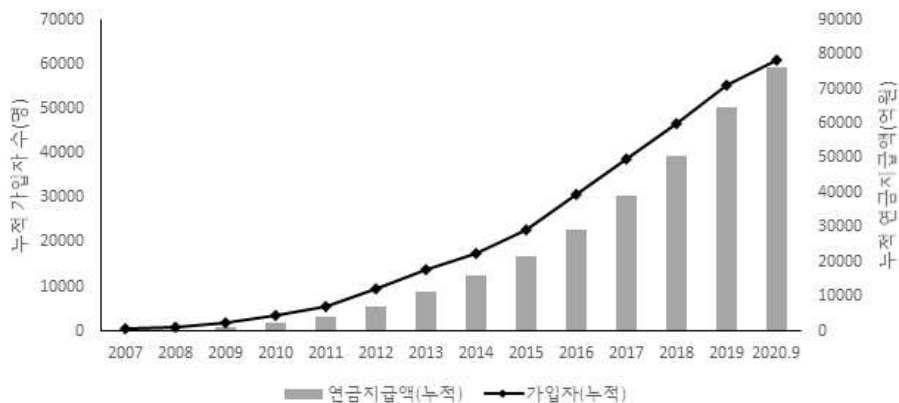
반면, 고령층은 자산의 약 77%를 부동산으로 보유하고 있는데(통계청 가계금융복지조사, 2019) 부동산의 경우 유동성이 매우 낮으므로 대다수 고령층의 노후생활을 안정시키기 위해서는 주택 등에 집중된 고령층의 자산을 유동화하여 현금흐름을 창출할 필요가 있다.

주택연금의 집은 있지만, 소득이 부족한 고령층의 노후생활 안정을 취지로 하여 2007년 7월 한국주택금융공사(이하 공사)에 의해 도입된 역모기지 상품이다. 주택연금을 통해 고령자는 보유주택에 계속 거주하면서도 이를 담보로 안정적인 연금을 받을 수 있다.

〈그림 1〉에서 볼 수 있듯이 주택연금은 도입 이후 꾸준히 성장하여 2020년 9월 말 기준 누적 가입자는 78,379명이며, 총 누적액 5.93조 원의 연금을 지급하였다. 지역별 가입현황을 보면 수도권 가입자의 비중이 68.0%로 높은 편이고, 비수도권 가입자의 비중 또한 2016년 우대형 주택연금²⁾ 도입 이후 26.4%에서 32.0%까지 증가하였다.

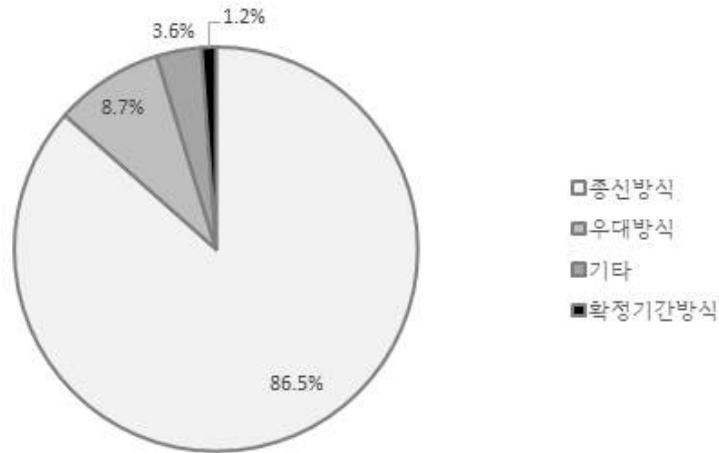
지급방식별로는 종신지급방식과 종신훘합방식이 86.5%로 가장 높으며, 우대지급방식과 우대혼합방식의 비중도 8.7%(우대형 도입 이후로 한정 시 13.8%)에 달한다. 종신훘과 확정 기간형으로 구분할 경우, 종신훘 가입자가 전체 가입자의 98.8%로서 가입자 대부분이 종신 지급을 주택연금의 중요한 특성으로 생각하고 있음을 보여준다(〈그림 2〉).

주택연금의 가장 큰 특징으로는 담보주택에 계속 거주하면서 연금을 종신지급 받으면서도 사망 이후에는 담보주택의 가치만큼만 책임을 부담하는 비소구성을 들 수 있다. 이러한 비소구성으로



〈그림 1〉 주택연금 누적 가입자 수 및 누적 연금지급액

2) 저소득 취약 고령층의 노후생활 안정을 위하여 보유주택 가격이 1.5억 원 미만인 기초연금 수급 대상자에게 일반형 주택연금 대비 최대 20%의 월지급금을 더 지급하는 상품.



〈그림 2〉 주택연금 지급방식별 누적 가입자 비율

인해 공사에게 손실이 발생할 수 있는데, 공사는 수지상등의 원칙에 따라 기대 보증료 수입과 기대 손실의 현재가치가 같아지는 월지급금을 산출하여 손실 가능성을 관리하고 있다.

이처럼 수지상등 원칙에 따라 지급되는 주택연금은 정부 재정 투입 없이도 고령층의 소득을 보완할 수 있게 해준다. 또한, 백인걸, 최경진(2019)에 따르면 주택연금의 소득대체율은 평균 70% 수준으로 공적연금의 소득대체율(31%)보다 매우 높고, 주택연금의 한계소비성향은 0.96으로 공적연금(0.76)과 일반 가구의 근로·사업소득(0.61)보다 높아 노인복지정책은 물론 경기 활성화 방안으로도 활용될 수 있다.

한편, 주택연금은 초장기 상품으로서 수지상등 원칙이 실제로 성립되기 위해서는 주택가격상승률, 이자율, 사망확률 등에 대한 장기적인 예측이 매우 중요하다. 이에 따라 한국주택금융공사법(이하 공사법)에서는 매년 1회 이상 주택가격상승률, 이자율, 사망확률을 재산정하여 신규 가입자에 대한 월지급금 산출에 반영하도록 하고 있다. 그런데 주택연금은 국민연금과 달리 가입이 의무사항이 아니므로 장기적인 재정 건전성을 점검하기 위해서는 거시경제 변수에 대한 예측과 함께 가입자 규모에 대한 예측 또한 중요하다.

본 연구의 목적은 주택연금 제도운영 및 리스크관리가 보다 정확히 이루어질 수 있도록 주택연금의 장래 중장기 수요에 대해 예측하는 것이다. 이를 위하여 본 연구에서는 다양한 산업 분야에서 신기술·신제품의 수요추정 모형으로 널리 쓰이는 Bass 확산모형을 활용하여 분석을 수행하려고 한다. 또한, 주택연금의 특징을 고려한 분석이 이루어질 수 있도록 인구구조의 변화 등을 반영한 적절한 추정 모형의 사용을 검토할 것이다.

본격적인 연구에 앞서 주택연금을 주제로 한 기존연구를 크게 분류하면 주택연금의 도입 및 제도개선 방안에 관한 연구, 주택연금 계리모형과 공급자인 공사에게 발생할 수 있는 리스크를 중심으로 한 연구, 주택연금 가입자에 대한 재무분석 및 전체 국민경제적인 효과에 관한 연구, 주택연금 가입수요 및 가입자 특성에 따른 이용행태에 관한 연구로 나눌 수 있다.

이 중 주택연금 가입수요 및 가입자 특성에 따른 이용행태에 대한 연구 중에는 설문조사 결과자료 또는 주택연금 가입자의 인구통계학적 특성을 바탕으로 한 실증분석 연구가 많았다. 기존연구에서 분석한 주택연금 가입에 영향을 미치는 요인을 살펴보면 주택 상속의향이 낮을수록 주택연금 이용 의사가 높은 것으로 나타났고(유선종, 구본영, 2005; 이선형, 김영훈, 2009; 김정주, 마승렬, 2011; 전희주, 2019), 주택가격이 하락할 것으로 예측하는 사람일수록 주택연금 이용 의사가 높게 나타났으며(김대환, 최경진, 2020), 주택가격 상승률 및 주택가격의 변동성이 낮을수록 주택연금 이용 의사가 높았다(임하나 외, 2016; 신현재, 2019). 그 외 가입가구의 연령이 낮을수록(이종의, 2010; 박근수, 김영훈, 2010; 김정주, 마승렬, 2011), 주택가격이 높을수록(이종의, 2010; 박근수, 김영훈, 2010; 김정주, 마승렬, 2011; 이달남 외, 2015) 주택연금 이용 의사가 높았으며, 서운규(2016)는 주택연금 가입조건, 월지급금, 초기보증료 등 상품의 특성 또한 유의함을 보였다.

한편, 설문조사 또는 주택연금 가입자의 인구통계학적 특성을 바탕으로 한 실증분석이 아닌 모형을 통한 수요예측 연구(민인식, 조만, 2009; 김정주, 2013; 양진아 외, 2017)도 있다. 그 중 양진아 외(2017)의 연구는 주택연금 수요추정을 위하여 마케팅 분야에서 신제품 및 신기술의 중장기 수요예측에 많이 사용되는 Bass 확산모형을 활용하여 2036년의 누적 가입자를 최소 74만 명으로 추정하였고, 양진아 외(2019)는 이용의향에 대한 응답 비율을 직접 반영하는 대신에 이용의향에 영향을 주는 다수의 요인을 포함하는 순위반응모형과 Bass 확산모형을 활용하여 2036년의 누적 가입자를 최소 86만 명으로 추정하였다.

본 연구는 주택연금의 중장기 가입 규모를 예측하기 위해 Bass 확산모형을 이용한다는 점에서 양진아 외(2017)의 연구와 관련이 있다. 다만 기존연구에서와 달리 장기간에 걸친 인구구조 변화를 잠재시장크기에 반영하였다는 점과 여러 선행연구에서 보고된 것과 같이 주택연금 가입 의사에 영향을 미치는 상속의향 및 주택가격의 변화를 고려하였다는 점에서 차이가 있다. 따라서 본 연구는 보다 현실에 가까운 수요예측을 수행하여 향후 주택연금의 경영목표 설정 및 재정 건전성에 대한 관리계획 수립에 기여할 수 있을 것으로 기대한다.

본 연구는 다음과 같이 구성된다. 제2장에서는 Bass 확산모형 및 모수 추정 방법을 소개하고, Bass 확산모형 적용 시 고려해야 할 주택연금의 특성이 반영될 수 있도록 연구에 적용할 구체적인 모형을 탐색한다. 제3장에서는 제2장에서 고려한 주택연금의 특성을 상세히 검토하여 실증분석을

위한 주요 사항들을 확정하고, 제4장에서는 이를 바탕으로 한 실증분석 결과를 정리하고, 중장기 예상수요 추정을 수행한다. 마지막으로 제5장에서는 결론 및 정책적 시사점을 제시하고, 향후 연구 방향에 대해 정리한다.

II. 연구모형의 선정

1. Bass 확산모형

Bass(1969)에 의해 제시된 확산모형은 혁신의 확산을 표현하는 대표적인 방법으로서 혁신의 채택은 대중매체 등 외부영향(혁신효과)과 구전 등에 의한 내부영향(모방효과)에 의하여 확산된다고 가정하고 이를 다음과 같은 위험함수(hazard function)의 형태로 모형화하였다.

$$\frac{f(t)}{1-F(t)} = p + qF(t) \quad \langle \text{식 1} \rangle$$

여기서 $f(t)$ 는 t 시점의 가입 확률을, $F(t)$ 는 t 시점까지의 누적 가입률을 나타내며, p 는 외부영향의 크기를 나타내는 혁신계수이고, q 는 내부영향의 크기를 나타내는 모방계수이다. 잠재시장의 크기를 m 이라고 할 때 <식 1>은 다음과 같이 표현된다.

$$\frac{n(t)}{m-N(t)} = p + q \frac{N(t)}{m} \quad \langle \text{식 2} \rangle$$

여기서 $n(t)$ 는 t 시점의 신규 가입자 수를, $N(t)$ 는 t 시점까지의 누적 가입자 수를 나타낸다. 한편, <식 1>로부터 미분방정식의 해를 구하면 다음의 식이 도출되고,

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}} \quad \langle \text{식 3} \rangle$$

<식 3>으로부터 $N(t)$ 및 $n(t)$ 를 구하면 다음과 같다.

$$N(t) = mF(t) = m \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}} \quad \langle \text{식 4} \rangle$$

$$n(t) = m \frac{(p+q)^2}{p} \frac{e^{-(p+q)t}}{(1 + \frac{q}{p} e^{-(p+q)t})^2} \quad \langle \text{식 5} \rangle$$

과거 관측자료로부터 m, p, q 의 값을 추정하여 <식 4, 5>에 대입하면 장래 t 시점의 수요를 예측할 수 있다. Bass 확산모형은 수식이 간단하면서도 설명력이 뛰어난 모형으로서 (Chandrasekaran and Tellis, 2004) 다양한 산업 분야의 수요 예측에 널리 활용되고 있다.

2. Bass 확산모형의 모수 추정 방법

Bass 확산모형의 모수 m, p, q 를 추정하는 방법을 제시한 여러 연구 중에서 대표적인 것들은 다음과 같다.

Bass(1969)는 최소제곱법(OLS : Ordinary Least Square)을 사용하여 모수를 추정하였고, Schmittlein and Mahajan(1982)과 Srinivasan and Mason(1986)은 각각 최대우도법(MLE : Maximum Likelihood Estimation)과 비선형 최소제곱법(NLS : Nonlinear Least Squares)을 이용할 것을 제안하였다(홍정식 외, 2011). 그 밖에도 Satoh(2001)는 리카티 방정식을 이용한 새로운 최소제곱법을 제안하였다.

Bass 확산모형의 모수추정 방법에 관한 국내연구로는 비교적 최근에 홍정식 외(2011)가 제시한 OLS와 NLS의 하이브리드 방법에 관한 연구가 있다. 홍정식 외(2011)는 새롭게 제시한 하이브리드 방법의 추정성능을 알아보기 위하여 시뮬레이션 데이터와 실제 제품 데이터를 이용해 하이브리드 방법과 전통적인 OLS, Satoh의 OLS, 그리고 NLS 방법으로 모수추정을 수행하여 비교분석하였다. 그 결과 전통적인 OLS와 Satoh의 OLS는 비슷한 추정성능을 보였으며 둘 다 NLS에 비해 추정성능이 떨어졌다. 하이브리드 방법의 추정성능은 NLS와 근접한 수준에서 다소 낮은 경향을 보였지만 시계열 데이터의 수가 적을 때는 NLS보다 더 나은 추정성능을 보였다. 또한, NLS 방법으로 추정에 실패한 데이터에 대해 하이브리드 방법은 추정에 성공한 다른 방법들(OLS, Satoh)보다 좋은 성능으로 추정에 성공하였다.

제3장에서 후술할 바와 같이 본 연구에서는 분석단위를 1년으로 선정하여 13개의 시계열 데이터(2007~2019년)로 분석을 수행할 예정이기 때문에 보다 안정적인 추정을 위해 하이브리드 방법을 사용하였다.

하이브리드 방법은 Bass 확산모형의 모수 m , p , q 중 m 을 고정한 상태에서 OLS 방법을 토대로 p , q 를 추정한 뒤, NLS 방법을 통하여 m 값을 추정한다. 이 경우 m , p , q 에 대한 3차원 탐색을 수행하는 NLS와는 달리 m 에 대한 1차원 탐색만을 시행하게 된다.

3. 주택연금의 특성을 반영한 연구모형 선정

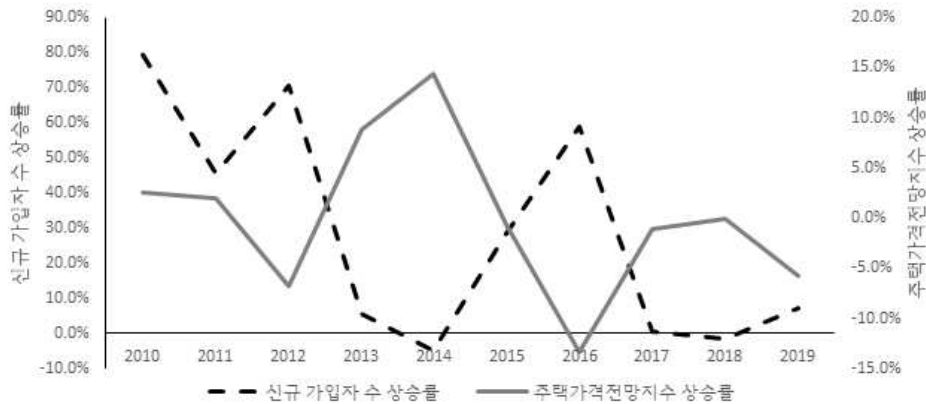
1) 주택연금의 특성

Bass 확산모형 적용에 앞서 주택연금의 특성 중 두 가지를 고려할 필요가 있다. 첫째는 주택연금은 고령층을 대상으로 하므로 인구구조의 변화에 따라 잠재시장크기(m)가 지속적으로 달라진다는 점이고, 둘째는 주택연금 가입 결정에는 과거 주택가격의 등락 및 장래 주택가격에 대한 전망이 큰 영향을 미친다는 점이다.

특히 장래 주택가격 전망과 주택연금 가입의향의 관계에 대하여 김대환, 최경진(2020)은 주택가격 하락을 예상한 사람이 그렇지 않다고 생각하는 사람에 비해 주택연금에 가입하고자 하는 의향이 높은 것을 주택연금 수요실태조사 자료를 활용하여 실증적으로 보인 바 있다. 이런 경향이 실제 주택연금 가입에도 영향을 미치는지 알아보기 위해 주택가격전망 CSI(60세 이상 대상, 한국은행) 연간 평균값의 상승률과 연간 주택연금 신규 가입자 수의 상승률을 <그림 3>으로 나타내었다.

<그림 3>을 통해 2011년 이후 주택가격전망지수의 상승률과 주택연금 신규 가입자 수의 상승률은 정확히 반대 방향의 움직임을 보임을 알 수 있다. 이를 주택연금의 월지급금과 관련하여 생각하면, 주택연금에 가입할 의사가 있더라도 향후 주택가격이 상승할 것으로 전망할 때는 더 높은 월지급금을 받기 위해 주택연금 가입 시기를 늦추고, 주택가격이 하락할 것으로 전망할 때는 주택가격이 하락하기 전에 주택연금 가입을 서두르기 때문으로 볼 수 있다.

위의 논의를 바탕으로 2011년~2020년(9월까지의 자료를 연환산)의 시계열 자료를 사용하여 연평균 주택가격전망지수(60세 이상) 상승률(X)이 연도별 주택연금 신규 가입자 수 상승률(Y)에 미치는 영향을 알아보기 위해 단순회귀분석을 수행하였다(<표 1>).



〈그림 3〉 주택가격전망지수의 상승률과 주택연금 신규 가입자 수 상승률 추세

〈표 1〉 주택가격전망지수 상승률과 주택연금 신규 가입자 수 상승률의 단순회귀분석 결과

	<i>B</i>	<i>S.E.</i>	<i>t</i>	<i>p</i>	95% C.I.		<i>F</i>	<i>R</i> ²
					Lower	Upper		
상수	0.2230	0.07068	3.155	0.0135	0.0600	0.3860	7.4014*	0.481
독립변수	-2.3082	0.84843	-2.721	0.0262	-4.2647	-0.3517		

* $p < .05$, ** $p < .01$, *** $p < .001$

분석 결과, $F = 7.4014$ ($p < .05$)로서 위 회귀모형이 적합하다고 할 수 있으며, $R^2 = 0.481$ 로 모형의 설명력은 48.1%로 나타났다. 주택가격전망지수 상승률의 회귀계수는 $B = -2.3082$ ($p < .05$)로서 연평균 주택가격전망지수(60세 이상)의 상승률은 연도별 주택연금 신규 가입자 수 상승률에 부(-)의 영향을 미치는 것으로 나타났다. 이는 주택가격이 상승할 것으로 전망하는 고령자의 비중이 전년보다 높을수록 주택연금 신규 가입자 수가 전년보다 많이 증가하지 못함을 의미한다. 한편, 〈표 1〉 회귀모형의 회귀식 $Y = 0.2230 - 2.3082X$ 를 활용하면 주택가격전망지수의 변화에 따른 영향을 제거한 주택연금 신규 가입자 수를 추정할 수 있으며, 그 결과는 제4장에 제시하였다.

2) 연구모형 선정

Bass(1969)는 잠재시장크기를 나타내는 m 을 고정된 상수로 가정했지만, Mahajan and Peterson(1978), Sharif and Ramanathan(1981)의 연구에서는 m 을 가구수나 인구수에 따라 변한다고 보고 확산모형을 적용하였으며(Mahajan et al., 1990), 채현석 외(2016) 또한 전기

자동차 수요예측에 관한 연구에서 인구수의 변화에 따라 m 이 변한다고 가정하였다.

앞서 살핀 바와 같이 인구구조 변화에 따라 주택연금 가입대상 가구의 규모가 변하기 때문에 본 연구에서는 기존연구들과 마찬가지로 잠재시장크기를 시간의 함수 $m(t)$ 로 가정하였다. 이 경우, $m(t)$ 는 t 시점의 60세 이상 가구수를 바탕으로 산출하게 된다. 또한, 사망자는 가구수 산정 대상에서 애초부터 제외되기 때문에 잠재시장에서 제외되며, 마찬가지로 누적 가입자 수에서도 제외된다. 따라서 누적 가입자 수 $N(t)$ 는 t 시점에서 생존해 있는 가입자 수로 보아야 한다. $N(t)$ 를 누적 생존가입자의 수로 놓는 데는 다음 두 가지 장점이 있다.

① Bass 확산모형의 기본 가정에 더욱 충실해진다. q 는 구전 등에 의한 모방계수로서 t 시점의 신규 가입자 수는 t 시점까지의 누적 가입자 수에 비례하는 모방효과를 받게 되는데 t 시점 전에 이미 사망한 가입자에 의한 모방효과는 거의 기대하기 어렵다는 점을 모형에 반영할 수 있다.

② 보다 현실적인 수요예측이 가능해진다. $N(t)$ 를 사망자를 포함한 누적 가입자로 놓고 Bass 확산모형에 따라 수요를 예측하면 t 가 증가함에 따라 신규 가입자 수 $n(t)$ 는 0에 수렴하게 된다. 그런데 주택연금의 가입자는 기간 경과에 따라 지속적으로 사망하므로, 신규 가입자 수가 0에 수렴하면 주택연금의 유지건수 또한 0에 수렴하게 되어 현실과 맞지 않는 수요예측을 하게 된다. $N(t)$ 에서 사망자를 제외할 경우, $n(t)$ 는 당해연도 신규 가입자 수에서 당해연도 사망자 수를 뺀 값이 되고, $n(t)$ 가 0으로 수렴하는 것은 장기적으로 신규 가입자 수와 사망자 수가 균형을 이룬다는 결론에 도달하게 된다.

한편, 주택연금을 주택을 제공하고 현금을 받는 거래로 보게 되면 가입자는 본인이 제공하는 주택의 가치를 연금을 받기 위해 지불하는 가격이라고 느낄 수 있다. 이 경우 주택가격의 전망은 가입자가 지불하는 가격에 대한 전망으로도 해석할 수 있고, 주택연금 가입 후 단기간 내에 본인이 제공한 주택의 가치가 상승한다면 연금을 받기 위해 적정 가격보다 비싸게 주고 거래를 했다고 느낄 수도 있다. 따라서 향후 1년간의 주택가격전망을 나타내는 주택가격전망 CSI 지수를 연금을 받기 위해 지불하는 소비자의 체감 가격을 표상하는 지수로 보는 것도 가능하다. 이와 같은 해석은 앞에서 다룬 주택가격전망 지수의 상승률과 주택연금 신규 가입자 상승률의 관계와도 부합한다.

이렇게 주택가격전망지수를 소비자의 체감 가격으로 간주하고 Bass, Krishnan and Jain (1994)이 제안한 GBM(Generalized Bass Model)을 변형하여 사용하는 방법 또한 고려할 수 있다. 양진아, 민대기(2017) 또한 GBM을 이용하여 주택연금의 수요를 추정한 바 있다.

다만, 주택가격전망은 Bass et al.(1994)에서 고려한 가격·광고와 달리 상품 공급자인 공사가 정책적으로 결정할 수 있는 변수가 아니고, 앞서 확인한 바 있는 주택가격전망 상승률과 신규 가입자 상승률의 선형함수를 Bass 모형에 포함할 경우 모형의 복잡성이 매우 증가하므로 본 연구에서는 이를 모형에 포함하지 않았다.

대신, 2011~2019년 동안 주택가격전망의 상승률을 0%로 가정하고, 이에 따라 같은 기간의 신규 가입자 수를 보정한 수요추정 모형과 실제 가입자 수를 그대로 사용한 수요추정 모형의 비교를 통해 주택가격전망이 수요예측에 미치는 영향을 살펴보았다.

Ⅲ. 연구방법

1. 분석단위 설정

주택연금 생존가입자 수에 대해 Bass 확산모형을 적용하기에 앞서 적절한 분석단위를 정하기 위해 월별 주택연금 신규 가입자 수 및 부부 모두가 사망하여 주택연금을 해지한 가입자(이하 사망해지자) 수의 추세를 살펴볼 필요가 있다.

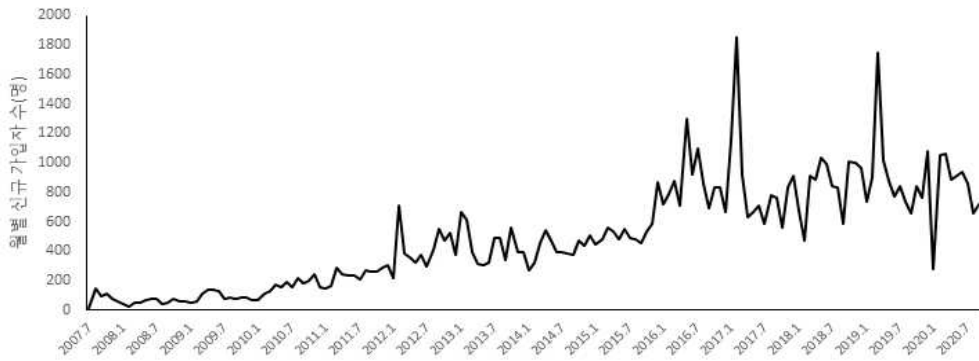
〈그림 4〉의 월별 주택연금 신규 가입자 수 추세에서는 다음과 같은 특징을 볼 수 있는데, 이는 특히 연간 신규 가입자가 1만 명 대에 진입한 2016년 이후부터 더욱 두드러진다. 참고로, 〈그림 4〉의 세로축은 월별 주택연금 신규 보증서 발급 건수이므로 이를 해석할 때는 주택연금 가입신청부터 보증서 발급까지는 통상 1개월 정도의 시차가 있음을 고려할 필요가 있다.

① 신상품 출시 등 제도개선 이후 신규 가입자 수의 증가

‘내집연금 3종세트’가 출시되었던 2016년 4월과 우대형 주택연금에 대한 지원이 확대된 2019년 12월 이후 신규 가입자가 증가하였다.

② 월지급금 조정 직전 시점에서 신규 가입자 수의 급격한 변화

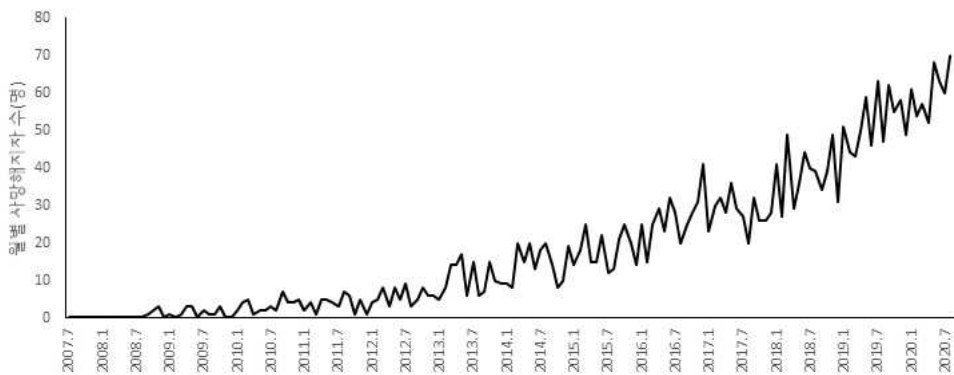
주택연금 월지급금은 2007년 7월부터 2011년까지는 변동이 없었으나, 2012년부터는 주택시장, 금융시장 및 사망확률의 변동을 반영하여 매년 초 월지급금을 조정하고 있다. 이에 따라 직전 연도보다 월지급금이 감소하는 경우에는 월지급금 변경 직전에 신규 가입자가 증가하고, 직전 연도보다 월지급금이 증가하는 경우에는 월지급금 변경 직전에 신규 가입자가 감소하였다.



〈그림 4〉 월별 주택연금 신규 가입자 수 추세

한편, 주택연금 가입자의 월별 사망해지 추세는 지속적으로 증가하고 있다는 점 외에 다른 뚜렷한 특징은 보이지 않는다. 사망해지가 지속적으로 증가하는 것은 누적 가입자 수의 증가 및 전체 가입자의 평균 이용기간의 증가에 따른 현상이다(〈그림 5〉).

이와 같은 특성을 고려하여 본 연구에서는 매년 초 월지급금 조정에 의한 신규 가입자 수의 급격한 변동으로 인하여 확산추세가 왜곡되지 않도록 분석단위를 1년으로 설정하였다. 그리고 이에 따라 분석대상 시계열 자료의 범위는 2007~2019년으로 설정하였다.



〈그림 5〉 월별 주택연금 사망해지 수 추세

2. 주택연금 누적 생존가입자 수의 산출

주택연금 누적 생존가입자 수는 누적 가입자 수와 누적 사망해지자 수의 차이로서 다음과 같이 표현할 수 있다.

$$N(t) = N_0(t) - N_d(t) \quad \langle \text{식 6} \rangle$$

$N(t)$: t 시점의 누적 생존가입자 수

$N_0(t)$: t 시점의 누적 가입자 수

$N_d(t)$: t 시점의 누적 사망해지자 수

한편, 주택연금의 해지는 크게 사망해지와 중도해지로 구분할 수 있다. 사망해지는 가입자와 배우자가 모두 사망하여 연금을 받을 사람이 없게 될 때 발생하고, 중도해지는 주택처분, 가족의 반대 등 개인적인 사정에 따라 연금을 그만 받고자 할 때 발생한다.

그런데 Bass 확산모형은 혁신에 대한 수용자의 최초 구매에 따른 확산추세를 나타내는 모형으로서, 주택연금 중도해지자는 추후 개인적인 사정으로 인해 이용을 중단하였을 뿐, 주택연금 전문상담 인력으로부터 주택연금 제도에 대한 충분한 정보를 획득하고 스스로 의사결정에 따라 주택연금에 가입하였다는 점에서 혁신 수용자에서 제외할 이유가 없다. 한편, 가입자가 중도해지 이후 재가입하게 되면 누적 생존가입자 수에 중복으로 계산될 우려가 있다. 다만, 중도해지 시에는 주택가격의 1.5%에 달하는 초기보증료를 환급하지 않는 점과 중도해지 후 3년간은 주택가격에 따라 재가입이 제한될 수 있는 점을 생각하면 중도해지 후 재가입하는 경우는 일반적이지 않은 것으로 볼 수 있으므로 본 연구에서는 이를 별도로 고려하지 않았다.

본 연구의 주요 목적이 인구구조의 변화에 따라 시점마다 달라지는 잠재시장크기($m(t)$)를 확산모형에 반영하여 누적 생존가입자의 비율을 추정하기 위한 점임을 고려하면, 중도해지자도 누적 가입자에 포함하되, 중도해지자가 해지 후 사망하는 경우에는 누적 가입자에서 제외하여야 한다. 다만, 중도해지 이후에는 공사가 가입자의 사망 여부를 추적하는 것이 불가능하므로 중도해지자 중에서 중도해지 이후에 사망한 사람의 수를 추정하기 위하여 전체 주택연금 가입자를 다음과 같이 두 그룹으로 나누었다.

A : 2019년 12월 31일까지 중도해지하지 않은 가입자

B : 2019년 12월 31일 이전에 중도해지한 가입자

이 경우, 사망해지자는 A 그룹에 속하며, B 그룹에 속한 가입자의 시점별 사망률은 A 그룹의 시점별 사망률과 같다고 가정하고 다음과 같이 B 그룹의 사망자 수를 추정하였다.

$$N_{B,d}(t) = \frac{N_{A,d}(t)}{N_{A,0}(t)} \times N_{B,0}(t) \quad \langle \text{식 7} \rangle$$

$N_{A,0}(t)$: A그룹의 t 시점 누적 가입자 수

$N_{A,d}(t)$: A그룹의 t 시점 누적 사망해지자 수

$N_{B,0}(t)$: B그룹의 t 시점 누적 가입자 수

$N_{B,d}(t)$: B그룹의 t 시점 누적 사망해지자 수

〈식 7〉을 〈식 6〉에 대입하면 다음의 식을 얻을 수 있다.

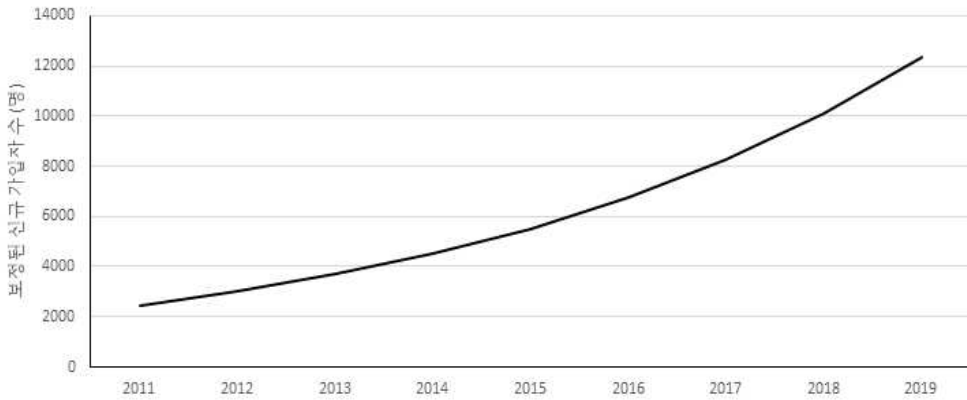
$$\begin{aligned} N(t) &= N_A(t) + N_B(t) \\ &= N_{A,0}(t) - N_{A,d}(t) + N_{B,0}(t) - \frac{N_{A,d}(t)}{N_{A,0}(t)} \times N_{B,0}(t) \end{aligned} \quad \langle \text{식 8} \rangle$$

1) 과거 주택가격전망이 일정함을 가정하는 경우

제2장에서 논의한 바와 같이 2011~2019년의 주택가격전망 상승률이 0%임을 가정하고 앞서 확인한 선형관계로부터 해당 기간의 신규 가입자 수를 보정한 결과는 〈표 2〉 및 〈그림 6〉과 같다.

〈표 2〉 보정된 신규 가입자 수

연도	$\tilde{n}_0(t)$ (명)
2011	2,466
2012	3,015
2013	3,688
2014	4,510
2015	5,516
2016	6,746
2017	8,250
2018	10,090
2019	12,340



〈그림 6〉 보정된 신규 가입자 수 추세

3. 잠재시장크기 산출

2019년 말을 기준으로 주택연금은 부부 중 1인 이상이 60세 이상인 자가보유자로서 실제로 거주하고 있는 시가 9억 원 이하의 주택을 담보로 하여 가입할 수 있다.³⁾ 따라서, 주택연금의 잠재시장 크기는 이러한 가입기준을 만족하는 가구수를 기준으로 산출해야 하며, 이는 다음과 같은 식으로 나타낼 수 있다.

$$\begin{aligned} m(t) &= H(t) \times \lambda_1(t) \times \lambda_2(t) \times c \\ &= M(t) \times c \end{aligned} \quad \langle \text{식 9} \rangle$$

$H(t)$: t 시점의 가구주 연령 60세 이상 가구수

$\lambda_1(t)$: t 시점의 가구주 연령 60세 이상 가구 중 자가거주 가구 비율

$\lambda_2(t)$: t 시점의 가구주 연령 60세 이상 자가거주 가구 중 거주주택 가격이 9억원 이하인 가구 비율

c 는 상수($c \in (0,1)$)로서 주택연금 가입기준을 만족하는 가구 중 주택연금 이용의향을 잠재적으로 보유한 가구의 비율을 뜻한다.

3) 참고로, 2020년 4월 1일부터 부부 중 1인 이상이 55세 이상이면 가입할 수 있게 되었으며, 공사법 개정으로 2020년 12월 8일부터 공시가격 9억 원 이하의 주택까지 가입이 허용되었다.

1) 가구주 연령 60세 이상 가구수 산출

가구주 연령이 60세 이상인 가구수는 인구총조사(통계청) 가구부문 자료를 활용하여 산출하였다. 인구총조사(등록센서스 방식 포함)가 시행되지 않은 2007~2009년, 2011~2014년의 60세 이상 가구수는 2005년, 2010년, 2015년의 60세 이상 가구수를 선형보간하여 추정하였다. 그 결과는 <표 3> 및 <그림 7>과 같다.

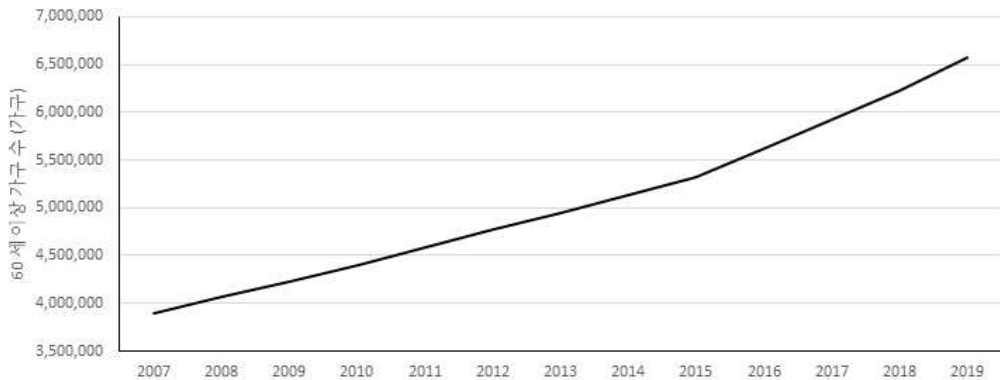
가구주 연령이 60세 이상인 가구수는 지속적으로 증가하여 2019년에는 2007년 대비 68.6% 증가하였고, 증가량도 매년 높아지고 있는 것을 알 수 있다. 이는 서론에서 언급한 바 있던 한국의 급격한 고령화 속도의 영향을 받은 것으로 보인다.

2) 자가거주 비율 및 주택가격 9억 원 이하 가구 비율 산출

가구 주 연령이 60세 이상인 가구 중 자가거주 가구의 비율과 이 중 거주주택 가격이 9억원 이하인 가구의 비율은 가계금융복지조사(통계청) 자료를 활용하여 산출하였다. 조사가 시행되지 않은 해인 2007~2010년의 비율은 선형보간법을 이용하여 추정하였다. 그 결과는 <표 4> 및 <그림 8>과 같다.

<표 3> 가구주 연령 60세 이상 가구수

연도	60세 이상 가구 수
2007	3,897,312
2008	4,063,718
2009	4,230,123
2010	4,396,528
2011	4,581,657
2012	4,766,786
2013	4,951,916
2014	5,137,045
2015	5,322,174
2016	5,622,780
2017	5,921,822
2018	6,230,137
2019	6,569,763

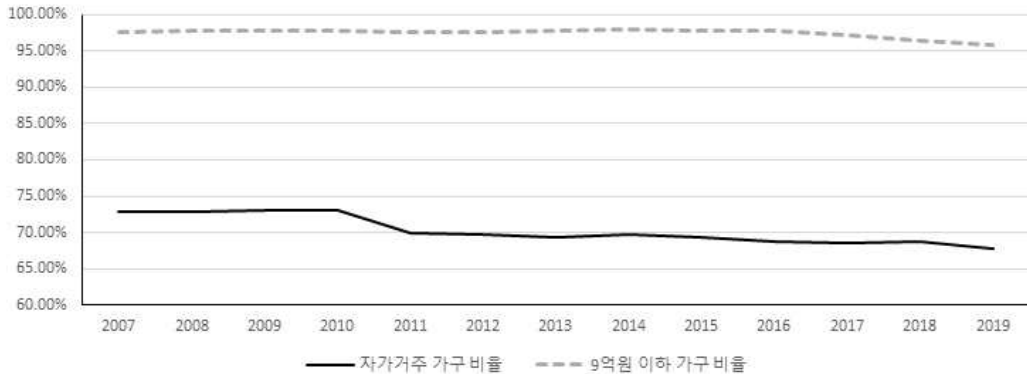


〈그림 7〉 가구주 연령 60세 이상 가구수 추세

〈표 4〉 자가거주 가구 및 주택가격 9억 원 이하 가구 비율

연도	자가거주 가구 비율(%)	9억 원 이하 가구 비율(%)
2007	72.8	97.6
2008	72.9	97.7
2009	73.0	97.7
2010	73.1	97.8
2011	69.9	97.5
2012	69.8	97.6
2013	69.3	97.8
2014	69.7	98.0
2015	69.4	97.7
2016	68.8	97.7
2017	68.6	97.1
2018	68.8	96.4
2019	67.8	95.9

60세 이상 가구 중 자가거주 가구의 비율 및 이 중 주택가격 9억 원 이하 가구 비율은 분석 기간동안 각각 약 68~73%, 96~98%로 일정한 수준임을 알 수 있다.



〈그림 8〉 자가거주 가구 및 주택가격 9억 원 이하 가구 비율 추세

3) 주택연금 이용의향 잠재보유 가구 비율 산출

하이브리드 방법에 따라 Bass 확산모형의 모수를 추정하기 위해서는 잠재시장크기의 상한과 하한을 정할 필요가 있다. 〈식 9〉에서 $H(t)$, $\lambda_1(t)$, $\lambda_2(t)$ 는 관측치에 의해 확정적으로 결정되므로 $m(t)$ 의 상한과 하한을 정하기 위해 c 의 상한($\bar{c}$)과 하한($\underline{c}$)을 각각 정하였다. $\bar{c}$ 와 $\underline{c}$ 를 산출하기 위해 공사가 실시하는 주택연금 수요실태조사(2015~2019) 자료를 활용하여 주택연금 이용의향을 잠재적으로 보유한 가구의 비율을 추정하였다.

제1장에서 기존의 연구들로부터 상속의향이 주택연금 가입에 영향을 미친다는 점을 알 수 있었다. 주택연금 가입 시에 보유주택을 상속받지 못할 수도 있다는 점에 대해서 자녀들과 충분한 상의를 거치고 가입이 이루어지는 것이 일반적이므로, 주택연금 이용의향뿐만 아니라, 보유주택 상속의향을 동시에 고려하여 잠재적 이용의향을 산출하였다. 주택연금 이용의향과 보유주택 상속의향에 대한 응답이 배치되는 경우에는 주택연금 가입에 대한 의사결정에 있어 자녀의 영향이 크고, 가입 의사결정 자체가 매우 신중하게 이루어지는 점을 고려하여 상속의향을 우선으로 고려하였다.

또한, 주택연금에 대한 인지가 거의 없음에도 불구하고, 막연하게 주택연금 이용/비이용 의향을 밝힌 응답자는 실제 가입으로 이루어질지를 판단하기 어렵다는 점과 후술할 바와 같이 본 연구에서는 장기적으로 전체 고령자가 주택연금에 대한 대략적인 인지 수준을 갖출 것으로 가정하고 있는 점을 고려하여 주택연금에 대한 대략적인 인지 수준을 갖춘 응답자만을 대상으로 비율을 산출하였다. 이를 위해 전체 응답자를 다음과 같이 분류하였다.

U_t : t 시점 조사의 전체 응답자 집합

C_t : t 시점 조사의 응답자 중 주택연금 인지 수준을 묻는 문항에서 ‘2. 주택연금이 무엇인지 정도만 알고 있다’, ‘3. 주택연금 가입이 가능한지 여부를 알고 있다’, ‘4. 받을 수 있는 대략의 연금 수준을 알고 있다’, ‘5. 방문, 전화상담, 홈페이지 방문 등을 통해 문의해서 잘 알고 있다’를 선택한 응답자 집합

$\overline{D}_t$: t 시점 조사의 응답자 중 주택연금 이용의향을 묻는 문항에서 ‘4. 이용하겠다’, ‘5. 반드시 이용하겠다’를 선택하고, 보유주택 상속의향을 묻는 문항에서 ‘2. 보유하고 있는 주택을 일부 물려주겠다’, ‘3. 전혀 물려주지 않겠다’를 선택한 응답자 집합

$\underline{D}_t$: t 시점 조사의 응답자 중 주택연금 이용의향을 묻는 문항에서 ‘4. 이용하겠다’, ‘5. 반드시 이용하겠다’를 선택하고, 보유주택 상속의향을 묻는 문항에서 ‘3. 전혀 물려주지 않겠다’를 선택한 응답자 집합

본 연구에서는 주택연금 가입 가능 가구 중 잠재적인 이용의향을 가진 가구의 비중은 시점과 관계없이 항상 일정하다는 가정하에 $\bar{c}$ 와 $\underline{c}$ 를 상수로 두었으며, 이 경우 주택연금 수요실태조사를 통해 확인되는 $\frac{n(C_t \cap \overline{D}_t)}{n(U_t)}$ 및 $\frac{n(C_t \cap \underline{D}_t)}{n(U_t)}$ 값은 혁신의 확산에 따라 증가하여 각각 $\bar{c}$ 와 $\underline{c}$ 에 수렴하는 것으로 생각할 수 있다. 이를 식으로 표현하면 다음과 같다.

$$\bar{c} = \lim_{t \rightarrow \infty} \frac{n(C_t \cap \overline{D}_t)}{n(U_t)} = \lim_{t \rightarrow \infty} \frac{n(C_t)}{n(U_t)} \times \lim_{t \rightarrow \infty} \frac{n(C_t \cap \overline{D}_t)}{n(C_t)} = \lim_{t \rightarrow \infty} P(C_t) \times \lim_{t \rightarrow \infty} P(\overline{D}_t | C_t) \quad \langle \text{식 10} \rangle$$

$$\underline{c} = \lim_{t \rightarrow \infty} \frac{n(C_t \cap \underline{D}_t)}{n(U_t)} = \lim_{t \rightarrow \infty} \frac{n(C_t)}{n(U_t)} \times \lim_{t \rightarrow \infty} \frac{n(C_t \cap \underline{D}_t)}{n(C_t)} = \lim_{t \rightarrow \infty} P(C_t) \times \lim_{t \rightarrow \infty} P(\underline{D}_t | C_t) \quad \langle \text{식 11} \rangle$$

한편, 2015년부터 2019년까지의 조사 결과를 바탕으로 시점별 $P(C_t)$ 와 $P(\overline{D}_t | C_t)$, $P(\underline{D}_t | C_t)$ 를 산출한 결과는 <표 5>와 같다.

전체 가입대상 가구 중 주택연금에 대한 대략적인 인지 수준을 갖춘 가구의 비중을 나타내는 $P(C_t)$ 는 지속적으로 증가하고 있으며, 본 연구에서는 향후 제도 운영 기간이 길어짐에 따라 공사의 홍보 노력 등으로 인해 고령층 대부분이 주택연금에 대한 대략적인 인지 수준을 갖추게 되어 $P(C_t)$ 가 장기적으로 1에 수렴한다고 가정하였다. 또한 <표 5>에서 관찰 가능한 가장 최근

〈표 5〉 시점별 주택연금 수요실태조사 분석 결과

t	2015	2016	2017	2018	2019
$P(C_t)$	0.5807	0.6647	0.7130	0.7133	0.7767
$P(\overline{D}_t C_t)$	0.1297	0.1235	0.1543	0.1506	0.1765
$P(\underline{D}_t C_t)$	0.0855	0.0794	0.1070	0.0992	0.1207

값인 0.1765와 0.1207을 $P(\overline{D}_t|C_t)$ 와 $P(\underline{D}_t|C_t)$ 가 수렴하는 값으로 가정하였다. 이를 〈식 10, 11〉에 대입하여 $\bar{c}=0.1765$, $\underline{c}=0.1207$ 을 도출하였다.

4. 장래 신규 가입자 수 추정 방법

장래 시점 기간별 신규 가입자 수($n_0(t)$, $t > 2020$)는 다음의 식과 같이 누적 생존가입자 수($N(t)$)와 신규 사망해지자 수($n_d(t)$)로부터 산출할 수 있다.

$$\begin{aligned}
 n_0(t) &= N_0(t) - N_0(t-1) \\
 &= (N(t) + N_d(t)) - (N(t-1) + N_d(t-1)) \\
 &= (N(t) - N(t-1)) + n_d(t)
 \end{aligned}
 \quad \langle \text{식 12} \rangle$$

1) $N(t)$ 의 추정

하이브리드 방법에 따라 추정한 모수 p , q 를 〈식 3〉에 대입하여 $F(t)$ 를 추정할 수 있다. 〈식 4〉에 따라 $F(t)$ 로부터 $N(t)$ 를 추정하기 위하여, 2047년까지 추세치가 발표된 장래가구추계(통계청)로부터 산출한 60세 이상 가구수($H(t)$)에 2019년의 자가거주 비율(67.8%) 및 주택가격 9억 원 이하 비율(95.9%)을 적용하여 $m(t)$ 를 추정하였다.

2) $n_d(t)$ 의 추정

$t-1$ 시점 기준 생존가입자 평균연령(a_{t-1})에 대한 장래생명표(통계청) 상 t 년도 사망확률

$(\delta_{t,a_{t-1}})$ 을 $t-1$ 시점 누적 생존가입자 수($N(t-1)$)에 곱하여 $n_d(t)$ 를 추정할 수 있다. 이 경우 a_{t-1} 은 신규 가입자의 가입 시점 평균연령을 72세(2019년 신규 가입자의 평균연령)로 놓고 다음과 같이 구할 수 있다.

$$a_{t-1} = \frac{(a_{t-2} + 1) \times (N(t-1) - n_0(t-1)) + 72 \times n_0(t-1)}{N(t-1)} \quad \langle \text{식 13} \rangle$$

IV. 분석 결과

1. 모수추정 결과

하이브리드 방법을 사용하여 $\underline{c}=0.1207$ 부터 $\bar{c}=0.1765$ 까지 0.0001씩 증가시키며 모수를 추정한 결과, 오차제곱합은 $c=0.1207$, $p=0.004398$, $q=0.141516$ 일 때 1314.9로 가장 작았다.

혁신계수(p)가 모방계수(q)보다 작은 것은 Bass 확산모형을 이용한 기존 대부분의 연구와 일치한다. 본 연구에서 q 와 p 의 비는 32.2로서 Bass 확산모형으로 주택연금의 수요를 추정한 기존연구(양진아 외 2인, 2017)의 비율(508.2)보다 낮았다. 이는 기존연구 이후의 기간 경과로 2014~2019년의 시계열 자료를 분석대상에 포함하게 된 결과, 해당 기간의 다양한 제도개선⁴⁾에 따른 혁신효과가 가입 결정에 반영된 것으로 보인다. 이를 확인하기 위해 본 연구에서도 2007~2013년의 시계열 자료만을 대상으로 별도 분석한 결과, q 와 p 의 비가 189.4로 높아지는 것을 알 수 있었다.

한편, <표 2>의 자료를 반영하여 주택가격전망에 의한 영향을 제거하고 모수를 추정한 결과 오차제곱합은 $c=0.1207$, $p=0.002848$, $q=0.172389$ 일 때 545.3으로 가장 작았으며, 이는 보정하지 않은 실제 관측치로 추정한 결과(1314.9)보다 낮았다. 모수추정결과에 따른 관측치와 추정치의 비교는 <표 6, 7> 및 <그림 9, 10>과 같다.

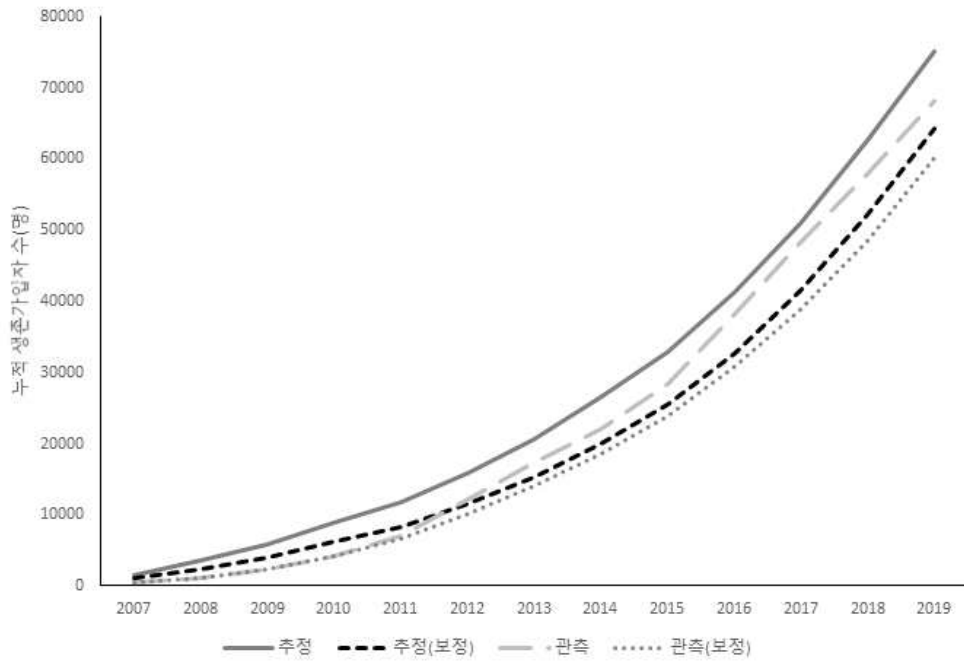
4) 초기보증료를 인하, 전후후박형 주택연금 출시, 가입연령 충족(60세 이상) 대상을 주택소유자에서 주택소유자와 배우자 중 연장자로 완화, 내집연금 3종세트 출시, 상환용 주택연금 인출한도 확대, 우대형 지원확대 등

〈표 6〉 관측치와 추정치 비교(실제 관측치 사용; 단위: 명)

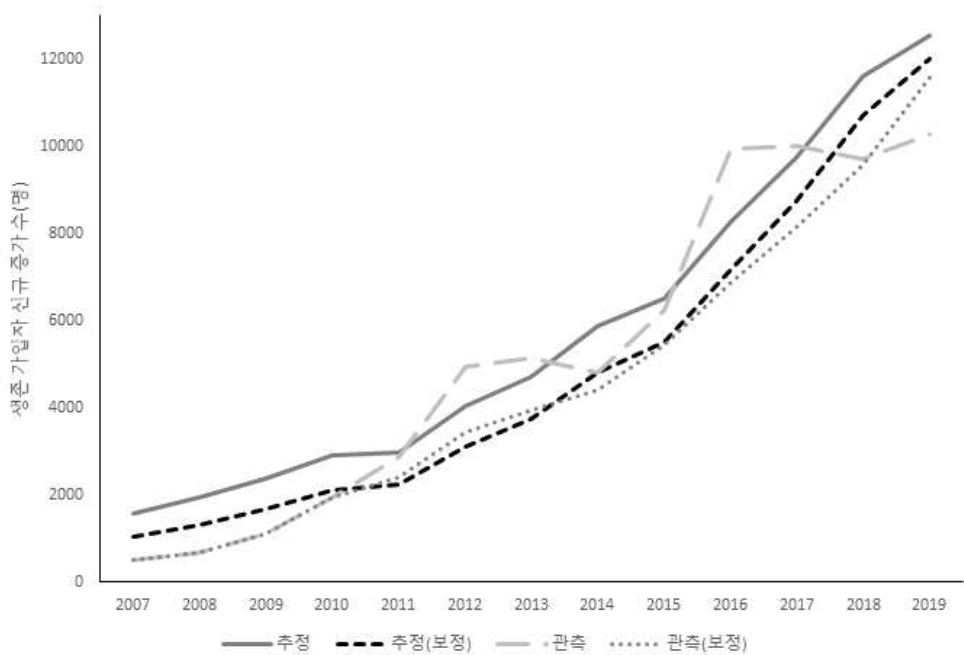
연도	누적(관측)	누적(추정)	신규(관측)	신규(추정)
2007	516	1,576	516	1,576
2008	1,203	3,532	687	1,955
2009	2,310	5,930	1,106	2,399
2010	4,272	8,846	1,962	2,915
2011	7,154	11,817	2,882	2,971
2012	12,098	15,859	4,945	4,042
2013	17,250	20,581	5,152	4,722
2014	22,081	26,460	4,830	5,879
2015	28,315	32,969	6,234	6,510
2016	38,254	41,254	9,939	8,285
2017	48,269	51,014	10,014	9,760
2018	57,980	62,623	9,712	11,609
2019	68,273	75,180	10,292	12,556

〈표 7〉 관측치와 추정치 비교(보정 관측치 사용; 단위: 명)

연도	누적(관측)	누적(추정)	신규(관측)	신규(추정)
2007	516	1,038	516	1,038
2008	1,203	2,367	687	1,329
2009	2,310	4,048	1,106	1,682
2010	4,272	6,156	1,962	2,108
2011	6,683	8,392	2,411	2,236
2012	10,136	11,503	3,453	3,111
2013	14,068	15,257	3,932	3,754
2014	18,481	20,061	4,413	4,805
2015	23,929	25,582	5,449	5,520
2016	30,813	32,776	6,883	7,194
2017	38,957	41,516	8,144	8,740
2018	48,533	52,214	9,576	10,699
2019	60,115	64,227	11,582	12,012



〈그림 9〉 누적 생존가입자 수 추세



〈그림 10〉 연도별 생존가입자 신규증가 수 추세

2. 장래 수요추정 결과

제3장에서 논의한 방법에 따라 장래 수요를 추정한 결과는 <표 8>과 같다.

한편, <표 2>의 자료를 반영하여 주택가격전망에 의한 영향을 제거한 장래 수요추정은 <표 9>와 같다.

<표 8> 장래 수요 추정(실제 관측치 사용; 단위: 명)

연도	누적 생존가입자	당기 사망	당기 신규 가입
2021	109,097	1,220	20,134
2022	129,457	1,513	21,873
2023	152,959	1,834	25,336
2024	177,541	2,209	26,791
2025	205,501	2,619	30,579
2026	235,819	3,101	33,419
2027	268,337	3,632	36,150
2028	303,421	4,200	39,284
2029	342,446	4,848	43,873
2030	382,637	5,545	45,736
2031	425,255	6,355	48,973
2032	468,396	7,168	50,309
2033	512,231	8,061	51,895
2034	555,873	9,096	52,738
2035	597,385	10,177	51,689
2036	636,910	11,300	50,825
2037	674,617	12,625	50,332
2038	711,182	14,008	50,573
2039	746,447	15,487	50,753
2040	782,739	17,188	53,480
2041	817,677	18,898	53,836
2042	850,707	20,704	53,734
2043	879,443	22,631	51,366
2044	903,153	24,757	48,468
2045	922,177	27,247	46,271
2046	939,208	29,838	46,869
2047	953,284	32,315	46,391

〈표 9〉 장래 수요 추정(보정 관측치 사용)

연도	누적 생존가입자(명)	당기 사망(명)	당기 신규 가입(명)
2021	97,787	1,152	20,009
2022	118,766	1,419	22,398
2023	143,509	1,718	26,461
2024	170,163	2,079	28,733
2025	200,936	2,479	33,252
2026	234,864	2,956	36,884
2027	271,730	3,495	40,360
2028	311,797	4,082	44,149
2029	356,362	4,754	49,319
2030	402,380	5,467	51,486
2031	450,949	6,327	54,896
2032	499,837	7,204	56,092
2033	549,007	8,165	57,335
2034	597,334	9,238	57,565
2035	642,606	10,421	55,692
2036	684,907	11,656	53,957
2037	724,414	12,999	52,507
2038	761,897	14,527	52,011
2039	797,265	16,096	51,464
2040	833,093	17,945	53,772
2041	866,939	19,839	53,686
2042	898,333	21,797	53,191
2043	924,884	23,932	50,482
2044	945,967	26,256	47,339
2045	962,075	29,019	45,127
2046	976,122	31,804	45,851
2047	987,183	34,550	45,611

두 경우 모두 누적 생존가입자($N(t)$)의 증가는 2033년을 정점으로 하여 둔화되는 추세를 보인다. 다만, 누적 생존가입자 수의 증가($n(t)$)는 당기 신규 가입자($n_0(t)$)에서 당기 사망자($n_d(t)$)의 숫자를 차감한 것과 같고, 기간의 경과에 따라 당기 사망자 또한 증가하므로 당기 신규 가입자 수는 2030년대에 정점에 이른 뒤 2040년대 중반부터는 안정화되는 모습을 보인다. 2040년대 중반

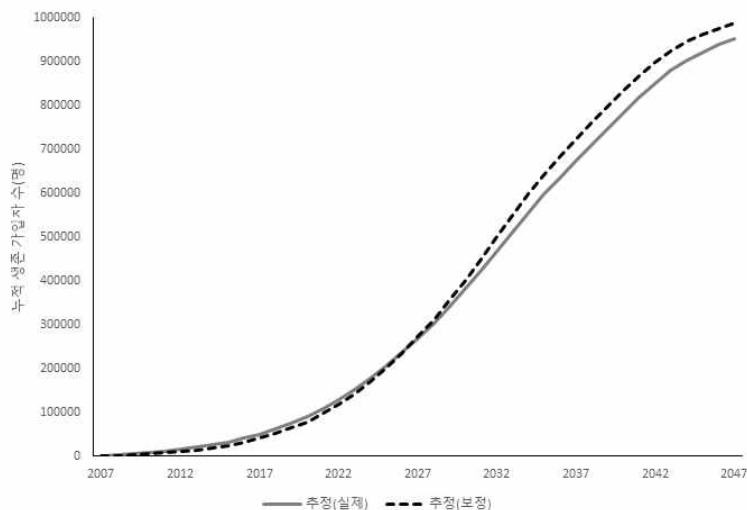
이후부터는 장기적으로는 당기 사망자와 당기 신규 가입자가 균형을 이루어 누적 생존가입자 수는 일정한 수준에서 유지될 전망이다.

한편, 주택가격전망을 반영하여 보정된 관측치를 사용하여 추정한 결과와 실제 관측치를 사용하여 추정한 결과를 비교하면 2021년 말의 누적 생존가입자 수(실제: 109,097명 > 보정: 97,787명)는 실제 관측치를 사용한 결과가 더 높지만, 2047년 말의 누적 생존가입자 수(실제: 953,284명 < 보정: 987,183명) 및 2047년까지의 최대 신규 가입자 수(실제 2041년: 53,836명 < 보정 2034년: 57,565명)는 보정된 관측치를 사용한 결과가 더 높게 나타났다.

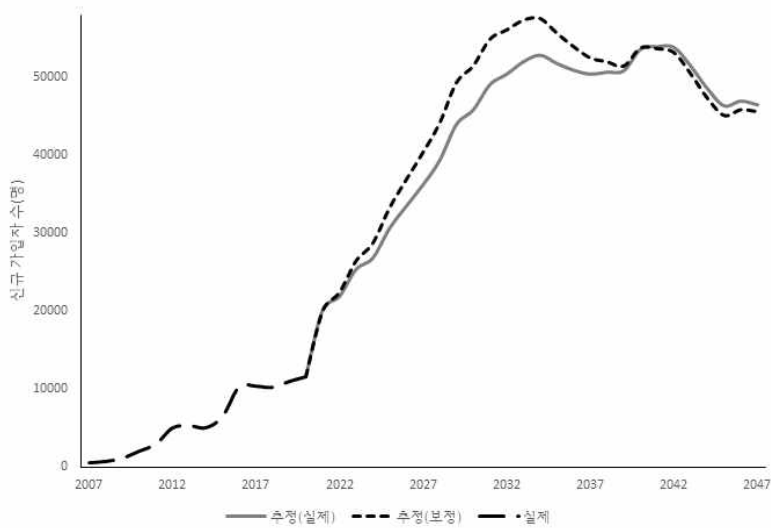
누적 생존가입자 및 신규 가입자 수 장래 추정치의 추세를 그래프로 나타내면 <그림 11, 12>와 같다.

V. 결론

주택연금은 공급자인 공사가 장기간 주택가격리스크, 금리리스크, 장수리스크를 부담하는 구조로서 제도의 건전한 운영을 위해서는 실제 현실을 잘 반영한 수요예측이 무엇보다 중요하다. 또한, 주택연금 가입자 수에 대한 전망은 공사의 목표설정 및 사업계획 수립에도 큰 영향을 주며 나아가 정부의 노인복지정책과도 관련된다. 주택연금 수요예측의 중요성을 고려하여 본 연구에서는 주택연금의 주요한 특징인 인구구조의 변화와 주택가격전망이 주택연금 가입자 수에 미치는 영향을 반영하여 Bass 확산모형을 통해 주택연금의 중장기 수요를 추정하였다. 연구의



<그림 11> 누적 생존가입자 수 장래 추정치 추세



〈그림 12〉 신규 가입자 수 장래 추정치 추세

결과를 정리하면 다음과 같다.

첫째, 사망자가 잠재시장 및 누적 가입자 그룹에서 제외되는 것을 고려할 경우, 주택연금의 누적 가입자 수는 장기적으로 안정적인 수준에서 유지될 것이며, 이는 신규 가입과 사망해지가 안정적인 수준에서 장기균형을 이루며 제도가 꾸준히 지속될 수 있음을 보여준다.

둘째, 주택가격에 대한 전망은 주택연금 가입에 결정적인 역할을 하며, 이는 일시적인 현상에 그치지 않고 장기적인 수요추정에도 영향을 줄 수 있음을 확인했다.

셋째, 주택연금의 Bass 확산모형은 높은 모방계수를 보인다. 고령층의 자산구조에서 주택이 차지하는 비중이 매우 높는데 주택연금 가입은 이런 주택의 처분으로도 이어질 수 있는 매우 신중한 행동인 점에서 구전효과는 중요하다. 다만, 최근 혁신계수의 비중이 더욱 커진 것을 통해 공사의 제도개선 노력 또한 가입에 영향을 주고 있음을 말해준다.

한편, 본 연구를 통해 얻을 수 있는 시사점은 다음과 같다.

첫째, 주택연금 또한 다른 일반적인 신기술·신제품과 마찬가지로 그 수요는 특정 시점에 정점을 기록하고, 이후로는 안정세에 접어든다. 본 연구에서는 2030년대 중반을 전후하여 정점에 이르는 것으로 나타났다. 다만, 여기서 말하는 안정세란 정적 과정이 아니고 신규 가입과 사망해지가 균형을 이루며, 계속 발생하는 동적 과정이다. 따라서 신규 가입 못지않게 이에 대한 사후관리가 중요하며, 보다 효율적이며 전문적인 사후관리 체계 구축방안을 고민할 필요가 있다.

둘째, 익히 알려진 바와 같이 주택연금 가입의사 결정은 주택가격에 대한 전망에 큰 영향을 받는다.

따라서 시시각각으로 변하는 주택시장 동향에 공사가 보다 적극적으로 대응할 수 있도록 주택가격 상승전망 시기와 하향전망 시기의 제도운영 계획을 수립·적용할 필요가 있다.

셋째, 주택연금 시행 초기에는 구전효과가 매우 중요했지만 초기 단계에서 벗어나면서 구전효과 대비 혁신효과의 비중이 증가하고 있다. 따라서 구전효과뿐만 아니라, 고령자에게 실질적인 도움이 될 수 있는 제도개선을 지속하여야 한다. 그리고 이러한 제도개선의 결과를 잘 홍보할 수 있는 효과적인 마케팅을 통해 지속적인 가입확대를 일구어낼 수 있을 것이다.

마지막으로 본 연구의 한계점 및 발전 방향은 다음과 같다.

첫째, 주택연금은 제도 도입 후 경과 기간이 짧아 현재 사용 중인 국민생명표에 대해 충분한 검증을 수행할 수 없으며, 자체 경험생명표만으로는 신뢰성 있는 사망확률을 산출하기 어렵다. 본 연구에서도 장래 사망률의 적용에 있어 통계청 작성 장래생명표를 이용하였으나, 더 많은 시계열 자료가 축적되어 경험생명표의 신뢰성이 높아진 후에는 주택연금 경험생명표를 통해 더욱 정확한 수요를 예측할 필요가 있다.

둘째, 2020년 4월 가입연령 기준을 60세에서 55세로 완화하였고, 공시가격 9억 원 이하 주택의 가입허용, 주거목적 오피스텔의 가입허용, 신탁형 주택연금 도입 등의 내용이 포함된 공사법 개정법률이 2020년 12월부터 시행되는 등 가입대상 확대를 위한 제도개선이 활발히 이루어지고 있지만, 본 연구에서는 2019년 12월의 기준에 따라 분석을 수행하였다. 추후 제도개선에 따른 효과를 반영한 분석 수행이 가능할 것이다. 다만, 제도개선으로 인해 새로 편입되는 가입 대상자들에 대한 수요예측 시에는 이들을 기존 가입 대상자들과 동질한 집단으로 볼지에 대해 추가적인 논의가 필요하다.

셋째, 본 연구에서는 주택가격전망지수를 분석에 활용하였지만, 이를 확산모형에 직접 반영하지는 못하였다. 주택가격전망 외에도 선행연구에서 제시한 주택연금 가입에 영향을 주는 여러 요인을 확산모형에 반영하는 방법을 검토할 수 있을 것이다. 다만, 해당 요인들이 확산추세에 따라 수용자가 증가하면서 자연스럽게 늘어나는 요인인지, 아니면 이와 별개로 독립적으로 수요에 작용하는 요인인지에 대해서 먼저 확인해야 한다.

넷째, 본 연구에서 추정한 누적 생존가입자 수는 중도해지를 고려하지 않은 수치이므로 실제로는 중도해지가 발생하는 정도에 따라 주택연금의 유효 이용자 수는 달라질 수 있음을 주의해야 한다.

다섯째, 본 연구는 Bass 확산모형을 통한 중장기 수요추정이 주요 과제였지만, Bass 확산모형 적용과는 별개로 주택가격전망 및 근래에 누적된 주택가격 상승 등 주택가격과 관련된 다양한 요인들이 주택연금의 가입 및 중도해지에 주는 영향을 분석하면 주택연금의 수요예측에 대한 보다 풍부한 논의가 이루어질 수 있을 것으로 기대한다.

참고문헌

- 김정주, 마승렬, 2011, “역주택저당대출 수요의 결정요인에 관한 연구”, 부동산학보 46권, 207-225.
- 김정주, 2013, “역모기지 수요 변화의 결정요인 분석과 정책적 시사점”, 서울도시연구 14권 2호, 13-33.
- 민인식, 조만, 2009, “역모기지의 Cross-over Risk와 잠재수요에 관한 연구”, 주택연구 17권 3호, 161-187.
- 박근수, 김영훈, 2010, “고령자의 주택연금 제도 이용의 필요성에 영향을 미치는 요인: 55세 이상 고령자를 중심으로”, 노인복지연구 50권, 291-312.
- 백인걸, 최경진, 2020, “주택연금의 소득보장 및 소비진작효과 분석”, 한국자료분석학회지 22권 1호, 227-241.
- 서윤규, 2016, “주택연금(Reverse Mortgage) 제도 이용의 결정요인 분석”, 부동산학보 66권, 231-245.
- 신현재, 2019, “주택가격 변동률이 주택연금 가입에 미치는 영향 분석”, 주택도시금융연구 4권 1호, 25-50.
- 안세룡, 2018, “주택연금 제도의 재무적 가치”, 한국증권학회지, 47권 5호, 823-847.
- 양진아, 민대기, 최형석, 2017, “Bass 확산모형을 활용한 국내 주택연금의 중·장기 수요 예측”, 한국경영과학회지 42권 1호, 29-41.
- 양진아, 민대기, 김은갑, 2019, “순위반응모형을 이용한 주택연금의 잠재시장 규모 추정”, 유라시아연구 16권 3호, 61-80.
- 여운경, 양재환, 2018, “상속동기를 반영한 주택연금의 평가와 가입전략 - 주택연금의 역할 관점에서”, 리스크관리연구 29권 1호, 33-87.
- 유선종, 구본영, 2005, “역모기지 제도 도입을 위한 고령자 의식에 관한 연구”, 국토연구 45권, 119-143.
- 이달남, 김수민, 신승우, 2015, “고령가구의 주택연금 가입 결정요인에 관한 연구”, 도시행정학보 28권 2호, 309-323.
- 이선형, 김영훈, 2009, “수도권 노인의 주택연금제도 이용의향에 관한 연구: 예비노인과의 비교를 중심으로”, 한국인구학 32권 3호, 73-101.
- 이종의, 2010, “특수지역 종사자들의 역모기지 이용의사에 관한 연구”, 주택연구 18권 1호,

127-157.

임하나, 신승우, 노승한, 2016, “지역별 주택연금 가입비율에 영향을 미치는 요인 연구 - 시·군·구를 중심으로”, 부동산연구 26권 3호, 39-47.

전희주, 2019, “주택연금 가입에 대한 가입 의향 요인”, 한국자료분석학회지 21권 3호, 1203-1216.

채현석, 정재우, 김종달, 2016, “Bass 확산모형을 이용한 전기자동차 수요예측”, 환경정책 24권 1호, 109-132.

최경진, 김대환, 2020, “주택 가격 예측에 따른 주택연금 가입 의향 분석”, 소비자문제연구, 51권 1호, 35-50.

홍정식, 김태구, 구훈영, 2011, “NLS와 OLS의 하이브리드 방법에 의한 Bass 확산모형의 모수추정”, 대한산업공학회지 37권 1호, 74-82.

Bass, F. M. 1969., “A new product growth model for consumer durables”, Management Science, 15, 215-227.

Bass, F. M., Krishnan, T. V. and Jain, D. C., 1994, “Why the bass model fits without decision variables”, Marketing Science, 13(3), 203-223.

Chandrasekaran, D. and Tellis, G. J., 2004, “A critical review of marketing research on diffusion of new products”, Review of Marketing Research, 3, 39-80.

Mahajan, V., Muller, E. and Bass, F. M., 1990, “New product diffusion models in marketing: A review and directions for research”, Journal of Marketing, 54, 1-26.

Satoh, D., 2001, “A discrete bass model and its parameter estimations”, Journal of the Operations Research Society of Japan, 44(1).

(논문 접수일: 2020.10.31. 수정논문 접수일: 2020.12.03. 논문 채택일: 2020.12.11.)

Long-Term Demand Estimation for Reverse Mortgage Considering Changes in Population Structure and House Price Prospect

Jonghoon Lee*

Abstract

For the stable operation of the Home Pension(Korean public reverse mortgage) system, policy establishment and risk management should be made based on demand estimates that reflects the main characteristics of reverse mortgage. In this study, changes in population structure and house price prospect were selected as the main factors that affect Home Pension subscriptions to estimate demand through the Bass diffusion model. The trend of cumulative number of surviving subscribers in Home Pension was investigated to reflect changes in population structure in the diffusion model. In addition, a comparative analysis was conducted to examine the impact of house price forecasts on new Home Pension subscriptions. The analysis shows that the number of new Home Pension subscribers will peak from the mid-2030s to 2040, after which the total number of new subscribers will remain stable with a long-term balance with the death toll.

Keywords : Home Pension, Reverse Mortgage, Bass Diffusion Model, Population Structure, House Price Prospect

* Jonghoon Lee, Corresponding author, Korea Housing Finance Corporation(HF) Reverse Mortgage Department, Assistant Manager, 1864@hf.go.kr

© Copyright 2020 Housing Finance Research Institute. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

부록

1. 논문작성지침
2. 외부논고 투고자 제출서류
3. 논고 심사체계 안내

논문작성지침

1. 논문작성양식

- ① 원고는 한글(HWP) 워드프로세서로 작성하며, 원고의 본문 분량은 A4용지 20페이지 내외로 한다.
- ② 원고는 표지, 본문, 참고문헌, 부록, 영문초록, 영문 핵심주제어 순으로 배열한다.
- ③ 표지에는 논문제목, 저자명(페이지 하단에 소속, 직위, 이메일주소를 기입하며, 공저일 경우 제1저자, 공동저자, 교신저자 등의 순서대로 이를 명시한다.), 국문초록(500자 이내), 국문 핵심주제어(5개 내외)를 표기한다.
- ④ 원고용지 편집은 다음에 의한다.

용지여백 (mm)		문단모양		글자모양 및 번호체계	
위쪽	10	좌우여백	0	제목	20pt
아래쪽	10	간격	줄간격	본문	12pt
왼쪽	20		문단위	글꼴	휴먼명조
오른쪽	20		문단아래	도표/그림 제목	12pt
머리말	15	들여쓰기		각주	10pt
꼬리말	15	정렬방식		양쪽혼합	
제본	0				

- ⑤ 본문에는 논문 저자명 등 투고자의 신분이 명시적으로 노출되지 않도록 주의하여야 한다.
- ⑥ 본문의 장절구분은 I, 1, 1), (1), ①, 가의 순서를 따르며, 본문 및 각주는 완전한 문장으로 작성한다.
- ⑦ 표와 그림의 제목에는 각각 일련번호 (예 : <표 1>, <그림 1>)를 매겨 각각 표와 그림의 상단 중앙에 배치하고, 자료 출처는 아랫부분에 배치한다.
- ⑧ 수식의 번호는 우측 정렬하여 일련번호 (예 : <식 1>)로 표기한다.

2. 본문 주와 참고문헌

- ① 인용 · 참고한 자료의 출처를 밝히는 참고주(Reference notes)는 본문 중에 괄호 속에 표기하고 (예 : 홍길동 외, 2010), 상세사항을 정리한 참고문헌(Reference List)은 논문의

끝부분에 첨부한다.

② 본문의 내용에 설명을 부연하기 위한 내용주(Content notes)는 해당 부분의 오른쪽 위에 일련번호를 매기고 (예 : 부동산업¹⁾은), 페이지 하단에 각주로 처리한다.

③ 참고문헌 작성은 다음에 의한다.

구분	세부내용
작성 방법	- 참고문헌은 본문의 마지막에 장을 달리하여 기록하며, 제1저자의 성을 기준으로 국내문헌은 가나다순으로, 외국문헌은 알파벳순으로 배열한다. - 같은 저자의 여러 문헌은 연도순으로 배치하되, 동일연도에 발간된 문헌이 둘 이상일 경우에는 문헌 제목의 가나다 혹은 알파벳 순서에 따라 발간연도에 a,b,c를 붙여 구분한다. - 기록순서는 정기간행물은 저자명, 출판년도, 논문제목, 간행물명, 권(Vol), 호(No.), 면(page)으로 기록하며, 정기간행물이 아닌 경우에는 저자명, 출판년도, 도서명, 권, 호, 출판사, 인용한 페이지 등의 순서로 기록한다. - 신문기사는 신문사명, 발간년, 월, 일을 표기하고, 인터넷 등 온라인 자료는 작성자, 자료생성일, 자료제목, 자료주소(URL)의 순서로 기재한다. - 참고문헌은 국문문헌, 영문문헌, 신문기사, 온라인 자료의 순서로 기재한다.
예시	- 강철수, 2013, 계량경제학, 도서출판 책, 22-43 - 고제현, 2017, “주택연금과 하우스세어링 연계방안”, 주택금융연구 1권 1호, 63-64 - Laurenson. J, 2016, “Generations together”, Stanford Social Innovation Review, 17-18.

3. 제출방법 등

① 홈페이지에 공지된 학술지 담당자 이메일로 제출한다.

② 투고논문은 다른 간행물에 발표되지 아니한 것이어야 하며, 또한 다른 간행물에 중복 심사 의뢰하여서는 아니된다. 단, 다른 간행물에 심사의뢰하였으나, 탈락이 확정된 것은 예외로 한다.

③ 논문 작성에 사용되는 언어는 한글을 원칙으로 한다.

④ 논문 투고시 심사비 및 게재료는 발생하지 않으며, 게재 확정시 소정의 원고료를 공사가 투고자에게 지급한다.

⑤ 투고된 원고는 반환하지 않으며, 투고자는 게재된 논문에 대한 판권을 포함한 저작권 일체를 한국주택금융공사에 위임한다.

⑥ 투고자는 한국주택금융공사의 내규인 「주택금융연구원 운영규정」 제6장(연구윤리)(별첨)을 준수한다.

외부논고 투고자 제출서류

□ 논문 투고 신청서

논문제목				
제 1 저 자	이름		교신저자 여부	(여, 부)
	소속 및 직위			
	전화			
	이메일			
공 동 저 자	이름		교신저자 여부	(여, 부)
	소속 및 직위			
	전화			
	이메일			

* 공동저자가 2인 이상인 경우 모든 저자의 이름, 소속 및 직위, 전화, 이메일을 기재하여 주십시오

저자들은 본 논문이 한국주택금융공사 학술지 「주택금융연구」에 게재되기를 희망합니다. 또한 이 논문은 한국주택금융공사 “주택금융연구원 운영규정 제6장(연구윤리)”를 준수하여 작성하였음을 확인합니다.

제1저자 서명 (인)

공동저자 서명 (인)

년 월 일

한국주택금융공사 귀중

□ 저작권 위임 동의서

논문제목 :

국문 :

영문 :

본 논문의 저자(들)는 한국주택금융공사 학술지「주택금융연구」에 논문 게재를 신청하면서 다음 사항에 대하여 확인 및 동의합니다.

1. 저자(들)는 본 논문이 학문적으로 독창성을 가지며, 다른 논문의 저작권을 침해하지 않았음을 확인합니다.
2. 저자(들)는 본 논문에 실질적인 공헌을 하였으며, 논문의 내용에 대하여 책임을 함께 합니다.
3. 저자(들)는 본 논문이「주택금융연구」에 게재된 후 표절로 판명되거나, 다른 학술지에 게재된 논문으로 확인된 경우, 귀 공사의 조치(원고료 반납 등)에 따르겠습니다.
4. 저자(들)는 본 논문이 투고일 현재 다른 간행물에 발표되지 않은 논문이며, 다른 간행물에 투고 중인 논문이 아님을 확인합니다.
5. 저자(들)는「주택금융연구」의 발행인이 저자(들)나「주택금융연구」발행인의 허락 없이 타인에 의해 이루어지는 저작권 침해에 대해서 이의를 제기할 권리가 있음을 동의합니다.
6. 저자(들)는 본 논문이「주택금융연구」에 게재될 경우, 본 논문에 따른 권리, 이익, 저작권 및 디지털 저작권에 대한 모든 권한행사(복사·전송권 포함) 등을 한국주택금융공사에 위임함을 확인합니다.

저자	성명	소속	연락처	자필서명
제1저자				
공동저자				

년 월 일

한국주택금융공사 귀중

논고 심사체계 안내

- ① 편집위원회를 구성하여 「주택금융연구」에 논고 수록을 요청한 논고의 게재 여부를 심사한다.
- ② 편집위원회는 편집위원장 1인과 편집위원장이 지명한 편집위원 3인으로 구성한다.
- ③ 편집위원 3인은 논문의 유용성, 내용의 창의성, 전개의 논리성, 자료의 신뢰성, 연구 윤리 등을 심사한다.
- ④ 심사결과에 따른 게재 여부는 다음 판정기준에 따른다.

※ 논문심사 판정기준 (○: 게재가능/수정후게재, △: 수정후재심, X: 게재불가)

심 사 결 과			판 정
○	○	○	게재가능/수정후게재 ^{주1)}
○	○	△	게재가능/수정후게재
○	○	X	게재가능/수정후게재
○	△	X	수정후재심 ^{주2)}
○	△	△	수정후재심
○	X	X	게재불가
△	△	△	수정후재심
△	△	X	수정후재심
△	X	X	게재불가
X	X	X	게재불가

주1. 수정후게재란 일부 자구의 수정이 필요하거나 내용상 수정사항이 경미하여 일부 수정 후 게재가 가능하다고 판단되는 것을 의미하며, 편집위원회는 심사자가 제시한 수정의견을 투고자에게 통보하고, 투고자는 이를 적극 반영하여 논문을 수정하여야 한다.

주2. 수정후재심이란 수정사항이 중요하거나 논문의 내용이 대폭 보완되어야 한다고 판단되는 경우로서, 투고자는 심사자의 의견을 반영하여 수정한 후 재심을 받아야 한다.

- ⑤논문수정은 10일 이내를 원칙으로 하되, 편집위원회는 수정의 난이도를 고려해 기간을 가감할 수 있다.
- ⑥수정 후 재심 판정에 따라 논문을 수정한 후 재심사한 결과, 수정후재심 의견이 다시 나온다면 게재불가 판정한다.

주택금융연구

통권 제 4호

등록일자 | 2017년 5월 25일(부산남. 사00005)

발행일 | 2020년 12월 30일

발행인 | 이정환

편집인 | 권오훈

발행처 | 한국주택금융공사 주택금융연구원(48400)

부산시 남구 문현금융로40(문현동)

부산국제금융센터

TEL : 051) 663-8155

FAX : 051) 632-9570

편집·인쇄 | 거목문화사 02-2277-3324

 **한국주택금융공사**

www.hf.go.kr