

트리 앙상블 기반 자동산정모형을 이용한 재고 주택가격 분포 관측: 전국 아파트 단지 전수 자료의 시계열 분석

홍정의*

요 약

본 연구는 부스팅 트리 알고리즘(XGBoost)에 기반한 자동산정모형을 활용하여 추정 기반 재고 주택가격 관측을 시도한다. 본 연구는 2012년 3월 1일부터 2024년 4월 1일까지의 아파트 실거래 자료 6,118,730건과 전국 12,517,131호(주택 총조사 기준 전체 아파트 단지의 99.1%)의 개별 단지(평형별) 정보를 활용하여, 지역별 평균 가격 수준, 가격 분포, 소득 대비 주택가격 비율(price to income ratio, PIR) 등을 도출하였다. 결과는 평균 아파트 가격이 2017~2022년 가격상승기에 기존 공표가격지수에 비해 더 빨리 상승했다는 것과, 분석 기간 동안 지역 간 가격 격차와 지역 내 가격 분위수 간 격차가 동시에 확대되었음을 보여주었다.

핵심어 : XGBoost, 기계학습, 자동산정모형(Automated Valuation Model, AVM), 주택가격분포, PIR(Price to Income Ratio)

I. 서론

주택 시장에 대한 분석에서 가장 핵심이 되는 정보는 아마도 주택의 가격일 것이다. 하지만 주택 가격은 주택이라는 자산의 특성상 상당히 제한적인 형태로만 관측이 가능했다. 주식이나 채권과 같은 금융자산은 기본적으로 동질적이며 따라서 그에 대한 거래가 각 시점마다 빈번하게 발생하므로,

* 홍정의, 주저자, 한동대학교 경영경제학부 부교수, hwgh024@handong.edu

© Copyright 2026 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.

개별 자산 가격의 변동성이 시계열 상 명확히 관측된다. 한편 모든 주택은 본질적으로 비동질적이며 개별 자산의 거래 빈도는 매우 낮다. 즉, 주택은 위치, 구조, 주변 환경과 같은 많은 특성들이 결합하여 개별 주택의 가격을 형성하며, 특정 시점에 거래가 이루어지는 주택은 전체 재고 주택의 극히 일부에 불과하다.

주택 시장의 이러한 특징이 "올해 서울의 평균 아파트 가격은 얼마인가?"와 같은 일견 평범해 보이는 질문에 대해서도 그 답변을 어렵게 만드는 이유가 된다. 만약 단순히 올해 거래된 아파트 가격의 평균을 답했다면, 그것은 올해 많이 거래된 아파트의 평균적 가격일 뿐 존재하는 전체 아파트의 실제 평균 가격이 아닐 것이다. 경제적 환경에 따라 많이 거래되는 아파트의 종류가 상당히 달라질 수 있음을 감안하면, 그러한 대답이 일반적으로 믿을만한 답변이 아닌 것은 이미 잘 알려져 있다. 이처럼, 주택 시장에서는 전체 자산 중 일부의 가격만이 관측되는 한계가 존재하며, 이를 극복하고 시장 전체의 가격 수준과 추이를 파악하기 위해서는 전통적으로 주택가격지수를 이용해왔다.

그러나 기존의 주택가격지수 역시 그 구성 방식에 내재하는 근본적 한계를 안고 있다. 가장 널리 사용되는 두 가지 접근인 반복매매지수(repeat-sales index)와 헤도닉 가격지수(hedonic price index)는 각각 다른 방식으로 주택의 질적 차이를 통제하고자 하지만, 양자 모두 해당 시점 거래된 주택만을 대상으로 한다는 공통적 한계를 갖는다. 반복매매지수는 동일 주택이 두 번 이상 거래된 경우의 가격 차이만을 이용하여 지수를 산출하므로 표본선택편의(sample selection bias)에 노출되며(Gatzlaff & Haurin, 1997), 헤도닉 가격지수는 함수 형태의 설정오류(functional form misspecification) 문제가 지적되어 왔다(Hill, 2013).

이보다 본질적으로, 주택가격지수는 모두 본질적으로 가격의 상대적 변화만을 측정하며, 가격의 절대적 수준에 대한 정보를 직접 제공하지 못한다. 예를 들어, 특정 연도 서울 지역의 주택가격지수가 110이라는 것은 기준 시점 대비 10% 상승했음을 알려줄 뿐, 실제 평균 가격의 수준 자체가 얼마인지에 대해서는 아무런 정보도 담고 있지 않다. 또한, 지수의 산출 과정에서 사용되는 표본이 고정되어 있으므로, 연구자가 원하는 기준에 따라 표본을 자유롭게 분할하여 하위 집단별 가격 추이나 가격 분포를 관측하는 것이 구조적으로 어렵다. 결과적으로, 기존의 주택가격지수는 시장의 가격 변화 방향과 크기에 대한 중요한 정보를 제공하지만, 시장 전체의 가격 수준, 분포 구조, 가격 불균등성, 소득 대비 가격 부담 등을 입체적으로 파악하는 데에는 근본적인 한계가 있었다.

만약 우리가 개별 주택들의 가격을 일정 수준 이상 정확하게 추정할 수 있다면, 이러한 한계를 상당 부분 극복할 수 있을 것이다. 특히, 거래된 주택뿐 아니라 거래되지 않은 주택을 포함한 전체 재고 주택의 가격을 추정할 수 있다면, 이는 거래 표본에 의존하는 전통적 가격 측정 방식을 보다 확장하는 관측 정보를 제공할 수 있다. 본 연구는 이러한 접근을 추정 기반 재고 주택 시장 관측으로 개념화한다.

이러한 전수적 가격 추정의 가능성은 대규모 주택가격 추정을 위한 자동산정모형(automated valuation model, AVM)의 효율성 확장과 밀접한 관계가 있다. 주택가격 자동산정모형에 대한 논의는 꽤 오랜 역사를 가지고 있다(IAAO, 2018; Tekouabou et al., 2024). 그럼에도 불구하고 시장 상황을 파악하는 데에 있어서 자동산정모형은 지수산출과 같은 영역에서의 활용에는 비교적 관심을 받지 못해 왔다. 그 주요한 이유는 회귀 모형에 기반한 전통적 자동산정모형은 개별 자산 단위에서 그 예측의 정확성이 충분히 높지 않다는 근본적인 한계를 지니고 있었기 때문이다(Hill, 2013; Malpezzi, 2003). 이는 비록 전통적 모형이 이론적 기반과 결과의 직관적 해석 가능성을 제공함에도 불구하고, 직관적 함수 형태를 가정하는 모형의 특성 상, 현실의 복잡성을 충분히 포착하기 어려울 뿐 아니라, 모형의 설정 오류(misspecification)나 변수 간 내생성 문제에서도 자유롭지 못하다는 것과 관련되어 있다.

그러나 최근 들어, 기계학습 기반의 자동산정모형이 주택 가격 평가의 정확성과 안정성을 획기적으로 향상시킬 수 있다는 사실이 보고되고 있다(이인호 · 이경섭, 2020; 홍정의, 2020; Antipov & Pokryshevskaya, 2012; Hong & Kim, 2022; Hong et al., 2020; Pace & Hayunga, 2020). 특히, 개별 주택의 가격 예측에서 두드러진 성과를 보이고 있는 접근법은 랜덤포레스트나 XGBoost와 같이 의사결정나무(decision tree)의 앙상블을 활용한 모형이다. 전통적인 모형과 달리, 트리 앙상블에 기반한 모형은 함수 형태에 대한 사전 가정 없이 주어진 데이터를 통해 비선형적이고 복잡한 특성 간의 관계를 자동으로 학습할 수 있으며, 이는 개별 주택 가격의 추정 정확도를 크게 높이는 것으로 나타나고 있다. 이러한 기계학습 모형은 기존의 선형회귀 기반 모형에 비해 상대적으로 정확성이 높을 뿐만 아니라, 절대적인 예측 성능 면에 있어서도 상당한 수준의 정확성을 확보할 수 있는 것으로 드러나고 있다(Hong et al., 2020). 그 중에서도 특히 기존 모형의 약점을 반복적으로 보강하는 부스팅 트리 알고리즘(XGBoost)이 특히 높은 정확성과 안정성을 보이는 것으로 알려져 있다(Hong & Kim, 2022).

본 연구의 목적은 XGBoost 알고리즘에 기반한 자동산정모형(AVM)을 이용하여 전국 단위의 개별 아파트 단지(단지 내 각 평형별) 가격을 각 시점마다(분기별) 전수적으로 추정하고, 이를 통해 지역별 주택시장의 가격 수준, 분포, 그리고 그 시계열적 변화를 직접 관측할 수 있는 방법을 제시하는 것이다. 본 연구가 제시하는 것은 가격지수가 아니라, ¹⁾ AV M의 정확성을 활용하여 시장 전체의 가격 분포를 시점별로 추정·관측하는 방법이며, 이를 통해 가격지수가 직접 제공하지 못하는 정보 차원에 접근한다.

이러한 작업을 수행하기 위해서는 실거래 자료뿐 아니라, 전국 단위의 개별 아파트 단지에 신뢰할 만한 정보가 필요하다. 본 연구에서는 한국부동산원으로부터 제공된 전국 개별 아파트 단지의 평형별

1) 전통적으로 가격지수는 시기별 거래 표본의 질적 차이뿐 아니라 재고 주택 자체의 시기별 변화(멸실 및 신규 입주)에 대한 보정과 가중 방식의 선택 등 별도의 방법론적 논의를 요구하나, 이는 본 연구의 초점은 아니다.

정보와 아파트 실거래 정보를 사용하였는데, 이는 현재 구축되어 있는 대규모 공동주택 정보 중 가장 신뢰도가 높고 포함 범위가 넓은 자료라고 할 수 있다. 본 연구의 자료는 2012년 3월 1일부터 2024년 4월 1일 사이의 아파트 실거래 정보 6,118,730건과 총 12,517,131호에 해당하는 개별 아파트 단지의 평형별 정보를 포함하고 있는데, 이는 우리나라에 존재하는 전체 아파트 세대의 약 99.1%(2023년 총주택 총조사 기준 우리나라 전체 아파트 12,631,608 호 중)에 해당하는 것이다. 따라서, 본 연구의 추정치는 우리나라의 아파트 시장 전체의 상태에 대한 실질적인 전수 가격 추정에 가까운 결과가 될 수 있다.²⁾

본 연구에서 실거래가 자료는 개별 아파트 가격의 시점별 추정을 위한 학습에 사용되었으며, 개별 단지의 평형별 정보는 추정의 대상으로서, 동일한 표본을 전 기간에 걸쳐 고정적으로 유지함으로써 표본 구성 변화의 영향을 배제하였다. 이렇게 추정된 아파트 단지별 가격을 바탕으로 본 연구는 주택 가격의 수준과 분포 및 그 추이를 탐색하였다. 첫째, 본 연구에서는 전국 시·도 등 지역별로 단위 가격(만 원/m²)의 수준을 찾고, 지역 간 및 지역 내 가격 분포의 추이를 추적하였다. 둘째, 지역별로 PIR(price to income ratio)을 산정하여 지역별 아파트 가격이 가구의 실제 구매력 대비 어느 정도의 수준이고 어떤 차이가 있으며 어떻게 변화하였는지를 구체적으로 제시한다.

본 연구가 제시하는 전수 추정 기반 시장 관측은 다음과 같은 이점을 갖는다. 첫째, 본 연구의 결과는 전통적인 가격지수와 달리 절대적 가격 수치에 대한 정보를 제공한다. 이는 서로 다른 지역의 가격 수준을 직접 비교할 수 있게 하며, 가격의 변화뿐 아니라 수준 자체에 대한 해석을 가능하게 한다. 둘째, 본 연구의 추정 결과는 실질적인 전수에 해당하는 표본을 통해 가격 추이를 관측하고 있으므로 반복매매지수의 표본선택편의 가능성(Gatzlaff & Haurin, 1997; Melser, 2023)에서 자유롭다. 셋째, 전수적 가격 추정 정보를 기반으로 연구자가 원하는 방식대로 표본을 분할하여 하위 집단별 가격 추이, 분위수, PIR, 지니계수 등 다양한 분석적 지표로 쉽게 확장될 수 있다. 이는 기존의 고정된 표본 바스켓에 의존하는 지수 체계에서는 구조적으로 불가능했던 것으로, 시장에 대한 정보 접근성을 질적으로 확장한다.

본 연구의 구성은 다음과 같다. II장은 본 연구에서 활용된 기계학습 기법인 의사결정나무의 앙상블 알고리즘에 대한 소개 및 관련 선행연구들을 검토하였다. III장에서는 본 연구에서 사용된 자료에 대한 개략적 정보를 소개한다. IV장에서는 개별 주택 가격에 대한 추정을 수행하기에 앞서, 먼저 본 연구에서 사용한 기계학습 기반 자동산정모형이 개별 아파트 가격을 얼마나 정확하고 안정적으로 추정하는지를 실거래가 자료를 통해 테스트하였다. V장에서는 지역별 아파트 평균 가격 추이, 가격 분포의 추이, 가격 불균등성, PIR의 추이를 제시한다. VI장에서는 결론 및 향후 연구 과제를 논의한다.

2) 특히, 거의 전수에 가까운 표본을 가지고 있다는 것은 추정의 일반적 편향성이 발생하지 않는 한 분석 대상에 대한 거의 실제에 가까운 통계량들을 얻을 수 있다는 것을 의미한다.

II. 모형 개요 및 선행연구 검토

전통적으로 주택가격의 자동산정모형으로 가장 널리 알려진 접근법은 헤도닉 가격 모형일 것이다. 헤도닉 모형은 Lancaster(1966)의 소비자 이론과 Rosen(1974)의 연구를 이론적 토대로 하여, 주택의 시장 가치를 크기, 구조, 주변 환경과 같이 그 효용을 구성하는 요소들에 대한 함수로 표현하는 방식이다. 이러한 접근 방식은 가치에 영향을 미치는 개별 특성과 주택 가격과의 명시적인 함수 관계를 사전에 설정하고(Sheppard, 1999), 주어진 자료를 통해 가정된 관계의 계수를 회귀분석을 통해 추정하는 방법으로 작동한다.

헤도닉 모형은 직관적이고, 계수에 대한 가설검정 등을 통해 쉽게 실증 분석에 적용될 수 있다는 특징으로 인해 학술 연구에서나 실용적으로나 널리 사용되어 왔다. 그러나 헤도닉 모형이 직관적이고 단순화된 함수 관계를 가정한다는 사실로 인해 개별 주택에 대한 정밀한 가격 추정을 수행하기에는 한계가 있다는 점도 알려져 왔다. 현실에서는 주택 가격이 개별 특성들 사이에 복잡한 상호작용과 비선형적 관계가 존재할 가능성이 크지만, 그 구체적 형태를 모두 연구자가 사전에 알거나 정형화된 식으로 반영할 수는 없기 때문이다(Malpezzi, 2003). 결과적으로 선형함수 가정에 기반한 헤도닉 모형은 모형식별문제에 노출된다(Ekeland et al., 2004). 이는 헤도닉 모형이 높은 해석 가능성을 제공하지만, 개별 주택 가격에 대한 예측의 정밀성에서는 근본적인 제약을 가지고 있음을 의미한다. 결과적으로 전통적인 자동산정모형의 활용은 일반적으로 제한적인 범위에서만 수행되어왔다.

이러한 맥락에서, 최근 데이터 분석 기술의 폭발적 발전과 함께, 기계학습에 기반한 자동산정모형의 활용 가능성이 주목되고 있다(Tekouabou et al., 2024). 그 중에서도 특히 의사결정나무(decision tree)의 앙상블 기법은 연구자가 주택 가격과 특성 간의 명시적인 함수 형태를 미리 가정할 필요가 없고, 주어진 데이터를 통해 비선형적이고 복잡한 특성 구조를 효과적으로 포착하여 주택 가격 예측의 정확성을 획기적으로 높일 수 있는 것으로 알려져 있다(홍정의, 2021; Antipov & Pokryshevskaya, 2012). 이러한 기계학습 모형은 전통적 모형과 달리 가격과 관련 변수 간의 내생적 관계를 직관적으로 해석하기에는 적합성이 떨어질 수 있지만, 그럼에도 자동산정모형의 주요 목표인 개별 주택 가격의 추정에 있어서는 기존 방법론의 정확성을 크게 상회하는 성능을 가질 수 있다. 다수의 선행 연구들은 의사결정나무의 앙상블을 통해 주택가격 예측의 정확성이 기존 전통적 모형에 비해 현저히 개선될 수 있음을 보고하고 있다(Hong et al., 2020; Kim & Hong, 2024; Pace & Hayunga, 2020; Park & Bae, 2015; Rico-Juan & de La Paz, 2021). 예를 들어, Hong et al.(2020)은 대표적인 앙상블 기법인 랜덤 포레스트를 이용하여 대한민국 서울의 강남 지역 아파트 가격을 분석하였는데, 기계학습 기반 모형의 예측 정확성이 평균백분율오차(mean absolute percentage error, MAPE) 기준으로

5% 내외인 것으로 나타나, 이러한 모형이 단순히 상대적으로 전통적 모형에 비해 나을 뿐 아니라, 절대적인 수준에서도 상당히 신뢰할 만한 예측을 수행할 수 있음을 보여준다.

의사결정나무는 특성 간의 명시적 함수 관계에 기초한 전통적 모형과 달리, 주어진 데이터를 특정 조건에 따라 반복적으로 나누어 가격의 차이를 가장 잘 설명할 수 있는 규칙을 자동적으로 탐색하는 알고리즘이다. 이는 마치 스무고개 놀이에서 세부적인 질문을 통해 점점 정답에 가까워지는 과정과 비슷한 형태로 작동하는데, 개별 주택의 특성에 따라 가장 유사한 그룹을 찾아 그 그룹의 가격 평균이나 중앙값을 예측값으로 삼는다. 특히 이 방식은 변수 간의 비선형 관계를 자연스럽게 포착할 수 있으며, 주택 가격에 영향을 미치는 위도와 경도 같은 변수도 별도의 변환 없이 직접 입력하여 유연하게 활용할 수 있다는 장점이 있다.

이러한 특성은 의사결정나무 기반 모형이 학습 자료에 분포의 양 극단에 위치한 비통상적 표본이 포함되어 있더라도 예측에서 상대적으로 강건하다는 점과도 연결된다. 의사결정나무는 분할 기준을 변수의 절대 수준이 아닌 순위(rank)에 따라 결정하므로, 일부 변수에 통상 범위를 크게 벗어난 값이 소수 존재하더라도 분할 지점의 선택이 그 값의 절대적 크기가 아닌 다른 표본들과의 상대적 순서에 의해 좌우되며, 모형 전체의 학습 결과가 소수의 극단값에 의해 크게 왜곡되지 않는다. 또한 의사결정나무의 예측값은 학습 자료에서 동일한 잎(leaf)에 도달한 표본들의 가격을 통계량(평균 또는 중앙값)으로 집계한 결과이므로, 예측 대상이 통상 범위를 벗어나는 입력값을 갖더라도 그 예측은 학습 표본 내 인접 그룹의 가격 범위 내로 자연스럽게 제약된다. 이러한 구조적 특성은 본 연구와 같이 전국 사실상 전수의 단지를 추정 대상으로 삼아 평균·분위수·지니계수·PIR 등의 집단 통계량을 산출하는 분석에서 특히 유리하게 작용한다. 즉, 추정 대상 표본의 일부에 비통상적 값이 포함되어 있더라도, 그러한 표본에 대한 추정치가 학습 자료의 인접 그룹 가격 범위 내로 제약되는 한, 전체 표본을 대상으로 집계한 결과 통계량에 미치는 영향은 제한적이다.

의사결정나무 단독으로 사용하면, 특정 훈련 자료에 지나치게 맞춰져 다른 자료에 적용할 때 성능이 떨어지는 과적합 문제가 발생할 수 있다. 이를 해결하기 위해 주택가격 예측 모형에서는 주로 다수의 의사결정나무를 결합하는 앙상블 기법이 주로 활용된다. 앙상블이란 여러 개의 의사결정나무를 동시에 또는 순차적으로 결합하여 보다 정확하고 안정적인 모형을 구성하는 방법을 말하는데, 이는 여러 개의 개별 모형의 예측 결과를 결합함으로써 개별 나무가 가진 한계나 오류를 감소시키고 보다 견고한 결과를 얻는 방식이다. 이러한 앙상블 기법은 크게 배깅(bagging)과 부스팅(boosting)으로 구분할 수 있다. 배깅 모형의 대표적인 예로는 Breiman(2001)에 의해 제시된 랜덤 포레스트가 있다. 이는 데이터를 서로 다른 방식으로 반복적으로 샘플링하여 독립적으로 훈련된 여러 나무의 평균 예측값을 사용하는 방식이다. 기계학습을 활용한 모형의 초기에 랜덤 포레스트를 활용한 연구가 다수 제시되었다. 예를 들어, Antipov & Pokryshevskaya(2012)는 러시아 상트페테르부르크의 아파트 2,848건

에 대한 데이터를 분석한 결과 랜덤 포레스트가 다른 여러 예측 모형에 비해 높은 예측 정확도를 나타낸다는 점을 확인하였다. 또한 Ho et al.(2021)은 홍콩에서 18년간 발생한 4만 건의 주택 거래 자료를 분석하여, 랜덤포레스트와 부스팅 기반의 모델들이 지지벡터기계(support vector machine, SVM)보다 더 예측 정확성이 높았음을 확인하였다. Yilmazer & Kocaman(2020)는 터키 앙카라의 복잡한 도시 구조를 지닌 마마크(Mamak) 지역을 사례로 선정하여 전통적 회귀모형과 랜덤 포레스트를 통한 대량평가의 결과를 비교하였다. 그 결과 랜덤 포레스트는 요인 변수가 복잡하게 형성되는 경우에도 그 구조를 자동으로 최적화하여 회귀모형보다 예측 정확성이 더 뛰어남(조정 R^2 0.734)을 확인하였다.

부스팅은 이전 단계에서 잘못 예측된 데이터 포인트를 다음 단계에서 보다 중점적으로 학습하여 점진적으로 성능을 높이는 방식이다. 배깅이 개별 나무를 독립적으로 생성하는 반면, 부스팅은 직전의 나무가 잘못 예측한 데이터에 초점을 맞추어 다음 나무가 이를 보완하도록 하는 원리이다. 이러한 방식으로 부스팅은 개별 나무의 민감성을 점진적으로 줄이고 전체 모형의 예측 정확성을 지속적으로 높여간다. 이러한 부스팅 방식의 대표적인 예가 최근 주택 가격 예측에서 뛰어난 성능을 보여주고 있는 것으로 알려진 XGBoost(eXtreme Gradient Boosting)이다. XGBoost는 순차적으로 새로운 의사결정나무를 추가하여 이전 단계의 예측 오류를 점진적으로 개선해나가는 방식으로 작동한다. 각 나무는 앞서 구성된 나무가 잘못 예측한 자료를 집중적으로 학습하도록 설계되어 전체적으로 예측력을 높여간다. XGBoost는 복잡한 비선형 관계를 효과적으로 탐색하여 데이터에 내재된 구조를 잘 포착할 뿐 아니라, 나무의 복잡성을 제어하는 정규화를 활용하여 과적합 문제를 보다 적극적으로 관리할 수 있다.

최근의 관련 연구들 역시 개별 주택가격 예측에 있어 XGBoost의 활용 가능성을 주목해왔다(Guliker et al., 2022; Hong & Kim, 2022; Ming et al., 2020). 예를 들어, Ming et al.(2020)은 XGBoost를 비롯한 부스팅 기반 모형들이 부동산 가격 예측에 탁월한 성능을 나타내며, 특히 가격에 영향을 미치는 복잡한 변수 간 관계를 보다 효과적으로 학습하여 전통적 모형과의 성능 차이가 명확하다는 것을 보여주었으며, Hong & Kim(2022)은 서울의 아파트 실거래가 자료를 기반으로 XGBoost를 비롯한 의사결정나무 모형들의 예측력이 평균백분율오차(MAPE) 기준으로 4% 내외에 불과해 실용적으로 활용이 가능한 수준임을 보여주었다.

관련 연구들은 해외뿐 아니라 국내에서도 빠르게 늘고 있는 추세이다. 예를 들어, 배성완·유정석(2018)은 의사결정나무 기반 모형인 랜덤 포레스트, 그래디언트 부스팅 회귀 트리를 비롯한 다양한 기계학습 기법을 활용하여 아파트 가격지수에 대한 예측 성능을 전통적 시계열 모형과 비교하였다. 분석 결과, 특히 의사결정나무 기반 모형들이 시계열 분석 모형에 비해 전반적으로 우수한 예측력을 나타냈으며, 시장이 급격히 변동하는 상황에서 추세 예측의 정확성이 더욱 두드러지게 향상되었음을 밝혀냈다. 이인호·이경섭(2020)은 서울의 아파트 실거래가 자료 71만건을 토대로 XGBoost, 랜덤 포레스트, LightGBM과 같은 의사결정나무 양상불 기법의 설명력을 분석하였다. 분석 결과는 의사

결정나무 기반 알고리즘을 통해 전통적 헤도닉 모형에서 나타날 수 있는 다중공선성으로 인한 예측 왜곡을 극복할 수 있음을 보여준다. 홍정의(2021)는 주택시장의 복잡성을 랜덤 포레스트 모형이 어떻게 포착하는지에 대해 분석하고, 모형의 효율적 설계 방법을 제시하였으며, 홍정의(2020)는 의사결정나무 앙상블 모형 중 부스팅 기반의 세 알고리즘(XGBoost, LightGBM, CatBoost)을 통해 주택가격 대량 평가 모형을 구축하고, 성능과 특징을 비교 분석하였다. 논문은 2009년부터 2019년까지 서울에서 거래된 아파트 매매 데이터 620,617건을 통해 분석한 결과, 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 불과한 것으로 나타났다. 이는 의사결정나무 기반 모형의 예측력이 단순히 상대적인 측면(즉, 전통적인 헤도닉 모형에 비해)뿐 아니라, 절대적인 측면에서도 상당히 높게 나타나, 그 실용적 잠재력이 상당히 높다는 것을 보여준다.

이처럼 최근의 연구들은 의사결정나무 기반 모형의 성능과 활용 가능성에 대해 보여주고 있다. 이 시점에서 관련 연구는 단순히 예측 성능의 제시를 넘어, 높은 예측 정확성 등을 활용하여 어떻게 주택 시장을 보다 심층적으로 분석하고 이해하는 데에 활용될 수 있는지에 대한 탐구로 나아가는 추세이다. 초기 AVM 연구가 주로 개별 자산의 감정평가 정확도 향상에 집중되었다면(Antipov & Pokryshevskaya, 2012; Park & Bae, 2015), 최근에는 AVM이 산출하는 정확한 추정치를 시장 전체의 가격 동학을 관측하는 도구로 확장하려는 시도가 나타나고 있다. 예를 들어, Gallin et al.(2021)은 Zillow의 AVM을 활용하여 미국의 총 주택자산 가치를 추정하면서, AVM 기반의 재고 중심(stock-based) 측정이 기존의 반복매매지수나 설문 기반 자가추정보다 더 안정적인 결과를 제공할 수 있음을 보여준 바 있다. 이는 AVM이 단순한 개별 자산 감정평가의 도구를 넘어 시장 전체의 가격 수준을 측정하는 도구로 활용될 수 있음을 시사한다.

김이환 외(2022)는 네이버 부동산을 통해 수집한 아파트 정보에 랜덤포레스트와 인공지능망을 적용하여 서울시 아파트의 추정 매매가격을 산출한 뒤, 시점·자치구·권역별 시간더미 추정치를 활용하여 반기별 매매가격지수를 산정하였고, 그 지수를 통해 나타난 변동성이 KB지수나 한국부동산원 실거래가지수에 비해 가격 상승기에 더 크게 나타남을 보고함으로써 기계학습 기반 추정이 기존 지수의 평활화 한계를 보완할 수 있음을 시사하였다. 이소영·김경민(2025)은 LightGBM 기반 AVM을 통해 서울시 다세대주택 약 71,970개 단지(약 65만 호)의 월별 잠재가격을 추정한 후, 단지·기초구역·자치구·생활권 단위의 라스파이레스 가격지수를 산정함으로써 반복매매모형으로는 표본 한계 때문에 산출이 어려웠던 소지역 단위 가격지수의 가능성을 실증하였다.

두 연구는 본 연구와 유사한 방법론적 골격(AVM의 추정가격을 활용한 가중평균 산출)을 공유하나, 산출물의 형태와 분석의 범위에서 본 연구와 구별된다. 두 선행연구가 모두 가격지수의 산정과 그 평활화 한계의 보완을 1차적 목적으로 삼은 한편, 본 연구는 전국 단위의 전수적 재고 기반 추정을

목적으로 하고 있으며 이를 통한 재고 주택 가격 분포·하위집단별 통계량의 직접 관측에 초점을 맞춘다. 또한 자료의 측면에서 두 선행연구가 서울시에 한정된 단일 주택유형(아파트 또는 다세대주택)을 대상으로 한 반면, 본 연구는 전국 17개 시·도의 아파트 재고 주택 약 1,250만 호(전체의 99.1%)를 추정 대상으로 하여 분석 범위가 가장 넓다.

본 연구에서는 전국에 있는 개별 아파트 단지(단지 내 각 평형별)의 가격을 추정한 뒤, 지역별로 아파트 가격 수준과 분포, 그리고 소득 대비 주택가격 비율(PIR)이나 지니 계수와 같은 주요한 관련 지표들을 산출할 것이다. 이를 통해 본 연구에서 제시하는 접근을 통해 기존의 통계로는 직접 관측이 어려웠던 가격 변화의 추이 및 지역 간 차이 또는 지역 내 분포에 대한 현황을 효과적으로 관측할 수 있을 것으로 기대한다.

본 연구가 분석적 예시로 다루는 PIR과 지니계수는 산정 방식에 따라 결과가 상당히 달라질 수 있는 지표인 만큼, 그간의 관련 논의를 간략히 정리해둘 필요가 있다. PIR은 가계의 주택구입 부담을 측정하는 가장 대표적인 지표로, 국내에서는 KB국민은행, 한국주택금융공사, 국토교통부 주거실태 조사 등이 각기 다른 표본·통계지표·산정방식으로 PIR을 산출·공표하고 있는데, 그 결과 동일 시점에 대해서도 기관 간 수치가 상당한 차이를 보인다. 이창무 외(2012)는 PIR이 표본·지역 구분·통계 지표에 민감한 지표라는 점을 이론적·실증적으로 분석하고, 국제 비교를 위해서는 평균 기반·세전 정의를 채택하는 일관된 산정 방식이 요구됨을 제시하였다. 또한 권진웅·김남정(2020)은 서울 25개 자치구의 약 78만 건 실거래 자료를 활용하여 PIR을 자치구 단위로 산정한 후, 그 시계열 변동 패턴이 자치구별로 4개 유형으로 분화됨을 보였는데, 이는 미시 자료를 활용한 PIR 분석이 시·도 단위 평균보다 풍부한 정보를 산출할 수 있음을 시사한다.

한편 주택가격 분포의 불평등성에 대한 분석은 가구 자산 단위와 주택 가격 단위의 두 흐름으로 나누어 이해할 수 있다. 통계청의 가계금융복지조사를 활용한 분석들은 가구 단위 순자산 지니계수가 2017년 0.584에서 최근 0.625 수준까지 지속적으로 상승하였음을 보고하며, 이는 주택가격 상승이 가구 자산 불평등 심화에 결정적 역할을 했음을 의미한다(김지원·마강래, 2021; 송영호·마강래, 2025; 정의철 외, 2009). 한편 주택 가격 자체의 분포가 얼마나 불균등한지를 자산(주택) 단위에서 측정하려는 시도도 늘어나고 있다. Kang & Koo(2023)는 국토교통부 실거래 자료를 활용하여 서울 25개 자치구의 주택가격 지니계수를 산출하고, 이를 사회자본 수준과의 관계 분석에 활용하였다. 서재원·진장익(2021) 또한 주택가격 불평등이 계층이동 가능성에 대한 기대에 미치는 영향을 분석하였다.

이러한 관련 분석들은 필연적으로 이질적인 주택 가격과 소득 및 자산의 변화를 대규모로 추적하거나 그것을 대표하는 지표를 환산하는 작업을 수반하므로, 상당한 기술적 비용이 요구되었다. 본 연구에서 제시된 접근은 이러한 관련 분석을 기술적으로 용이하게 하며, 그 기술적 편의성을 바탕으로 보다 넓은 범위에 대해 접근할 수 있도록 한다. 본 연구는 이창무 외(2012)가 정리한 평균 기반·세전

정의를 따름으로써 국제비교성을 확보하면서, 동시에 17개 시·도 단위의 분기별 추정치를 일관된 방식으로 산출하며 지역 내 변화뿐 아니라 지역 간 비교를 가능하게 한다. 이처럼 본 연구는 주택 가격에 대한 보다 종합적이고 입체적인 정보를 제공함으로써 추후 다양한 후속 연구를 위한 토대가 될 수 있을 것으로 기대한다.

III. 분석 자료

본 연구에서는 전국 단위의 아파트 단지의 평형별 정보와 아파트 실거래 자료이다. 두 자료는 그 성격과 본 연구에서의 역할이 서로 구별된다. 단지 자료는 분석 시점 기준으로 전국에 존재하는 아파트 단지의 평형별 정보를 담고 있어 거래 여부와 무관하게 시장의 재고 상태 그 자체를 기술하며, 자료로 포착된 단지의 총 주택 호수는 12,517,131호로 2023년 인구주택총조사 기준 전국 아파트 12,631,608호의 약 99.1%에 해당하여 사실상 전국 아파트 재고의 거의 전수를 포괄한다. 한편 실거래 자료는 같은 단지들 가운데 2012년 3월 1일에서 2024년 4월 1일까지 약 12년의 관측 기간 동안 거래가 발생한 사례들의 가격 기록이며, 그 규모는 6,118,730건이다. 한 호의 주택이 해당 기간 동안 한 번도 거래되지 않을 수도, 여러 차례 거래될 수도 있으므로, 두 자료의 표본 수는 서로 단순한 부분집합 관계를 갖지 않는다. 본 연구의 분석 절차에서도 두 자료는 서로 다른 역할을 수행한다. 실거래 자료는 가격이 관측된 사례를 통해 주택 특성과 가격 사이의 관계를 학습시키는 데에 사용되며, 단지 자료는 학습된 모형의 추정 대상으로서, 모형이 단지 자료의 모든 단지·평형에 분기별로 적용됨으로써 거래 여부와 무관하게 전 단지의 시점별 추정 가격이 얻어진다.

본 연구의 단지 자료는 분석 시점을 기준으로 한 단일 시점의 단지·평형별 정보로 구성되며, 입지·구조 특성 또한 분석 시점 기준의 상태로 모든 추정 시점에 동일하게 적용된다. 즉, 분석 기간(2012년 3월~2024년 4월) 중에 발생한 지하철 노선의 개통, 학교의 신설·폐교, 재건축·재개발 추진단계의 변화 등은 분석 시점(2024년 4월 1일)의 상태로 통일되어 모형의 입력 변수로 사용된다. 이는 본 연구가 측정하고자 하는 대상이 시점에 따라 함께 변화한 입지·재고 특성을 반영한 시장 가격이 아니라, 동일하게 정의된 표본 바스켓에 대해 시점별 시장 환경이 부여한 가격 평가이기 때문이다. 이러한 의미에서 본 연구의 추정은, 표본 바스켓을 고정하여 산출하는 기존 가격지수의 확장된 형태에 가깝다. 이러한 설계는 본 연구가 산출하는 시점별 통계량(평균, 분위수, 지니계수 등)이 시점 간 표본 구성의 변화로부터 영향을 받지 않도록 함으로써, 전수적 추정에 기반하더라도 추정된 지표가 최대한 지수적 성격을 갖도록 가격 변화의 해석을 단순화하기 위한 것이다.

실거래 자료는 학습 기간으로 포함된 기간 중 발생한 모든 거래를 학습에 사용하였으며, 별도의

단지별 최소 거래 건수 기준은 설정하지 않았다. 즉 분석 기간 중 거래 빈도가 낮은 단지의 거래도 모두 학습에 포함되며, 단지·시점 단위의 거래 빈도 차이로 인한 표본의 불균형은 별도의 가중 조정 등을 적용하지 않은 채 모형의 학습에 반영하였다.

분석에서는 다양한 선행 연구들로부터 가격과의 관련성이 있는 것으로 알려진 총 27개의 변수가 사용되었다. 이들은 변수의 형식에 따라서는 수치형 변수와 범주형 변수로 나뉘며, 속성에 따른 분류로는 건물 특성과 단지 특성 및 입지적 특성으로 분류할 수 있다(〈표 1〉). 먼저, 건물 특성으로는 공급면적, 전용면적, 건축 구조, 층, 승강기의 유무, 복층의 유무, 1층 전용정원의 유무, 다락방의 유무가 포함된다. 이는 건축물 자체의 특성과 관련이 있다. 단지 특성으로는 건폐율, 용적률, 단지 내의 총 세대수, 대지 지분, 단지의 건설사 브랜드가 포함된다. 단지 특성은 광의의 건축적 특성에 해당하나, 개별 세대보다는 단지 전체에 포괄적으로 적용되는 특성들을 분류하였다. 입지적 특성은 가장 기본적인 입지 구분자인 위도와 경도, 배정되는 학교(초등학교, 중학교, 고등학교), 입지의 경사 수준, 용도지역, 가장 가까운 초등학교/중학교/고등학교/지하철로부터의 거리, 그리고 재개발 관련 상태를 나타내는 변수(사업유형, 안전진단등급, 추진단계)가 포함되었다.

〈표 2〉와 〈표 3〉은 이러한 변수들에 대한 기초 통계량을 보여주고 있다. 기초 통계량은 단지 자료와 실거래 자료 중, 단지 자료를 중심으로 계산하였는데 이는 실거래 자료가 모든 단지에 대한 정보가 아닌 해당 시점에 거래된 단지의 자료만을 담고 있을 뿐 아니라, 동일한 단지에 대한 중복된 거래 정보를 담고 있으므로 보다 거래가 빈번하게 이루어지는 단지의 정보를 더 많이 담고 있기 때문이다.

먼저 수치형 변수에 대한 기초 통계량을 살펴보자. 아파트 개별 세대 및 단지 특성을 나타내는 주요 변수들이 비교적 넓은 분포 범위를 가지고 있는 것으로 나타났는데, 이는 본 연구에서 사용된 자료가 전국의 모든 단지를 대상으로 하고 있기 때문이다. 공급면적의 경우 전체 표본의 평균은 약 94.7m^2 이며, 중앙값은 89.7m^2 로 나타났다. 최소값은 9.3m^2 , 최대값은 $1,023.9\text{m}^2$ 에 이르며, 표준편차는 42.4m^2 로 상당히 높은 편이다. 이는 소형 원룸형 주택부터 대형 펜트하우스까지 다양한 규모의 주택이 대상으로 포함되어 있음을 보여준다. 전용면적 역시 평균은 76.4m^2 , 중앙값은 74.9m^2 이며, 최소값은 7.3m^2 , 최대값은 464.1m^2 로 공급면적과 유사한 분포 특성을 보인다. 이러한 면적 관련 변수들은 개별 세대의 가격 결정에 반영될 수 있는 핵심 요소로 알려져 있다. 단지 특성 변수 중 세대수는 평균 약 360세대이며, 중앙값은 136세대로 나타난다. 그러나 최대값은 9,510세대에 이르며, 표준편차 또한 521로 나타난다. 이 역시 자료에 포함된 아파트 단지가 수십 세대 규모의 소규모 단지부터 수천 세대를 포함한 대규모 단지까지 매우 다양한 형태로 존재한다는 점을 보여준다. 용적률은 평균 301.3%, 중앙값 245.1%이며, 최대 1,472%까지 관측되었다. 이는 고밀도 개발이 허용된 도심 지역 일부 단지에서 나타나는 특성으로, 전체적으로는 평균과 중앙값의 차이가 큰 비대칭 분포를 보인다. 건폐율 역시 평균이 37.0% 수준이며, 중앙값은 31.6%, 최대값은 98.4%로 나타났다.

〈표 1〉 변수의 목록

분류	변수명	형식	조작적 정의
건물특성	공급면적	수치형	전용면적과 공용면적의 합(m ²)
	전용면적	수치형	해당 세대만이 독점적으로 사용하는 면적(m ²)
	층	수치형	해당 세대가 위치한 층 (지하층은 음수로 표기, 예: 지하 1층=-1)
	승강기	범주형(0/1)	해당 단지의 승강기 설치 여부(더미: 1/0)
	복층여부	범주형(0/1)	해당 세대가 복층 구조인지 여부(더미: 1/0)
	1층 전용정원	범주형(0/1)	1층 세대에 부속된 전용정원의 유무(더미: 1/0)
	최상층 다락방	범주형(0/1)	최상층 세대에 부속된 다락방의 유무(더미: 1/0)
단지특성	구조	범주형(43범주)	건축 구조의 유형 (예: 철근콘크리트구조, 철골철근콘크리트구조 등)
	건폐율	수치형	단지 대지면적 대비 건축면적의 비율(%)
	용적률	수치형	단지 대지면적 대비 건축 연면적의 비율(%)
	세대수	수치형	단지 내 총 가구 수(호)
	대지지분	수치형	개별 주택에 귀속되는 단지 내 대지지분(m ²)
입지특성	브랜드	범주형(17범주)	단지의 건설사 브랜드(예: 일반, 힐스테이트 등)
	위도	수치형	단지 대표 좌표의 위도(도)
	경도	수치형	단지 대표 좌표의 경도(도)
	초등학교	범주형(4,082범주)	단지에 학군 배정된 초등학교 식별자
	중학교	범주형(2,398범주)	단지에 학군 배정된 중학교 식별자
	고등학교	범주형(939범주)	단지에 학군 배정된 고등학교 식별자
	고저	범주형(7범주)	단지 입지의 경사 수준(예: 급경사지, 완경사지 등)
	용도지역	범주형(20범주)	단지가 속한 「국토의 계획 및 이용에 관한 법률」상 용도지역(예: 제1종일반주거지역, 제2종일반주거지역 등)
	사업유형	범주형(8범주)	단지에 적용되는 정비사업의 유형(예: 주거환경개선, 공공재개발 등). 해당 없는 단지는 별도 해당없음 범주로 처리
	안전진단등급	범주형(6범주)	재건축 안전진단을 받은 단지에 부여된 등급(예: C, D 등). 미진단 단지는 별도 해당없음 범주로 처리
	추진단계	범주형(14범주)	정비사업의 추진 단계(예: 해당 없음, 정비구역지정 등)
	초등학교 거리	수치형	단지에서 가장 가까운 초등학교까지의 직선거리(m). 2km 초과 시 2km로 상한 처리
	중학교 거리	수치형	단지에서 가장 가까운 중학교까지의 직선거리(m). 2km 초과 시 2km로 상한 처리
	고등학교 거리	수치형	단지에서 가장 가까운 고등학교까지의 직선거리(m). 2km 초과 시 2km로 상한 처리
	지하철 거리	수치형	단지에서 가장 가까운 지하철역까지의 직선거리(m). 2km 초과 시 2km로 상한 처리. 지하철이 부재한 지역에서는 모형에서 제외

주 : 1) 지하철이 없는 지역의 경우, 지하철 거리 변수를 제외하고 모형을 학습 및 예측하였음.
 2) 모든 변수는 분석 시점 기준(2024.4.1.)의 단지·평형 상태를 기록한 값이며, 분석 기간 전체에 걸쳐 동일하게 적용된다.

〈표 2〉 기초 통계: 수치형 변수

	평균	최소값	중앙값	최대값	표준편차
건폐율	36.97	1.17	31.58	98.38	19.42
용적률	301.27	0.09	245.14	1,472	177.61
세대수	359.76	1	136	9,510	521.41
공급면적	94.66	9.26	89.69	1,023.91	42.43
전용면적	76.44	7.32	74.9	464.11	35.15
대지지분	39.39	0	37.25	31,257.4	75.34
층	11.83	-4	10	84	8.34
초등학교 거리	434.1	34	383	1,992	256.91
중학교 거리	613.02	40	546	1,999	360.41
고등학교 거리	606.89	0	574	1,998	508.67
지하철 거리	500.55	19	466	1,997	259.17
총 재고 표본수 (모든 평형)	1,128,355				

〈표 3〉 범주형 변수의 빈도

변수명	범주값	호수
승강기	N	718,762
	Y	11,798,369
복층여부	N	12,502,134
	Y	14,997
1층전용정원	무	12,516,740
	유	391
최상층다락방	무	12,506,782
	유	10,349
고저	고지	18,761
	급경사지	33,817
	없음	754
	완경사지	1,313,740
	저지	6,385
	평지	10,704,984
	미표기	438,690
사업유형	재개발(주택)	201,189
	재개발(도시환경)	18,657

〈표 3〉 계속

변수명	범주값	호수
사업유형	재건축(공동)	336,068
	재건축(단독)	36,381
	주거환경개선	43,426
	공공재개발	4,560
	공공재건축	270
	해당없음	11,876,580
안전진단등급	C	6,298
	D	218,512
	E	12,556
	F	4,703
	G	4,865
	해당없음	12,270,197

이외에도 수치형 변수로는 층수나 대지지분, 위도·경도 등 수치형들이 포함되어 있다.³⁾ 초등학교, 중학교, 고등학교, 지하철로부터의 거리는 최대 2,000미터까지만 계산되었으며 그 이상인 경우 최대 값으로 분류되었다. 또한, 분석대상 지역 중, 지하철이 존재하지 않는 지역에서는 모형의 학습 및 예측 과정에 해당변수(지하철역으로부터의 거리)를 제외하였다.

또한, 위도와 경도는 다른 수치형 변수와 달리 특정한 속성의 다소를 의미하는 정보라기보다는 위치값의 구분자 그 자체를 의미한다. 본 연구가 위도·경도를 별도의 변환 없이 입력 변수로 그대로 사용하는 것은 다음의 두 가지 근거에 따른 것이다. 첫째, 본 연구의 목적은 개별 변수의 가격 효과에

3) 언급된 것처럼 본 연구의 자료는 분포의 양 극단에 위치하는 비통상적 표본을 일부 포함하고 있다. 본 자료는 한국부동산원이 사전 검수를 거친 것이나, 자료 입력 단계의 정합성이 곧 모든 표본의 분석상 적합성을 보증하는 것은 아니므로, 본 연구에서는 비통상적 값을 갖는 표본에 대해 추가적인 검토를 수행하였다. 구체적으로, 층수, 공급·전용면적, 세대수, 대지지분 등 주요 변수의 분포에서 극단에 위치한 표본을 추출하여 단지명과 주소를 식별한 뒤, 공개 자료와의 대조를 통해 해당 단지의 실재 여부를 확인하였다. 다만 개별 단지의 변수가 입력 단계에서 얼마나 정확하게 코딩되었는지를 변수 단위로 완전히 검증하기 위해서는 사실상 현장 조사에 준하는 작업이 요구되며, 이는 본 연구의 범위를 넘어선다. 비통상적 값을 갖는 표본 중 일부에 입력 단계의 미세한 부정확성이 존재할 가능성은 100% 배제하기 어려우며, 이는 대규모 행정 자료에 기반한 연구가 일반적으로 갖는 한계의 하나로 이해될 수 있다. 한편 본 연구는 전국 사실상 전수의 아파트 단지를 추정 대상으로 삼고 있어, 통계적 기준이나 임의의 절단 기준을 적용하여 분포의 양 극단에 위치한 표본을 일률적으로 "이상치"로 분류·제외하는 처리는 전수 추정을 핵심 기여로 하는 본 연구의 설계와 상충되는 측면이 있다. 또한 II장에서 설명한 바와 같이 트리 앙상블 알고리즘은 분포의 양 극단에 위치한 표본에 대해 구조적으로 강건한 특성을 가지므로, 비통상적 표본이 결과 통계량에 미치는 영향은 제한적일 것으로 판단된다. 이상의 검토를 바탕으로, 본 연구에서는 비통상적 값을 갖는 표본을 별도로 분류·제외하지 않고 자료 전체를 그대로 분석에 활용하였다.

대한 해석이 아니라 시점별·단지별 가격 추정 그 자체에 있으며, 트리 양상블 알고리즘은 위도·경도와 같이 변환되지 않은 좌표 자체를 입력으로 받아 입지 효과를 비선형적·국지적으로 포착하는 데에 효율적임이 이론적으로 알려져 있다. 둘째, 기존 연구들은 트리 양상블 모형에서 위도·경도를 직접 입력으로 사용하는 방식이 일관적으로 가장 높은 예측 정확성을 제공함을 보고해 왔다(홍정의, 2020, 2021). 특히 위도·경도를 입력 변수로 사용한 모형에 행정구역 더미 등을 추가로 투입할 경우 예측력이 추가로 개선되지 않으며, 오히려 변수 수의 증가로 인한 효율성 저하가 발생할 수 있음이 보고된다(홍정의, 2021). 이러한 점들을 종합적으로 고려하여, 본 연구에서는 위도·경도를 직접 입력 변수로 사용하되 별도의 행정구역 단위 위치 변수는 추가하지 않는 방식을 채택하였다.

범주형 변수의 경우, 전반적으로 이진형 범주값을 가지며, 분포는 다소 불균형한 편이다. 예를 들어, 승강기 유무는 ‘있음’이 전체의 약 94%를 차지하고 있으며, ‘없음’은 약 6%에 불과하다. 이는 최근 수십 년간 건설된 아파트 대부분에 승강기가 설치되어 있는 현실을 반영한 결과이다. 한편, 복층 여부와 1층 전용정원의 경우, ‘있음’ 응답은 전체의 각각 약 1%와 0.1% 수준에 불과해 상당히 극히 드문 특성으로 확인되었다. 이는 대부분의 아파트가 단층 구조로 이루어져 있으며, 이러한 구조는 고급 주택이나 특정 단지에서 제한적으로 나타나는 선택적 구조임을 보여준다. 또한 초등학교·중학교·고등학교 변수는 분석 시점(2024년 4월 1일) 기준으로 해당 단지에 학군 배정된 학교의 학교명으로, 이는 입지 가치에 주요한 설명력을 가질 수 있는 것으로 알려져 있다(김세울·유선종, 2022). 이러한 학교 변수는 위도·경도와 상보적인 지리 식별자로서 모형에 포함되었다. 우리나라의 학군 배정 체계는 행정구역 단위보다 세밀한 공간적 구획을 자연적으로 형성하며, 그 경계는 연속적 변수인 위도·경도가 학습된 트리 구조 내에서 직접 포착하기 어려운 미시적 불연속성을 갖는다. 즉, 위도·경도가 연속적 입지 효과를 비선형적으로 포착하는 데에 효율적이라면, 학교 변수는 학군 경계에서 발생하는 단절적 입지 효과를 범주형으로 직접 식별하는 역할을 수행한다.⁴⁾⁵⁾

- 4) 이는 본 연구의 분석에 행정구역 더미를 추가하지 않은 결정과 모순되지 않는데, 행정구역 더미는 위도·경도가 이미 포착하는 광역 단위 입지 효과와 상당 부분 중복되는 반면, 학교 배정 단위는 행정구역보다 세밀하고 광역 좌표가 직접 포착하기 어려운 단절적 경계를 식별하기 때문이다.
- 5) 학교 변수는 위와 같이 4,082개에 이르는 범주 수를 가지나, 본 연구의 사용 맥락에서 범주 수 자체가 과적합 위험을 특별히 높인다고 보기는 어렵다. 본 연구가 사용하는 부스팅 트리 알고리즘은 분할 기준의 선택 단계에서 정규화 항을 통해 트리의 복잡성을 제어하며, 학습 표본의 규모(약 612만 건의 실거래) 대비 학교 변수의 범주 수는 범주당 평균 1,000건 이상의 학습 표본을 확보할 수 있는 수준이다. 또한 범주의 수가 많다는 것은 곧 학교 변수가 단지의 미시적 위치를 더 세밀하게 식별한다는 것을 의미하므로, 학습 표본의 밀도가 충분히 확보되는 한 미시적 식별력 자체가 모형의 일반화 성능을 저해할 이유는 없다. 본 변수의 시간적 자료 누설(temporal leakage) 가능성에 대해서도, 학교 배정 정보는 위에서 밝힌 바와 같이 분석 시점(2024년 4월 1일)의 상태로 전 시점에 동일하게 적용되므로, 추정 대상 시점의 정보를 사후적으로 사용하는 형태의 누설이 발생하지 않는 구조이다. 이러한 처리 방식의 적절성은 IV장의 평가 결과를 통해 실증적으로 뒷받침된다. 무작위 3-겹 교차검증 및 시점 기반 평가 모두에서 학습-평가 표본 간 성능의 일관성이 확인되며, 학교 변수에 기인한 과적합의 징후가 별도로 관측되지 않는다.

마지막으로, 본 연구에서는 특정 변수에 대한 결측치가 포함된 표본의 경우, 이를 제외하지 않고 해당 변수의 값에 '결측'에 해당하는 값을 할당한 뒤 그대로 모형의 학습에 사용하였다. 구체적으로, 범주형 변수의 결측치에는 '관측안됨'이라는 별도의 범주를 부여하였고, 수치형 변수의 결측치에는 해당 변수의 최소값보다 작은 특수값을 부여하여 결과적으로 XGBoost의 내장 결측 처리 방식과 동일한 분기가 이루어지도록 하였다. 다음 장에서는 예측 모형이 실제 가격을 얼마나 정확하고 안정적으로 추정하는지에 대한 평가 결과를 검토한다.

IV. 기계학습 기반 모형의 정확성 및 안정성 평가

1. 평가 개요 및 지표

본 연구는 XGBoost 알고리즘에 기반한 자동산정모형을 활용하여 개별 아파트의 가격을 추정하고, 이를 통해 전체 시장의 가격 분포와 관련 지표들을 도출하는 것을 목적으로 한다. 이러한 분석은 자동산정모형의 성능에 대한 신뢰를 전제로 이루어지며, 만약 모형의 예측 정확성이 일정 수준에 미치지 못할 경우, 해당 모형을 기반으로 한 추정 결과 역시 신뢰하기 어려울 것이다. 그러므로, 결과 자체를 제시하기에 앞서 본 연구에서 사용된 모형의 성능을 평가하는 것이 필수적이다.⁶⁾

따라서, 본 장에서는 자동산정모형을 통한 추정결과를 제시하기 이전에, 먼저 XGBoost 알고리즘에 기반한 자동산정 모형을 통해 분석하려는 자료에 대해 실제로 예측을 수행하고 그 정확성과 편향성을 검증한다. 본 연구에서는 이를 위해 3겹 교차 검증(three-fold cross-validation)을 실시하였다. 본 연구에서는 2012년 3월 1일부터 2024년 4월 1일 사이에 이루어진 아파트 실거래 자료를 각 지역별로 임의 분할 방식에 따라 3개의 하위 그룹으로 나누며, 각 반복에서 하나의 그룹을 평가 표본으로 설정하고 나머지 두 개를 학습 표본으로 사용하여 순차적으로 세 그룹 모두에 대한 평가를 수행하였다. 이와 같은 교차 검증은 모형의 일반화 능력을 보다 신뢰성 있게 판단할 수 있는 방법론으로 널리 사용된다. 자동산정모형을 활용한 주택 가격 예측 연구에서는 종종 시점을 기준으로 학습 표본과 평가 표본을 구분하기도 하지만, 본 연구는 과거 다양한 시점의 가격 분포를 재구성하려는 목적을 가지고 있으며, 이에 따라 전체 시점을 아우르는 임의 분할 방식이 실제 모형의 성능을 가장 잘 반영할 수 있는 평가 방식이라 판단된다.

6) 본 연구는 추정 모형의 변수 간 관계의 해석을 목적으로 하지 않으며, 모형의 추정치를 도구로 삼아 시장 전체의 가격 수준, 분포, 격차, 불균등성, 부담 등을 관찰하는 데에 분석의 초점을 둔다. 따라서 모형의 적절성에 대한 평가는 추정치의 정확성·편향성·안정성을 중심으로 이루어지며, 본 절에서는 이를 위한 평가 지표와 결과를 제시한다.

모형의 성능 평가는 다음과 같은 지표들을 통해 이루어졌다: 평균절대백분오차(MAPE), 산포계수(coefficient of dispersion, COD), 평균백분오차(mean percentage error, MPE), 오차 허용 구간 내 예측 비율(hit rate). 먼저, 가장 대표적인 오류 측정 지표인 MAPE는 예측값과 실제값 간의 차이를 백분율 기준으로 절대값 처리한 뒤 평균을 낸 지표로, 아래와 같이 표현된다.

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

이때 $n, \hat{p}_i, p_i$ 는 각각 표본수, i 번째 주택에 대한 예측값, i 번째 주택의 실제값이다. MAPE는 예측의 전반적인 정밀도를 쉽게 가늠하게 해준다. 이 지표는 개별 주택 단위에서 예측 오차가 얼마나 발생했는지를 직관적으로 보여주며, 가격 추정의 평균적 정확성을 가장 쉽게 파악할 수 있는 지표로 널리 활용된다.

한편 MAPE가 발생한 예측 오류의 평균적 수준을 측정하는 지표라면, COD와 MPE는 오류의 특성을 더 구체적으로 파악하는 데에 적합하다. 즉, 예측 오차는 비일관성(실제는 상대적으로 높는데 상대적으로 낮은 예측을 하는 경향)과 편향성(실제에 대해 일관적으로 더 높거나 낮은 값을 내는 경향)과 모두에 의해 발생할 수 있는데, COD는 비일관성(또는 불균형성)을 MPE는 편향성을 측정하는 데에 적합하다.

COD는 국제평가관협회(International Association of Assessing Officers, IAAO)가 대량평가 모형의 표준 평가 지표로 권고하는 산포 지표(IAAO, 2017)로, 추정가격과 실거래가의 비(sales ratio)가 그 중앙값으로부터 평균적으로 얼마나 벗어나 있는지를 백분율로 측정한다. 이때 COD를 계산하는 식은 아래와 같이 표현된다.

$$COD = \frac{100}{SR_m} \left[\frac{\sum_{i=1}^n |SR_i - SR_m|}{n} \right], \quad \langle \text{식 2} \rangle$$

이때 SR_i, SR_m, n 은 각각 i 번째 예측값과 실제값의 비율, 예측값-실제값 비율의 중앙값, 표본 크기를 의미한다. COD는 값이 작을수록 추정의 일관성이 높음을 의미하며, IAAO(2017)는 주거용 부동산의 경우 일반적으로 15 이하, 그 중에서도 비교적 동질적인 시장(예: 아파트)에서는 10 이하의 COD 값을 정확성·일관성이 양호한 수준으로 권고한다. MAPE가 평균적 오차의 절대적 크기를 측정한다면, COD는 sales ratio의 중앙값을 기준으로 한 산포의 정도를 측정하므로, 예측이 일관적으로 편향되어 있더라도(즉 모든 추정값이 일정한 비율로 과대 또는 과소 평가되더라도) 그 편향이 표본

전반에 일관적으로 나타나는 한 COD 값은 작게 유지될 수 있다는 특징이 있다. 즉, COD는 모형의 편향 여부와는 별개로 추정의 일관성·정밀도를 평가하는 데 적합하다.

한편 MPE(평균백분오차)는 예측값과 실제값 간의 차이를 백분율로 나타내되 그 부호를 유지한 채 평균을 계산하는 지표로 아래와 같이 표현된다.

$$MPE = \frac{100}{n} \sum_{i=1}^n \frac{\hat{p}_i - p_i}{p_i} \quad \langle \text{식 3} \rangle$$

이때 $n, \hat{p}_i, p_i$ 는 각각 표본수, i 번째 주택에 대한 예측값, i 번째 주택의 실제값이다. MPE는 오차의 크기 보다는 예측값이 체계적으로 과소평가되었는지, 혹은 과대평가되었는지를 판단하는 데 사용될 수 있다.

특히 본 연구와 같이 시장의 평균 가격, 분위수, 지역별 통계량 등 집단 단위의 지표를 생성하려는 목적에서는, 개별 예측에 다소의 오차가 있더라도 MPE가 0에 가까우면 예측 오차들이 서로 상쇄되어 집단의 평균적 특성은 왜곡 없이 잘 반영될 수 있다. 예를 들어, 어떤 지역 내 아파트의 가격에 대한 개별 예측치들이 실제보다 다소 높거나 낮게 나온다고 하더라도, 만약 그 예측치의 편향성(MPE가 0에 가까우면)이 낮다면 예측 오류는 평균적으로 상쇄되어 지역의 평균치는 실제의 평균에 가깝게 관측될 수 있다. 반대로 MPE가 일정하게 뚜렷한 양의 값을 가진다면, 예측값은 일관되게 실제보다 높게 나타나며, 그 집단의 평균 가격이나 전체 분포의 중심이 실제보다 상향 왜곡되는 문제가 발생한다. 따라서 본 연구와 같이 개별주택에 대한 추정을 통해 지역별 평균이나 분포 등 통계량을 관측하려는 경우, 특히 예측모형의 MPE가 신뢰할만큼 낮아야 할 것이다. 마지막으로 hit rate는 예측값이 실제값과 비교하여 특정 허용 오차 범위($\pm 10\%$, $\pm 20\%$, $\pm 30\%$) 이내에 포함되는 비율을 측정한 것이다. 이는 연구자가 허용 가능한 오차 기준에 따라 예측 결과의 수용 가능성을 판단할 수 있도록 하는 보조적 지표가 될 수 있다.

2. 자동산정모형의 정확성 및 불편향성 평가 결과

〈표 4〉는 전국 17개 행정구역(시·도)를 대상으로 실시한 XGBoost 기반 자동산정모형의 성능 평가 결과를 요약한 것이다. 아파트 실거래 자료를 3겹 교차 검증 방식으로 나누어 총 세 차례의 반복 실행을 진행하였으므로, 각 지역마다 3회의 결과가 나타나있다. 따라서 본 연구에서 제시된 성능평가의 결과는 모형이 특정 지역이나 시점이 아닌 일정 수준 이상의 성능을 일관적으로 유지할 수 있는지에 대한 관측 결과를 동시에 제공한다.

〈표 4〉 XGBoost 기반 AVM의 성능평가(3-fold 평가)

지역	분할	MAPE	COD	MPE	〈10%〉	〈20%〉	〈30%〉
서울	1	4.36	4.36	0.51	91.88	98.66	99.57
	2	4.30	4.29	0.46	91.88	98.63	99.56
	3	4.29	4.29	0.48	91.96	98.68	99.57
부산	1	5.25	5.25	0.56	86.85	97.37	99.10
	2	5.27	5.27	0.48	86.75	97.40	99.12
	3	5.29	5.29	0.51	86.89	97.37	99.07
대구	1	4.57	4.57	0.39	90.19	98.26	99.41
	2	4.57	4.57	0.36	90.17	98.32	99.42
	3	4.56	4.56	0.36	90.32	98.34	99.41
인천	1	4.27	4.27	0.32	91.52	98.64	99.59
	2	4.28	4.28	0.31	91.57	98.62	99.58
	3	4.27	4.27	0.27	91.67	98.63	99.59
광주	1	5.04	5.04	0.41	87.82	97.90	99.27
	2	5.04	5.05	0.40	87.66	97.86	99.27
	3	5.07	5.07	0.44	87.56	97.84	99.23
대전	1	4.48	4.48	0.31	90.59	98.66	99.62
	2	4.49	4.49	0.37	90.57	98.62	99.57
	3	4.46	4.46	0.27	90.66	98.70	99.63
울산	1	5.27	5.27	0.47	86.47	97.43	99.21
	2	5.27	5.27	0.46	86.36	97.42	99.19
	3	5.37	5.37	0.58	86.06	97.20	99.06
세종	1	5.07	5.07	0.56	87.54	97.86	99.35
	2	5.06	5.06	0.45	87.87	97.90	99.27
	3	5.12	5.12	0.50	87.65	97.67	99.25
경기	1	3.95	3.95	0.27	93.25	99.03	99.69
	2	3.94	3.94	0.26	93.35	99.04	99.71
	3	3.97	3.97	0.30	93.24	99.02	99.70
강원	1	5.87	5.88	0.58	83.46	95.84	98.47
	2	5.83	5.83	0.64	83.57	95.89	98.50
	3	5.84	5.85	0.62	83.43	95.94	98.48
충북	1	5.72	5.73	0.55	84.72	96.22	98.49
	2	5.73	5.73	0.64	84.80	96.09	98.42
	3	5.72	5.72	0.59	84.91	96.21	98.53

〈표 4〉 계속

지역	분할	MAPE	COD	MPE	〈10%〉	〈20%〉	〈30%〉
충남	1	4.81	4.81	0.44	88.45	97.65	99.25
	2	4.82	4.82	0.37	88.42	97.68	99.19
	3	4.84	4.84	0.41	88.31	97.74	99.23
전북	1	5.96	5.97	0.55	82.84	96.18	98.70
	2	5.94	5.94	0.61	82.97	96.18	98.68
	3	5.96	5.96	0.64	82.97	96.16	98.71
전남	1	6.08	6.08	0.61	81.62	95.76	98.64
	2	6.06	6.06	0.65	81.89	95.83	98.58
	3	6.09	6.09	0.60	81.85	95.82	98.59
경북	1	6.67	6.67	0.76	79.97	94.54	97.94
	2	6.67	6.67	0.86	79.81	94.58	97.93
	3	6.73	6.73	0.85	79.81	94.57	97.86
경남	1	5.46	5.46	0.56	85.53	96.87	98.93
	2	5.44	5.44	0.51	85.58	96.99	99.03
	3	5.44	5.44	0.52	85.65	97.01	99.00
제주	1	7.33	7.33	1.29	77.31	93.15	97.18
	2	7.41	7.41	1.17	76.82	92.87	97.13
	3	7.24	7.24	0.89	77.46	93.24	97.32

주: 소수점 3자리 이하는 반올림되었다.

AVM, automated valuation model; MAPE, mean absolute percentage error; COD, coefficient of dispersion; MPE, mean percentage error.

먼저 평균절대백분오차(MAPE)를 기준으로 살펴보면, 수도권 3개 시·도(경기 3.95, 서울 4.32, 인천 4.27)와 대구(4.57), 대전(4.48), 충남(4.82) 등 6개 지역에서 4%대 또는 그 이하의 낮은 오차율이 관찰되었다. 그 외 광역시(부산, 광주, 울산, 세종)와 충북·경남·강원·전북 등 8개 지역은 5%대로 나타났으며, 전남(6.08), 경북(6.69), 제주(7.33)는 6%를 상회하였다. 비수도권 도지역, 특히 제주·경북·전남에서 상대적으로 큰 오차가 관측되는 것은 이들 지역의 시장 구조가 단지 간 이질성이 크고 거래 빈도가 낮아 학습 표본의 밀도가 낮은 데에서 비롯된 것으로 해석된다. 그럼에도 가장 오차율이 큰 제주조차 7%대에 머무르며 모든 지역에서 MAPE가 한 자릿수 범위 내에 머무른다는 점은, 모형의 예측 정밀도가 비교적 안정적인 수준에서 유지되고 있음을 보여준다.

이러한 경향은 산포계수(COD)에서도 유사하게 관측된다. 경기(3.95), 인천(4.27), 서울(4.31), 대전(4.48), 대구(4.57), 충남(4.82) 등 6개 지역은 COD가 5 미만으로 나타났으며, 부산·광주·세종·

울산·경남·충북·강원·전북 등 8개 지역은 5~6 수준이었다. 전남(6.08), 경북(6.69), 제주(7.33)의 3개 지역에서만 6을 다소 상회하였다. IAAO(2017)는 주거용 부동산에 대해 일반적으로 15 이하, 동질적 시장에서는 10 이하의 COD를 양호한 수준으로 권고하는데, 본 연구의 추정치는 가장 높은 제주(7.33)도 10 이하 수준을 큰 폭으로 하회하여 17개 시·도 전체에서 IAAO 권고 기준을 충족한다.

또 한 가지 추정치의 활용 가능성을 높이는 지표인 평균백분오차(MPE)의 경우, 제주(1.12)를 제외한 16개 지역 모두에서 $\pm 1\%$ 이내로 수렴하였다. 이는 모형이 지역별로 체계적인 과대 또는 과소 예측 없이 거의 불편향적 예측을 수행함을 의미한다. 특히 경기(0.28), 인천(0.30), 대전(0.32), 대구(0.37)와 같이 거래량이 풍부한 수도권·광역시에서는 MPE가 0.3% 내외의 매우 낮은 수준에서 안정적으로 유지되었으며, 강원·전남·경북 등 비수도권 지역에서도 0.6%~0.8% 범위 내에서 예측 편향이 관리되었다. 유일하게 1%를 상회한 제주의 경우에도 절대 수준 자체는 1.12%에 머물러, 지역 단위 평균이나 분포 통계량의 산출에 실질적 왜곡을 초래할 정도의 편향은 아닌 것으로 평가된다. 이는 본 연구가 추구하는 '지역 단위 가격지표 산출'의 전제조건이 대체로 충족됨을 보여준다.

마지막으로 예측값과 실제값의 차이가 일정 수준 이하로 유지되는지를 평가하는 hit rate 분석에서는 지역 간 분포의 폭이 비교적 뚜렷하게 관측된다. $\pm 10\%$ 이내 hit rate는 수도권 3개 시·도와 대구·대전에서 90% 이상(경기 93.3%, 서울 91.9%, 인천 91.6%, 대전 90.6%, 대구 90.2%)을 보였고, 그 외 광역시 및 충남·경남에서는 85%~89% 수준(부산 86.8%, 광주 87.7%, 울산 86.3%, 세종 87.7%, 충남 88.4%, 경남 85.6%)이었다. 반면 강원·충북·전북·전남 등 비수도권 도지역은 81%~85% 수준에 머물렀으며, 경북(79.9%)과 제주(77.2%)에서는 80% 미만의 가장 낮은 hit rate가 관측되었다. 허용 오차 기준을 완화하면 지역 간 격차는 빠르게 좁혀진다. $\pm 20\%$ 기준에서는 경북(94.6%)과 제주(93.1%)를 제외한 15개 지역에서 95% 이상, $\pm 30\%$ 기준에서는 경북(97.9%)과 제주(97.2%)를 제외한 모든 지역에서 98.5% 이상이 포함된다. 이는 모형이 일부 비수도권 도지역에서 $\pm 10\%$ 라는 좁은 허용 오차 기준에 대해 다소 낮은 적중률을 보이긴 하나, 극단적인 예측 오차의 발생 자체는 전 지역에 걸쳐 매우 제한적임을 보여준다.

종합하면, 본 연구에서 사용된 XGBoost 기반 자동산정모형은 전국 17개 광역 시·도 모두에서 IAAO(2017) 권고 기준을 충족하는 산포 수준과, 제주를 제외한 전 지역에서 $\pm 1\%$ 이내의 낮은 예측 편향을 일관적으로 보여주었다. 지역 간 정확성에서 일정한 격차가 존재하며(특히 거래 빈도가 낮고 단지 간 이질성이 큰 일부 비수도권 도지역에서 좁은 허용 오차 기준의 적중률이 다소 낮음) 이는 학습 표본의 밀도 차이로부터 비롯된 자연스러운 결과로 해석된다. 그럼에도 모든 지역에서 산포 및 편향성 기준이 일정 수준 이상의 안정성을 유지하고 있다는 점은, 본 모형의 추정이 개별 주택 가격 예측뿐 아니라 지역별 평균·분위수·분포 특성 등 집단 통계량 추정에도 활용 가능한 신뢰성을 갖춤을 시사한다.

3. 시점 기반 평가를 통한 강건성 검토

본 연구의 IV.2절에서 활용한 무작위 3-겹 교차검증은 본 연구의 분석 목적, 즉 과거 각 시점의 가격 분포를 재구성하기 위한 예측에 부합하는 평가 방식이다. 분석 모형은 시점 정보를 공변량으로 포함하고 있으며, 본 연구가 추정하고자 하는 가격은 학습 자료가 관측된 시점 범위 내(in-sample period)에 위치하는 가격이기 때문이다. 그러나 이러한 임의 분할 방식은 동일 시점의 거래가 학습 평가 표본에 동시에 포함될 수 있어, 성능 지표가 시점 누설(temporal leakage)에 의해 다소 낙관적으로 평가될 가능성을 완전히 배제하기는 어렵다. 따라서 본 절에서는 이러한 가능성에 대한 보완적 검토로서 시점 기반 평가(time-based holdout)를 추가로 수행한다.

평가 절차는 다음과 같다. 평가 시점은 1개월 단위로 설정하고, 학습 표본은 해당 평가 시점 직전 3년 간의 실거래 자료로 한정한다. 즉 평가 표본의 거래 시점은 학습 표본의 시점 범위와 중복되지 않으며, 이를 통해 시점 누설의 영향을 원천적으로 통제한다. 본 절에서는 분석 기간 전체에서 4개의 대표 평가 시점—2015년 1월, 2018년 1월, 2021년 1월, 2024년 1월—을 학습 기간과 동일한 3년 간격으로 선정하여 결과를 제시한다. 이 네 시점은 시장의 안정기, 상승 직후, 급등기 직후, 조정기를 각각 포괄하며, 다양한 시장 국면에서 모형 성능의 강건성을 함께 점검할 수 있게 한다.

〈표 5〉는 17개 시도별 시점 기반 평가 결과를 정리한 것이다. 결과를 요약하면 다음과 같다. 첫째, 17개 시·도의 평균 MAPE는 2015년 1월 6.07%, 2018년 1월 6.24%, 2021년 1월 8.13%, 2024년 1월 6.61%로 나타났다. 이는 무작위 3-겹 교차검증 결과인 평균 4% 내외에 비해서는 다소 상승한 수준이나, IAAO(2017)가 동질적 시장에 대해 권고하는 산포 기준 등을 고려할 때 절대적 수준에서 여전히 양호한 범위 내에 위치한다.

둘째, 시점별 결과를 비교하면 2021년 1월의 평가 시점에서 성능이 가장 낮게 나타나는데, 이는 해당 시점이 2020~2021년의 급격한 가격 상승 구간에 위치한 것과 관련된다. 특히 울산(MAPE 11.34), 경북(9.12), 부산(9.10), 제주(9.74) 등 가격 변동성이 크게 나타난 일부 지역에서 9%~11% 대의 높은 MAPE가 관측된다. 즉, 학습 표본(2018년 1월~2020년 12월)이 후속 시기의 급격한 구조 변화를 사전적으로 반영하지 못하기 때문에 일시적인 성능 저하가 발생한 것으로 해석된다. 이는 시점을 가로지르는 외삽(extrapolation)이 모형의 일반화에 도전적인 과제임을 시사하나, 본 연구가 학습 시점 범위 내 시장의 가격 분포를 재구성하는 것을 목적으로 하고 있다는 점에서 본 연구의 분석 결과 자체의 신뢰성에 대한 직접적 위협으로 해석되지는 않는다. 한편 2024년 1월 시점의 부산(MAPE 10.17, MPE +4.14)에서도 비교적 큰 양의 편향이 관측되는데, 이는 학습 기간(2021년 1월~2023년 12월) 동안 부산 시장이 조정기에 들어선 후 평가 시점에서 일부 회복 또는 재상승의 신호가 발생하여 모형의 외삽이 어려웠던 결과로 해석된다.

〈표 5〉 XGBoost 기반 AVM의 성능평가(time-holdout 평가)

지역	시점	MAPE	COD	MPE	〈10%〉	〈20%〉	30%
서울	2015/01	4.00	3.84	-1.27	93.91	99.31	99.85
	2018/01	5.59	5.38	-1.80	85.09	97.95	99.32
	2021/01	6.51	6.07	-2.60	80.86	96.84	99.08
	2024/01	5.34	5.33	1.22	86.15	97.09	99.12
부산	2015/01	5.08	4.85	-1.94	88.58	98.43	99.55
	2018/01	6.35	6.35	-0.38	80.50	95.00	98.94
	2021/01	9.10	9.04	-1.70	65.62	91.64	96.91
	2024/01	10.17	10.19	4.14	77.58	92.92	95.47
대구	2015/01	6.04	5.31	-3.37	82.62	97.71	99.44
	2018/01	4.83	4.54	-2.01	89.51	99.08	99.72
	2021/01	7.60	7.42	-2.08	72.67	95.15	98.74
	2024/01	5.64	5.65	-0.04	84.46	97.65	99.59
인천	2015/01	4.56	4.34	-1.68	91.15	99.01	99.77
	2018/01	4.30	4.19	-1.26	92.75	98.97	99.76
	2021/01	6.14	5.69	-2.89	81.38	97.65	99.59
	2024/01	5.24	5.23	-0.44	86.82	98.10	99.64
광주	2015/01	6.41	5.95	-3.02	79.35	97.70	99.48
	2018/01	5.91	5.78	-1.16	85.40	97.29	98.41
	2021/01	7.04	7.01	-0.78	77.66	95.85	98.80
	2024/01	6.46	6.44	-1.30	79.63	96.56	99.63
대전	2015/01	4.76	4.53	-1.75	90.13	98.07	99.78
	2018/01	4.41	4.39	-0.74	91.38	99.19	99.94
	2021/01	7.01	6.48	-3.47	75.20	96.95	99.47
	2024/01	5.39	5.34	-1.17	85.70	98.42	99.61
울산	2015/01	6.02	5.54	-2.32	84.46	96.87	98.77
	2018/01	5.46	5.47	-0.46	86.30	98.72	99.74
	2021/01	11.34	11.06	-0.56	62.66	90.14	97.33
	2024/01	7.15	7.06	-1.24	78.91	95.09	97.91
세종	2015/01	5.62	5.49	-1.09	87.77	97.84	99.28
	2018/01	6.48	6.44	-2.19	80.71	94.64	98.93
	2021/01	8.28	8.12	-2.93	73.51	91.65	96.90
	2024/01	5.10	5.10	0.07	88.48	97.94	99.18
경기	2015/01	4.32	4.17	-1.62	92.02	98.89	99.78
	2018/01	5.02	4.87	-1.73	88.62	98.47	99.67
	2021/01	7.67	7.27	-3.35	72.03	94.89	99.04
	2024/01	4.96	4.96	-0.07	89.13	98.58	99.68

〈표 5〉 계속

지역	시점	MAPE	COD	MPE	〈10%〉	〈20%〉	30%
강원	2015/01	6.80	6.30	-3.18	78.57	95.54	98.77
	2018/01	6.41	6.29	-1.10	82.11	96.57	98.10
	2021/01	7.73	7.60	-1.96	72.75	94.60	98.08
	2024/01	6.94	6.95	-0.61	78.33	96.06	98.58
충북	2015/01	7.53	7.41	0.67	82.02	92.60	94.56
	2018/01	6.80	6.79	0.22	79.74	94.29	98.55
	2021/01	9.01	8.65	-2.55	66.58	93.00	97.01
	2024/01	5.95	5.88	-0.70	83.96	97.39	99.06
충남	2015/01	5.60	5.42	-1.72	86.03	96.80	99.26
	2018/01	6.35	6.19	-1.55	80.16	97.44	99.33
	2021/01	7.60	7.23	-2.94	73.91	94.68	98.41
	2024/01	6.69	6.68	-0.76	79.92	96.77	98.87
전북	2015/01	6.52	6.31	-1.92	79.59	95.98	98.90
	2018/01	6.19	6.00	-1.46	81.99	97.15	99.21
	2021/01	8.33	8.31	-0.82	69.28	93.31	98.09
	2024/01	7.29	7.31	-0.57	77.11	93.98	97.07
전남	2015/01	7.40	7.19	-1.76	75.05	93.44	97.57
	2018/01	7.59	7.51	-0.96	74.90	92.72	98.48
	2021/01	8.14	8.00	-2.22	70.88	93.01	98.53
	2024/01	8.34	8.37	0.17	73.30	93.13	97.00
경북	2015/01	7.48	7.32	-1.87	76.17	93.60	97.75
	2018/01	7.98	7.95	-1.16	72.72	93.59	98.09
	2021/01	9.12	8.78	-2.61	66.98	90.94	97.19
	2024/01	8.26	8.26	-0.83	71.86	92.72	97.37
경남	2015/01	5.15	4.91	-2.02	87.62	98.15	99.57
	2018/01	8.06	8.08	1.54	72.43	91.62	96.71
	2021/01	7.86	7.61	-1.95	73.06	94.22	98.29
	2024/01	6.62	6.60	-0.94	79.39	96.42	98.87
제주	2015/01	9.88	8.66	-5.51	53.73	91.07	98.36
	2018/01	8.35	8.30	-0.88	70.85	92.38	97.31
	2021/01	9.74	9.16	-3.97	60.05	88.94	98.24
	2024/01	6.83	6.76	-1.77	76.73	96.04	99.50

AVM, automated valuation model; MAPE, mean absolute percentage error; COD, coefficient of dispersion; MPE, mean percentage error.

셋째, 평균백분오차(MPE)의 절대값을 17개 시·도 평균으로 보면 2018년 1월(1.21%)과 2024년 1월(0.94%)에는 1% 안팎으로 낮게 유지되었으나, 2015년 1월(2.16%)과 2021년 1월(2.32%)에는 다소 상승하였다. 이는 시점 기반 평가에서는 무작위 교차검증에 비해 편향성이 다소 확대됨을 보여준다. 개별 지역·시점 단위에서도 $\pm 3\%$ 를 초과하는 편향이 8건(68건 중) 관측되며, 특히 제주 2015년 1월(-5.51%)과 부산 2024년 1월(+4.14%)에서 가장 큰 편향이 발생하였다. 그럼에도 $\pm 3\%$ 초과 사례의 대부분은 위에서 언급한 급격한 시장 국면 변화(2020~2021년 급등기, 2024년 부산 시장의 부분적 반전)에 집중되어 있으며, 여타 시점·지역에서는 편향이 $\pm 3\%$ 이내에서 관리된다.

넷째, 허용 오차 $\pm 30\%$ 이내의 적중률은 충북 2015년 1월(94.56%)과 부산 2024년 1월(95.47%) 두 조합을 제외한 모든 시점·지역 조합에서 96% 이상을 유지하며, 78%(53/68)에 해당하는 조합에서 98%를 상회한다. 이는 외삽 조건에서 일부 시점·지역의 평균 정확도가 다소 저하될지라도, 극단적 예측 오차의 발생 빈도 자체는 시점 기반 평가 조건에서도 제한적임을 보여준다(〈표 5〉).

종합하면, 시점 누설을 통제한 보다 엄격한 평가 조건에서도 본 모형은 시장의 다양한 국면에 걸쳐 일정 수준 이상의 정확성을 유지한다. 시점 기반 평가에서는 무작위 교차검증에 비해 편향성(MPE)과 평균 오차(MAPE)가 다소 확대되며, 이는 시장의 구조 변화 직후 시점(특히 2020~2021년 급등기 직후)에서 두드러진다. 그러나 이러한 성능 저하는 학습 시점 범위 밖으로의 외삽 조건에서 발생하는 자연스러운 현상이며, 본 연구가 IV.2절에서 채택한 평가 방식은 분석 목적에 부합한다는 판단을 보완적으로 뒷받침한다. 또한 $\pm 30\%$ 적중률 측면에서는 시점 기반 평가 조건에서도 압도적 다수의 조합이 96% 이상의 안정성을 유지함이 확인되어, 본 연구의 추정 결과가 시점 누설 가능성으로부터 비롯할 수 있는 낙관적 평가에 뚜렷하게 의존하지 않음을 시사한다. 이러한 결과를 바탕으로, 다음 장에서는 본 모형을 활용한 각 지역의 가격 수준 및 분포 특성을 비교·분석한다.

V. 분석 결과

1. 지역별 평균 가격의 추이

본 연구에서는 구축된 자동산정모형을 활용하여, 2012년 3월 1일부터 2024년 4월 1일까지 약 12년간 전국 및 17개 시·도 아파트 단지의 각 평형별 가격을 분기별로 추정하였으며, 이를 통해 전국 및 지역별 가중 평균(각 평형별 세대수를 기정한 기중평균)을 도출하였다. 이는 거래된 주택만의 가격 변화를 추적하는 전통적 지표와 달리, 동일하게 고정된 전수 표본의 가격을 각 시점마다 추정함으로써 시장 전체의 가격 수준을 직접 관측하는 접근이다. 이는 단순히 각 시점에 거래가 이루어진 주택의 가격이 아닌 모든 주택을 대상으로 한 실질적인 시장 전체의 평균가격을 도출한 데에 그 의미가 있다. 다만 이는 재고 주택의

특성 변화를 고려한 통계적 의미의 가격지수가 아니라 전수 추정에 기반한 가중평균값임을 미리 밝혀둔다.

먼저, <표 6>은 전국 및 지역별 평균 가격 수준을 보여준다. 2024년 1분기말 기준, 전국 17개 시도의 공급면적 기준 평균 아파트 가격은 약 470만 원/m²으로 나타난다. 서울이 약 1,121만 원/m²으로 가장 높은 평균가격을 기록했으며, 그 다음으로는 경기도(약 520만 원/m²)가 위치했다. 가장 평균 가격이 낮은 지역은 경북(약 181만 원/m²)이었으며 그 다음으로 낮은 지역은 전남(186만 원/m²)이었다. 전국 평균 아파트 가격은 2012년 1분기말 약 240만 원/m² 수준이었으나, 2024년 1분기말까지 전반적으로 상승했던 것으로 나타났다. 특히, 시기별 추이를 보면 2012년부터 2018년까지는 완만한 상승세를 보이다가 2018년 후반부터 2021년 상반기에 걸쳐 급등하고 2022년 이후 천천히 조정되는 형태를 보여준다.

<표 6> 아파트 평균가격(단위: 만 원/m²)

시점	전국	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2012/03	243	471	218	170	228	149	205	204	188	266	128	151	153	142	122	124	182	167
2012/06	240	464	216	171	228	151	202	206	190	262	127	153	155	141	123	125	178	167
2012/09	238	454	214	176	224	152	200	206	203	258	128	153	157	139	123	128	177	171
2012/12	237	453	213	177	224	152	201	204	217	256	128	153	158	138	122	132	176	170
2013/03	239	455	214	182	222	155	203	205	220	256	129	155	161	138	127	134	180	170
2013/06	241	459	214	188	224	157	203	205	224	258	128	154	163	137	125	137	179	172
2013/09	244	460	215	196	229	161	206	209	226	261	130	158	163	140	126	142	183	173
2013/12	247	465	218	205	231	163	208	212	238	264	131	159	166	140	125	145	185	177
2014/03	251	474	221	211	235	167	207	214	238	269	132	163	168	142	127	148	188	183
2014/06	252	474	224	215	232	167	208	218	238	270	133	166	169	141	127	150	190	190
2014/09	257	483	227	221	241	172	209	220	240	273	134	171	174	143	128	153	194	194
2014/12	261	490	230	230	245	178	208	226	241	278	136	175	176	143	128	157	197	203
2015/03	266	500	234	240	245	186	210	231	250	283	139	175	178	144	130	161	200	210
2015/06	273	514	240	253	253	192	211	236	261	290	142	178	179	146	131	164	205	225
2015/09	280	529	248	266	261	198	214	245	263	298	146	179	180	147	135	168	208	248
2015/12	283	540	254	257	264	198	213	252	269	301	150	176	181	147	137	166	209	266
2016/03	284	540	258	253	265	201	213	253	282	301	151	176	180	148	139	163	209	277
2016/06	287	553	261	249	268	200	213	254	287	306	154	175	179	149	140	160	208	287
2016/09	296	581	272	250	275	202	216	255	280	315	160	176	179	152	144	161	210	300
2016/12	301	589	286	253	283	202	220	256	291	320	165	174	179	154	147	161	210	307
2017/03	303	597	290	253	281	204	221	255	316	320	167	171	180	155	148	159	209	311
2017/06	308	620	293	256	281	207	223	254	324	326	171	174	179	156	151	158	206	314
2017/09	316	646	295	266	287	210	226	252	335	334	174	174	179	158	156	159	204	315
2017/12	320	674	293	270	289	211	227	249	321	335	173	174	178	159	158	159	202	316
2018/03	331	729	292	275	287	216	229	245	346	344	173	173	176	158	159	157	197	319
2018/06	333	746	287	280	286	220	230	237	348	347	171	173	174	158	162	154	195	314

〈표 6〉 계속

시점	전국	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2018/09	349	822	288	284	289	232	233	233	354	360	171	167	173	157	164	154	193	315
2018/12	350	816	286	289	288	242	243	230	356	365	169	166	175	157	165	151	191	314
2019/03	346	802	286	290	288	241	247	226	365	362	166	163	174	156	165	149	188	309
2019/06	350	823	285	290	289	240	254	226	377	363	163	161	173	154	166	149	186	310
2019/09	360	870	285	294	292	242	262	226	374	373	164	161	174	156	167	149	185	306
2019/12	376	926	299	300	299	242	279	234	393	389	164	163	178	157	169	150	189	308
2020/03	394	963	307	306	319	246	298	243	453	418	166	169	182	160	171	153	193	309
2020/06	400	966	310	308	337	249	311	246	473	428	169	177	184	161	174	154	194	310
2020/09	429	1,063	326	319	344	251	332	258	599	460	174	183	192	166	178	157	199	314
2020/12	452	1,098	372	353	358	263	348	291	614	486	179	191	200	176	183	166	215	323
2021/03	487	1,169	395	375	387	275	374	306	631	539	192	205	212	183	188	175	227	343
2021/06	505	1,207	408	378	423	281	388	307	617	565	198	215	218	189	189	180	232	366
2021/09	538	1,272	431	381	472	298	408	319	622	614	212	227	230	200	194	187	244	397
2021/12	544	1,280	433	376	485	317	411	320	597	618	224	236	238	208	199	192	252	404
2022/03	534	1,260	427	363	470	318	394	321	572	603	227	236	237	210	198	192	254	406
2022/06	531	1,255	425	355	459	320	388	317	562	598	233	236	237	212	198	192	255	413
2022/09	503	1,191	407	339	432	307	370	308	515	557	232	230	232	211	195	189	251	409
2022/12	470	1,105	382	325	395	293	347	294	487	516	221	220	221	201	187	183	237	395
2023/03	461	1,090	370	310	384	287	339	288	483	507	220	217	216	196	185	181	233	391
2023/06	468	1,110	372	310	388	287	346	290	498	517	220	218	216	196	186	181	234	394
2023/09	477	1,139	373	314	395	290	354	294	504	529	223	222	219	199	188	183	236	392
2023/12	472	1,125	369	314	393	287	353	293	494	521	222	223	218	197	186	182	234	390
2024/03	470	1,121	365	309	393	282	351	293	483	520	221	222	216	197	186	181	231	387

주: 소수점 이하는 반올림되었다.

결과에 의하면, 해당 기간 동안 각 지역의 가격 변화에 상당한 차이가 있었음을 알 수 있다. 〈표 7〉은 〈표 6〉의 결과를 토대로 가격의 변화율을, 〈표 8〉은 최초 시점으로부터의 누적가격 변화율을 보여준다. 특히, 수도권 지역과 세종시에서는 2018년에서 2021년 사이의 가격 상승폭이 매우 컸다는 것을 알 수 있으며, 2022년 이후에는 대부분의 지역에서 조정 국면에 접어든 것으로 관찰된다. 2012년과 2024년 사이의 누적 상승률 상위 지역은 각각 서울(138%), 경기(95%), 세종(156%), 제주(131%)이며, 이들 지역은 공통적으로 인구 증가, 또는 수도권 연계성이 높은 특성을 가진다. 특히, 서울의 단위 면적 당 아파트 평균가격은 2012년 초 기준으로 470만 원이었으나, 2024년 초에는 약 1,121만 원에 달해 세종시를 제외하고는 해당 기간 가장 높은 누적성장률을 보여준다. 반면 하위 지역은 경남(27%), 충남(40%), 울산(43%), 경북(46%), 전남(52%)으로 나타났다.

〈표 7〉 아파트 평균가격 변화율(단위: %)

시점	전국	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2012/03	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2012/06	-0.91	-1.44	-0.92	0.82	-0.08	1.70	-1.58	0.98	1.20	-1.64	-0.64	1.71	0.96	-0.27	0.73	1.57	-1.83	-0.08
2012/09	-1.09	-2.15	-0.83	2.69	-1.66	0.54	-1.10	-0.17	6.62	-1.67	0.58	-0.24	1.33	-1.43	0.03	2.16	-0.65	2.19
2012/12	-0.25	-0.25	-0.64	0.63	0.05	-0.01	0.53	-0.88	6.91	-0.80	0.19	-0.02	0.51	-0.91	-0.16	2.61	-0.87	-0.67
2013/03	0.68	0.49	0.42	2.96	-0.91	1.80	1.27	0.54	1.55	0.11	0.99	1.12	2.07	0.18	3.66	1.85	2.18	0.03
2013/06	0.81	0.88	0.10	3.44	1.03	1.32	-0.10	0.11	1.61	0.91	-0.42	-0.17	1.03	-0.58	-1.39	2.48	-0.08	1.21
2013/09	1.17	0.23	0.68	4.08	1.93	2.82	1.33	1.95	0.91	1.08	1.29	2.32	0.23	1.59	0.98	3.51	2.03	0.93
2013/12	1.28	1.09	1.32	4.40	1.13	0.96	0.85	1.03	5.44	1.11	0.45	1.05	1.99	0.51	-0.89	1.83	1.00	2.00
2014/03	1.77	1.96	1.39	2.87	1.79	2.78	-0.13	1.12	-0.34	1.77	1.25	2.04	1.30	0.87	1.57	2.36	1.63	3.32
2014/06	0.47	-0.01	1.18	2.09	-1.58	-0.25	0.16	1.96	-0.01	0.62	0.30	2.17	0.61	-0.50	0.11	1.37	1.18	4.18
2014/09	1.80	1.93	1.53	3.03	4.18	2.79	0.86	0.96	1.02	1.15	1.24	2.60	2.62	1.46	0.06	1.89	2.05	1.93
2014/12	1.62	1.48	1.09	4.07	1.36	3.92	-0.70	2.83	0.28	1.53	0.92	2.62	1.38	-0.20	0.00	2.85	1.64	4.75
2015/03	1.84	1.92	2.00	3.94	0.07	4.17	0.90	2.10	3.80	1.81	2.26	-0.06	0.70	0.69	2.30	2.29	1.47	3.49
2015/06	2.66	2.79	2.53	5.59	3.24	3.20	0.46	2.18	4.32	2.70	2.49	1.46	0.57	1.64	0.11	1.72	2.22	7.04
2015/09	2.82	3.07	3.32	5.01	3.38	3.28	1.38	3.72	0.95	2.56	2.69	0.70	0.96	0.80	3.52	2.76	1.70	10.13
2015/12	1.01	1.90	2.19	-3.10	1.05	0.19	-0.46	2.69	2.34	1.03	2.94	-1.59	0.28	-0.18	1.59	-1.27	0.34	7.32
2016/03	0.25	0.10	1.57	-1.55	0.59	1.34	-0.07	0.49	4.75	0.21	0.53	0.05	-0.45	0.91	1.24	-2.08	0.14	4.05
2016/06	1.19	2.40	1.17	-1.64	1.07	-0.69	0.39	0.42	1.85	1.52	1.78	-0.45	-0.40	0.53	0.80	-1.48	-0.48	3.52
2016/09	3.02	5.11	4.45	0.35	2.40	1.21	1.39	0.57	-2.67	2.95	4.01	0.59	-0.32	1.97	2.48	0.44	0.78	4.50

〈표 7〉 계속

시점	전국	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2016/12	1.60	1.41	5.16	1.23	2.87	-0.02	1.52	0.41	3.97	1.46	3.32	-1.14	0.27	1.62	2.46	-0.12	0.11	2.37
2017/03	0.60	1.35	1.39	0.12	-0.52	1.08	0.52	-0.39	8.59	0.13	1.20	-1.62	0.25	0.54	0.52	-0.81	-0.28	1.45
2017/06	1.83	3.77	1.09	0.91	-0.02	1.27	1.20	-0.73	2.79	1.88	2.22	1.26	-0.46	0.61	2.17	-1.13	-1.61	0.79
2017/09	2.44	4.26	0.56	3.82	2.14	1.51	0.98	-0.69	3.38	2.32	1.76	0.36	0.30	1.07	2.99	0.78	-0.75	0.57
2017/12	1.41	4.36	-0.81	1.60	0.50	0.59	0.55	-1.23	-4.19	0.50	-0.39	-0.32	-0.57	0.58	1.76	0.08	-1.36	0.11
2018/03	3.29	8.11	-0.28	2.06	-0.47	2.09	0.86	-1.65	7.80	2.63	0.14	-0.44	-1.35	-0.15	0.67	-1.48	-2.11	0.96
2018/06	0.84	2.29	-1.58	1.55	-0.37	2.04	0.60	-3.04	0.61	0.89	-1.25	-0.14	-0.82	-0.35	1.86	-1.93	-1.17	-1.42
2018/09	4.67	10.23	0.36	1.74	1.14	5.69	1.22	-1.65	1.69	3.84	-0.08	-3.23	-0.69	-0.61	0.80	-0.08	-0.90	0.19
2018/12	0.18	-0.75	-0.75	1.50	-0.32	4.18	4.33	-1.33	0.60	1.14	-1.38	-0.46	0.97	0.10	0.70	-1.47	-1.13	-0.08
2019/03	-0.99	-1.73	-0.04	0.30	-0.12	-0.53	1.58	-2.00	2.29	-0.82	-1.94	-2.13	-0.62	-0.91	0.38	-1.77	-1.47	-1.66
2019/06	1.02	2.60	-0.42	0.06	0.22	-0.37	2.92	0.11	3.34	0.46	-1.71	-0.87	-0.20	-0.85	0.40	0.23	-0.95	0.16
2019/09	3.03	5.79	0.21	1.56	1.22	0.69	3.23	0.26	-0.63	2.72	0.68	0.03	0.51	0.85	0.48	0.21	-0.56	-1.15
2019/12	4.44	6.45	4.61	2.06	2.44	0.31	6.40	3.50	4.83	4.36	0.02	1.00	2.28	0.63	1.32	0.66	2.03	0.55
2020/03	4.75	3.95	2.76	1.75	6.51	1.64	6.93	3.62	15.31	7.35	1.27	3.57	2.10	2.23	0.86	1.53	1.96	0.22
2020/06	1.57	0.27	0.99	0.68	5.63	1.03	4.25	1.44	4.39	2.48	2.02	4.92	1.00	0.75	2.15	1.01	0.47	0.34
2020/09	7.26	10.08	5.32	3.72	2.21	0.87	6.63	4.63	26.70	7.36	2.95	3.28	4.55	2.85	2.25	1.66	2.87	1.45
2020/12	5.35	3.32	13.99	10.52	3.99	4.82	4.87	12.80	2.61	5.68	2.59	4.35	4.21	5.80	2.90	5.76	7.85	2.86
2021/03	7.67	6.48	6.26	6.32	8.22	4.60	7.53	5.17	2.66	10.95	7.45	7.26	5.60	4.34	2.53	5.72	5.59	6.12
2021/06	3.70	3.17	3.11	0.79	9.19	2.05	3.69	0.50	-2.17	4.82	3.16	4.82	3.17	3.22	0.40	2.73	2.00	6.72

〈표 7〉 계속

시점	전국	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2021/09	6.48	5.40	5.63	0.69	11.67	6.25	5.03	3.84	0.77	8.68	7.12	5.98	5.36	5.55	2.73	4.04	5.53	8.40
2021/12	1.07	0.67	0.54	-1.17	2.55	6.36	0.78	0.46	-4.01	0.64	5.30	3.89	3.34	4.19	2.43	2.63	3.32	1.88
2022/03	-1.76	-1.61	-1.35	-3.41	-3.00	0.22	-3.99	0.15	-4.09	-2.39	1.67	-0.29	-0.48	1.09	-0.26	-0.06	0.76	0.49
2022/06	-0.65	-0.39	-0.38	-2.22	-2.36	0.56	-1.55	-1.18	-1.85	-0.95	2.64	0.18	0.29	0.93	0.03	-0.15	0.37	1.61
2022/09	-5.11	-5.09	-4.30	-4.70	-5.86	-4.00	-4.58	-2.94	-8.36	-6.72	-0.59	-2.62	-2.17	-0.62	-1.58	-1.33	-1.88	-0.86
2022/12	-6.66	-7.17	-6.10	-4.14	-8.48	-4.64	-6.32	-4.67	-5.35	-7.41	-4.63	-4.12	-4.98	-4.81	-4.09	-3.19	-5.28	-3.46
2023/03	-1.87	-1.41	-3.21	-4.56	-2.99	-1.94	-2.44	-1.89	-0.78	-1.74	-0.48	-1.59	-2.17	-2.55	-0.85	-1.42	-1.75	-0.90
2023/06	1.44	1.85	0.58	0.17	1.17	-0.12	2.20	0.71	3.09	1.85	-0.18	0.40	0.25	-0.02	0.32	0.54	0.46	0.70
2023/09	2.06	2.65	0.21	1.26	1.91	1.24	2.43	1.48	1.18	2.35	1.30	1.80	1.21	1.63	0.97	0.92	0.82	-0.62
2023/12	-1.16	-1.28	-1.03	-0.08	-0.73	-1.32	-0.47	-0.34	-2.03	-1.53	-0.15	0.39	-0.61	-1.04	-0.82	-0.87	-0.85	-0.37
2024/03	-0.43	-0.33	-1.18	-1.49	0.04	-1.59	-0.44	-0.07	-2.33	-0.15	-0.56	-0.03	-0.90	0.05	-0.13	-0.30	-1.19	-0.87

주: 1) 직전 시점으로부터의 변화율(분기 변화율)을 의미한다.

2) 소수점 2자리 이하는 반올림되었다.

〈표 8〉 아파트 평균가격 누적변화율(단위: %)

시점	전국	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2012/03	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2012/06	-0.9	-1.4	-0.9	0.8	-0.1	1.7	-1.6	1.0	1.2	-1.6	-0.6	1.7	1.0	-0.3	0.7	1.6	-1.8	-0.1
2012/09	-2.0	-3.6	-1.7	3.5	-1.7	2.2	-2.7	0.8	7.9	-3.3	-0.1	1.5	2.3	-1.7	0.8	3.8	-2.5	2.1
2012/12	-2.2	-3.8	-2.4	4.2	-1.7	2.2	-2.1	-0.1	15.4	-4.1	0.1	1.4	2.8	-2.6	0.6	6.5	-3.3	1.4
2013/03	-1.6	-3.3	-2.0	7.3	-2.6	4.1	-0.9	0.5	17.2	-3.9	1.1	2.6	4.9	-2.4	4.3	8.4	-1.2	1.4

〈표 8〉 계속

시점	전국	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2013/06	-0.8	-2.5	-1.9	10.9	-1.6	5.4	-1.0	0.6	19.0	-3.1	0.7	2.4	6.0	-3.0	2.8	11.1	-1.3	2.7
2013/09	0.4	-2.3	-1.2	15.5	0.3	8.4	0.3	2.5	20.1	-2.0	2.0	4.8	6.3	-1.4	3.8	15.0	0.7	3.6
2013/12	1.7	-1.2	0.1	20.6	1.4	9.5	1.2	3.6	26.7	-0.9	2.4	5.9	8.4	-0.9	2.9	17.1	1.7	5.7
2014/03	3.5	0.8	1.5	24.0	3.3	12.5	1.0	4.7	26.2	0.8	3.7	8.0	9.8	-0.1	4.5	19.9	3.4	9.2
2014/06	4.0	0.7	2.7	26.6	1.6	12.2	1.2	6.8	26.2	1.4	4.0	10.4	10.5	-0.6	4.7	21.5	4.6	13.8
2014/09	5.8	2.7	4.3	30.5	5.9	15.4	2.1	7.8	27.5	2.6	5.3	13.3	13.4	0.9	4.7	23.8	6.7	16.0
2014/12	7.5	4.2	5.4	35.8	7.3	19.9	1.4	10.9	27.9	4.2	6.3	16.2	14.9	0.7	4.7	27.4	8.5	21.5
2015/03	9.5	6.2	7.5	41.1	7.4	24.9	2.3	13.2	32.7	6.1	8.7	16.2	15.7	1.4	7.1	30.3	10.1	25.7
2015/06	12.4	9.2	10.2	49.0	10.9	28.9	2.7	15.7	38.5	8.9	11.4	17.9	16.4	3.0	7.2	32.5	12.5	34.6
2015/09	15.6	12.5	13.9	56.5	14.6	33.1	4.2	20.0	39.8	11.7	14.4	18.7	17.5	3.8	11.0	36.2	14.5	48.2
2015/12	16.8	14.7	16.4	51.6	15.8	33.4	3.7	23.2	43.0	12.9	17.8	16.8	17.8	3.7	12.8	34.5	14.8	59.1
2016/03	17.1	14.8	18.2	49.3	16.5	35.1	3.6	23.8	49.8	13.1	18.4	16.8	17.3	4.6	14.2	31.7	15.0	65.5
2016/06	18.5	17.5	19.6	46.8	17.8	34.2	4.0	24.3	52.6	14.8	20.5	16.3	16.9	5.2	15.1	29.7	14.5	71.3
2016/09	22.0	23.5	24.9	47.3	20.6	35.8	5.4	25.0	48.5	18.2	25.3	17.0	16.5	7.2	17.9	30.3	15.3	79.1
2016/12	24.0	25.3	31.4	49.2	24.0	35.8	7.0	25.6	54.4	19.9	29.5	15.7	16.8	9.0	20.8	30.1	15.5	83.3
2017/03	24.7	27.0	33.2	49.3	23.4	37.3	7.6	25.1	67.7	20.1	31.0	13.8	17.1	9.6	21.5	29.1	15.1	86.0
2017/06	27.0	31.7	34.6	50.7	23.4	39.0	8.9	24.2	72.4	22.3	33.9	15.2	16.5	10.2	24.1	27.6	13.3	87.4
2017/09	30.1	37.4	35.4	56.4	26.0	41.1	10.0	23.3	78.2	25.2	36.3	15.6	16.9	11.4	27.8	28.6	12.4	88.5
2017/12	31.9	43.3	34.3	59.0	26.6	42.0	10.6	21.8	70.7	25.8	35.8	15.3	16.2	12.1	30.1	28.7	10.9	88.7

〈표 8〉 계속

시점	전국	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2018/03	36.3	55.0	33.9	62.2	26.0	44.9	11.5	19.8	84.0	29.1	36.0	14.8	14.7	11.9	30.9	26.8	8.6	90.5
2018/06	37.4	58.5	31.8	64.7	25.6	47.9	12.2	16.1	85.2	30.3	34.3	14.6	13.7	11.5	33.4	24.4	7.3	87.8
2018/09	43.9	74.7	32.3	67.6	27.0	56.3	13.6	14.2	88.3	35.3	34.2	10.9	12.9	10.8	34.4	24.3	6.3	88.2
2018/12	44.1	73.4	31.3	70.1	26.6	62.8	18.5	12.7	89.4	36.8	32.3	10.4	14.0	10.9	35.4	22.4	5.1	88.0
2019/03	42.7	70.4	31.2	70.6	26.4	62.0	20.3	10.4	93.8	35.7	29.7	8.0	13.3	9.9	35.9	20.3	3.6	84.9
2019/06	44.1	74.8	30.7	70.7	26.7	61.4	23.8	10.6	100.2	36.3	27.5	7.1	13.1	9.0	36.4	20.6	2.6	85.2
2019/09	48.5	85.0	31.0	73.4	28.3	62.5	27.8	10.8	99.0	40.0	28.4	7.1	13.7	9.9	37.1	20.8	2.0	83.0
2019/12	55.1	96.9	37.0	77.0	31.4	63.0	36.0	14.7	108.6	46.1	28.4	8.2	16.3	10.6	38.9	21.6	4.1	84.0
2020/03	62.5	104.7	40.8	80.1	40.0	65.7	45.5	18.9	140.5	56.9	30.0	12.1	18.7	13.1	40.1	23.5	6.1	84.4
2020/06	65.0	105.2	42.2	81.3	47.8	67.4	51.6	20.6	151.1	60.8	32.7	17.6	19.9	13.9	43.1	24.7	6.6	85.0
2020/09	77.0	125.9	49.8	88.0	51.1	68.8	61.7	26.2	218.1	72.6	36.6	21.4	25.4	17.2	46.3	26.8	9.7	87.7
2020/12	86.5	133.4	70.7	107.8	57.1	77.0	69.6	42.3	226.4	82.4	40.1	26.7	30.7	24.0	50.6	34.1	18.3	93.1
2021/03	100.8	148.5	81.4	121.0	70.1	85.1	82.3	49.7	235.1	102.4	50.6	35.9	38.0	29.3	54.4	41.8	24.9	104.9
2021/06	108.2	155.4	87.0	122.7	85.7	88.9	89.1	50.4	227.8	112.1	55.3	42.5	42.3	33.5	55.0	45.6	27.4	118.7
2021/09	121.7	170.3	97.6	124.2	107.4	100.7	98.6	56.2	230.4	130.5	66.4	51.0	50.0	40.9	59.2	51.5	34.5	137.1
2021/12	124.1	172.1	98.6	121.6	112.6	113.4	100.1	56.9	217.1	132.0	75.2	56.9	55.0	46.8	63.1	55.5	38.9	141.5
2022/03	120.1	167.7	95.9	114.1	106.3	113.9	92.1	57.2	204.1	126.5	78.1	56.4	54.2	48.4	62.7	55.4	40.0	142.7
2022/06	118.7	166.6	95.2	109.3	101.4	115.1	89.1	55.3	198.5	124.3	82.8	56.7	54.7	49.8	62.7	55.2	40.5	146.6
2022/09	107.5	153.1	86.8	99.5	89.6	106.5	80.5	50.7	173.6	109.2	81.7	52.6	51.3	48.9	60.2	53.1	37.9	144.5

〈표 8〉 계속

시점	전국	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2022/12	93.7	134.9	75.4	91.2	73.5	96.9	69.1	43.7	158.9	93.7	73.3	46.3	43.8	41.7	53.6	48.2	30.6	136.0
2023/03	90.0	131.6	69.8	82.5	68.3	93.1	65.0	41.0	156.9	90.4	72.5	44.0	40.7	38.1	52.3	46.1	28.3	133.9
2023/06	92.8	135.9	70.8	82.8	70.3	92.9	68.6	42.0	164.8	93.9	72.2	44.6	41.0	38.1	52.8	46.9	28.9	135.6
2023/09	96.7	142.1	71.1	85.1	73.6	95.3	72.7	44.1	168.0	98.5	74.4	47.2	42.7	40.3	54.3	48.3	30.0	134.1
2023/12	94.5	139.0	69.4	84.9	72.3	92.7	71.9	43.6	162.5	95.4	74.2	47.7	41.9	38.9	53.0	47.0	28.9	133.2
2024/03	93.6	138.2	67.4	82.2	72.3	89.6	71.1	43.5	156.4	95.1	73.2	47.7	40.6	39.0	52.8	46.5	27.3	131.2

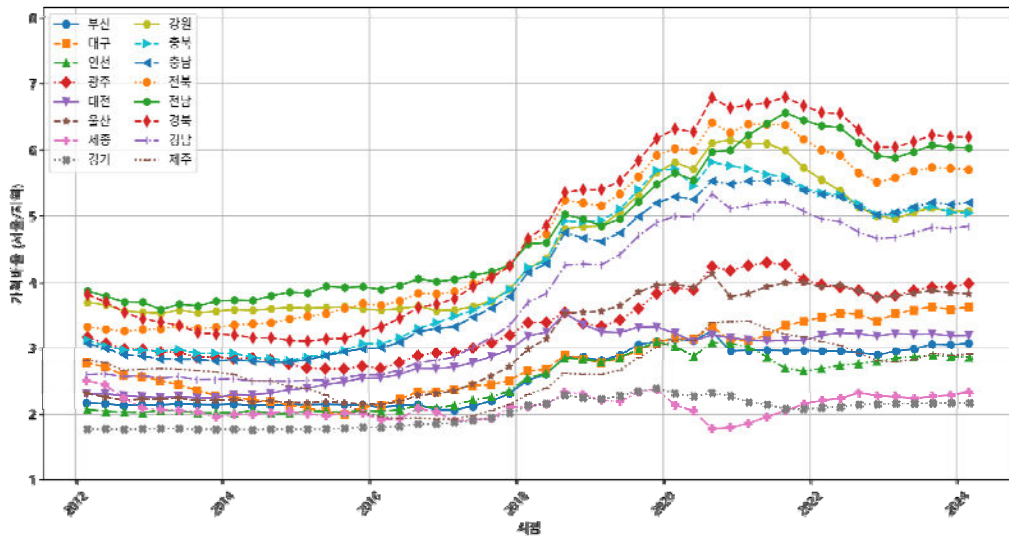
주: 1) 누적변화율은 시계열의 최초시점(2012.3.1.)으로부터 각 시점까지의 누적된 변화율을 의미한다.

2) 소수점 2자리 이하는 반올림되었다.

결과적으로, 이는 서울과 기타 지역 간의 절대적 가격 수준 격차가 매우 클 뿐 아니라 격차 자체가 지속적으로 확대되어왔음을 의미한다. 〈그림 1〉은 서울과 지역의 가격 차이(=서울가격/지역가격)가 어떻게 변화하는지를 시계열 그래프로 제시하고 있다. 그림을 보면, 2012년 초 기준으로 가장 높은 평균가격을 기록한 지역(서울)과 두 번째로 높은 지역(경기)의 배율은 1.76배 수준이었고, 가장 낮은 지역(전남)과의 배율은 3.86배 수준이었다. 그러나 2024년 3월에는 서울과 두 번째로 높은 지역(경기)와의 배율은 2.15배, 가장 낮은 지역(경북)과의 배율은 6.19배에 달하게 되었다. 이는 수도권, 특히 서울과 지역 간 가격의 격차 자체가 10여년간 큰 폭으로 상승했음을 직접적으로 보여준다.

본 연구의 추정 내용은 기존 관련 연구의 관찰 내용과도 부합한다. 예를 들어 허원제 외(2020)는 수도권과 비수도권 지역의 주택 가격 격차에 대한 분석에서, 2012년에서 2018년 이전에는 아파트 연평균 가격이 다소간 상승하는 가운데에서도 지역간 격차가 어느정도 유지되었으나, 2018년에 그 격차의 폭이 심화되기 시작했다는 것을 관측하였다. 본 연구의 추정치는 그 이후에도 격차의 심화가 3~4년간(2018년에서 2021년) 지속되었다는 것을 관측할 수 있도록 해준다.

다음으로, 서울 내 지역의 가격 수준을 살펴보자. 이는 〈표 9〉과 〈표 10〉에 제시되어 있다. 2024년 3월 기준으로 서울에서 가장 높은 평균 가격을 기록한 지역은 강남구로, 평균 1,634만 원/㎡에 달하였다. 이는 서울 전체 평균인 1,121만 원/㎡을 크게 상회하며, 가장 낮은 평균가격을 기록한 도봉구(약 524만 원/㎡)와 비교하면 3.12배에 달하는 수준이다. 강남구 다음으로는 용산구(1,454만 원/㎡), 서초구(1,391만 원/㎡), 송파구(1,053만 원/㎡) 등의 가격 수준이 높게 나타났다.



주: 가격배율은 서울의 단위 면적(1m²) 당 아파트 가격을 지역의 단위 면적(1m²) 당 아파트 가격으로 나눈 값을 의미한다.

〈그림 1〉 지역 대비 서울 가격배율의 추이

〈표 9〉 서울 평균 아파트 가격: 강남구~동작구(단위: 만 원/m²)

시점	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구	동대문구	동작구
2012/03	704	386	299	329	380	441	309	294	327	287	333	418
2012/06	693	378	302	321	381	438	308	298	320	283	334	414
2012/09	674	376	294	319	372	431	304	297	316	277	327	408
2012/12	670	378	295	321	375	431	301	299	311	276	329	404
2013/03	689	380	291	323	378	428	305	301	312	278	327	402
2013/06	672	381	302	325	377	433	306	298	312	277	326	405
2013/09	677	384	315	324	376	438	307	304	316	281	330	399
2013/12	689	386	307	328	382	438	311	308	316	284	336	407
2014/03	698	393	323	332	385	442	309	309	320	285	340	410
2014/06	701	390	324	333	387	448	312	305	319	285	340	412
2014/09	718	400	329	336	389	443	311	315	324	291	343	415
2014/12	725	404	329	340	392	445	317	312	330	288	349	419
2015/03	739	410	327	351	396	453	316	317	333	292	350	423
2015/06	762	421	329	364	405	458	323	319	341	297	364	437
2015/09	782	429	337	377	413	466	331	329	349	302	376	447
2015/12	802	439	344	383	423	469	333	342	353	314	379	450

〈표 9〉 계속

시점	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구	동대문구	동작구
2016/03	801	438	343	387	423	478	334	336	351	311	390	464
2016/06	822	440	346	398	433	475	338	346	354	316	389	467
2016/09	866	454	358	413	443	492	345	344	367	328	403	484
2016/12	876	460	361	420	449	503	353	352	372	329	411	494
2017/03	885	460	361	423	445	505	354	356	376	330	414	492
2017/06	915	479	362	434	456	511	357	369	384	333	425	512
2017/09	950	512	370	449	462	527	364	378	398	343	439	529
2017/12	996	518	378	459	473	539	370	381	403	349	442	545
2018/03	1,071	543	384	473	481	573	374	385	408	354	465	582
2018/06	1,088	557	415	485	488	588	383	397	413	359	485	599
2018/09	1,172	593	432	526	515	615	405	418	447	377	510	650
2018/12	1,154	592	440	529	518	617	415	422	463	386	524	647
2019/03	1,129	586	436	522	515	617	412	425	459	382	520	630
2019/06	1,175	598	426	525	529	630	418	430	470	387	529	653
2019/09	1,249	624	442	543	542	661	432	442	487	394	547	685
2019/12	1,317	651	452	562	562	698	447	462	513	407	565	735
2020/03	1,320	675	471	581	581	722	458	482	539	420	596	750
2020/06	1,325	683	475	585	589	722	472	494	556	425	605	746
2020/09	1,418	741	517	622	631	781	505	522	628	472	650	805
2020/12	1,441	780	540	648	641	808	533	536	656	492	667	839
2021/03	1,491	819	547	675	666	843	568	565	704	525	707	876
2021/06	1,556	842	559	689	678	870	598	586	738	553	725	899
2021/09	1,627	872	597	739	705	910	627	622	782	589	748	963
2021/12	1,668	882	607	745	708	907	634	621	777	585	746	975
2022/03	1,654	884	601	744	708	919	631	623	763	577	749	967
2022/06	1,664	870	614	736	709	921	635	620	770	582	754	988
2022/09	1,657	835	588	717	691	902	613	606	715	550	743	943
2022/12	1,571	779	577	672	664	841	590	576	680	530	699	883
2023/03	1,588	787	577	671	666	848	574	575	667	530	699	883
2023/06	1,615	789	580	680	670	867	577	572	668	532	700	890
2023/09	1,641	806	596	692	681	900	580	582	671	536	712	915
2023/12	1,636	798	585	690	669	872	573	575	656	526	706	911
2024/03	1,634	795	562	679	669	882	567	565	653	524	712	904

주: 소수점 이하는 반올림되었다.

〈표 10〉 서울 평균 아파트 가격: 마포구~중랑구(단위: 만 원/m²)

시점	마포구	서대문구	서초구	성동구	성북구	송파구	양천구	영등포구	용산구	은평구	종로구	중구	중랑구
2012/03	433	335	592	514	340	482	365	441	547	330	427	419	324
2012/06	428	329	590	516	339	478	360	438	555	329	419	411	319
2012/09	423	330	579	503	327	471	356	425	561	324	421	427	313
2012/12	425	327	578	502	328	471	355	425	564	324	416	421	318
2013/03	425	333	579	507	331	468	355	424	553	323	416	415	314
2013/06	432	330	586	508	333	475	360	431	552	329	415	421	320
2013/09	431	334	583	514	333	475	360	431	568	330	419	410	320
2013/12	440	333	591	516	338	477	362	434	570	332	420	430	316
2014/03	442	340	602	521	340	487	367	442	576	337	429	415	328
2014/06	449	348	609	524	340	485	366	438	574	338	430	422	326
2014/09	455	349	618	529	343	497	369	444	588	344	442	429	323
2014/12	456	356	633	530	354	502	375	447	591	347	458	438	330
2015/03	464	356	648	526	356	502	384	455	600	350	455	455	331
2015/06	479	368	659	546	367	511	392	457	623	357	463	452	339
2015/09	494	374	675	560	382	530	401	475	631	362	472	483	346
2015/12	505	383	691	571	390	536	408	478	643	372	481	485	348
2016/03	505	383	696	578	385	542	408	482	642	371	486	484	344
2016/06	512	391	710	595	395	549	412	494	656	380	496	492	352
2016/09	530	403	731	628	401	569	433	524	670	395	513	517	359
2016/12	546	419	743	647	411	577	439	535	690	402	523	521	363
2017/03	555	419	750	658	409	580	441	540	701	406	527	543	366
2017/06	572	434	789	680	414	608	453	562	722	414	529	547	368
2017/09	583	451	804	699	427	628	463	570	742	425	546	549	378
2017/12	595	460	834	729	441	657	476	598	785	429	549	569	387
2018/03	632	477	879	787	452	707	493	630	816	443	579	594	394
2018/06	652	500	898	823	470	710	503	649	848	452	588	593	401
2018/09	705	535	964	886	509	755	546	721	912	485	608	612	435
2018/12	703	545	954	892	528	750	550	718	908	491	623	617	440
2019/03	707	539	955	879	520	735	543	717	923	488	624	611	449
2019/06	717	548	975	902	523	761	551	734	935	496	622	622	455
2019/09	754	574	1,033	986	550	805	579	780	971	505	650	649	471
2019/12	785	603	1,079	1,042	578	849	620	820	1,040	520	662	681	483
2020/03	813	617	1,093	1,085	611	877	646	848	1,081	540	669	700	508
2020/06	819	628	1,095	1,096	627	887	648	856	1,105	550	697	715	521

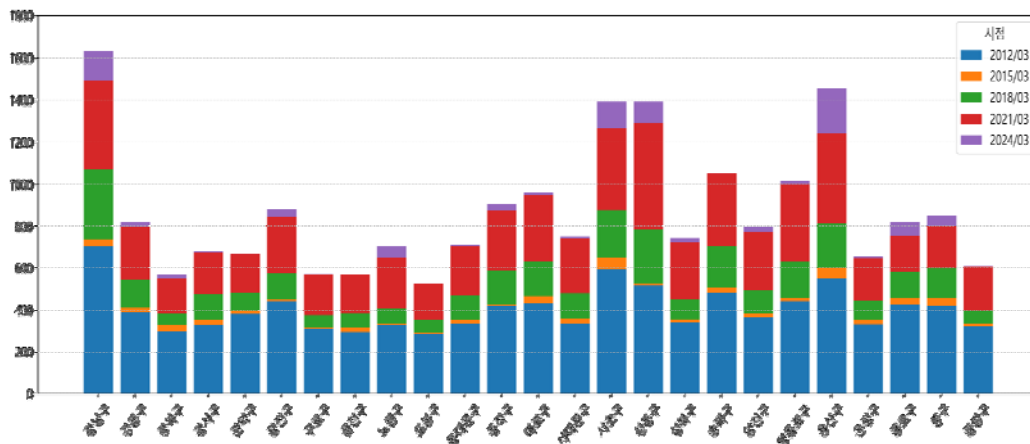
〈표 10〉 계속

시점	마포구	서대문구	서초구	성동구	성북구	송파구	양천구	영등포구	용산구	은평구	종로구	중구	중랑구
2020/09	876	677	1,183	1,179	684	962	704	923	1,166	587	713	752	568
2020/12	914	709	1,219	1,212	709	1,008	741	950	1,199	618	735	779	585
2021/03	947	744	1,267	1,291	745	1,050	774	1,002	1,240	647	755	800	605
2021/06	968	782	1,299	1,325	764	1,081	802	1,035	1,285	661	777	816	630
2021/09	1,020	809	1,368	1,377	809	1,123	837	1,072	1,374	698	801	847	667
2021/12	1,024	811	1,390	1,405	813	1,133	844	1,090	1,394	700	809	879	668
2022/03	1,013	803	1,413	1,416	790	1,137	844	1,077	1,411	698	805	878	665
2022/06	1,026	797	1,435	1,428	790	1,137	848	1,085	1,438	700	811	874	672
2022/09	977	787	1,400	1,400	755	1,092	824	1,054	1,404	683	808	857	646
2022/12	956	757	1,365	1,336	710	1,018	784	991	1,413	660	805	870	625
2023/03	948	745	1,364	1,314	712	1,022	781	982	1,392	649	805	866	613
2023/06	953	749	1,378	1,327	710	1,043	790	1,008	1,398	653	813	863	620
2023/09	985	767	1,417	1,372	730	1,075	810	1,035	1,418	660	830	846	626
2023/12	964	756	1,403	1,363	728	1,063	802	1,030	1,410	655	818	851	601
2024/03	963	751	1,391	1,391	723	1,053	795	1,017	1,454	655	819	851	612

주: 소수점 이하는 반올림되었다.

서울 내 지역의 가격 변동 양상을 살펴보면, 전체적으로는 완만한 상승세가 2012년부터 2017년까지 이어졌으며, 2018년부터 2021년 상반기까지는 모든 자치구에서 동반적으로 급등세를 보였다. 그러나 2022년 이후로는 대부분의 자치구에서 가격 상승세가 꺾이거나 조정 국면에 진입하였으며, 일부 지역에서는 소폭 하락세도 나타났다. 다만, 이러한 조정기의 가격 변화는 고가 지역일수록 완만한 양상을 보였고, 중저가 지역에서 더 두드러진 조정이 일어난 것으로 관측된다. 가격 상승이 두드러진 곳은 주로 고가 지역으로 강남구, 서초구, 성동구, 용산구로 각각 2012년 3월부터 2024년 3월까지 약 132%, 134%, 170%, 165% 상승한 것으로 나타난다. 반면, 강북구(88%), 구로구(83%), 도봉구(82%), 금천구(92%), 종로구(92%)의 경우 상대적으로 가격 상승의 속도가 완만했다. 결과적으로, 해당 기간 동안 지역 내 가격 격차 역시 심화된 것으로 나타난다(〈그림 2〉). 강남구와 비교할 때, 금천구와의 배율은 2012년 3월 2.39배에서 2024년 3월 2.89로, 강북구와의 배율은 2012년 3월 2.35배에서 2024년 3월 2.90배로 확대된 것으로 나타났다.

본 연구에서는 재고 주택의 표본을 측정 시점에 고정하고 있으므로, 전수적 추정이라는 것을 제외하면 고정 바스켓 기반 가격 지수와 유사한 성격의 지표라고 볼 수 있다. 그러므로 본 연구의 추정 결과를 기존의 한국부동산원 매매가격지수 및 KB부동산 매매가격지수와 같은 기존 공표 지수와 정량적으로 비교해 보는 것도 본 추정의 특성을 이해하는 데에 도움이 될 수 있다. 〈그림 3〉은 세 시계열을 18개

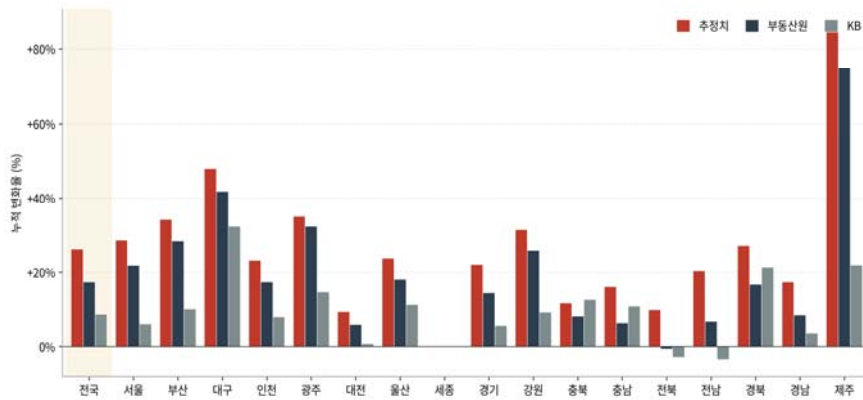


주: 그래프에서 바(bar)의 높이는 각 지역의 단위 면적(1m²) 당 평균 아파트 가격의 시점별 값을 의미한다.

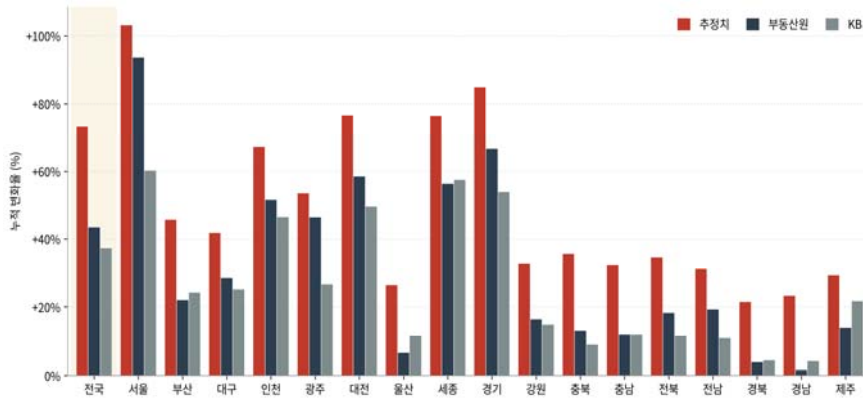
〈그림 2〉 서울 지역별 아파트 평균가격의 변화

지역(전국 및 17개 시·도)별로 2012년 1분기=100으로 재지수화한 후, 분석 기간을 회복기(2012년 2분기~2017년 1분기), 폭등기(2017년 2분기~2022년 1분기), 조정기(2022년 2분기~2024년 1분기)의 세 구간으로 구분하여 각 구간의 누적 변화율을 비교한 결과이다(〈그림 3〉).

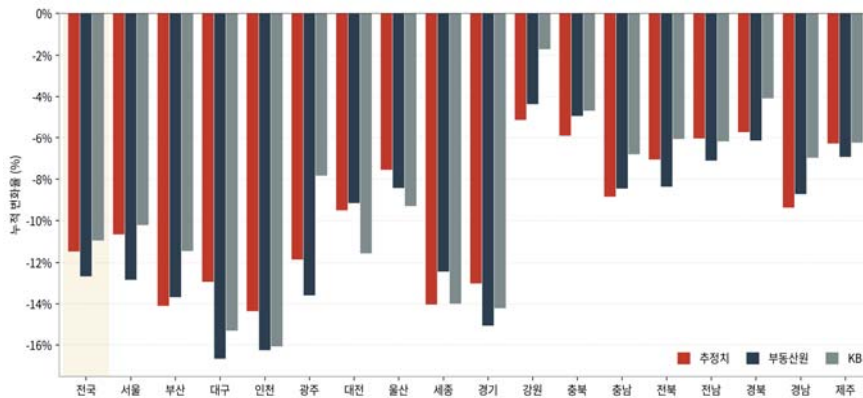
본 연구의 추정치를 통한 지수와 다른 두 가격지수의 누적 상승률, 분기 변화율의 표준편차로 측정한 변동성, 그리고 변화율 상관계수를 비교할 수 있다. 비교 결과의 주요 양상은 다음과 같다. 첫째, 본 연구 추정치의 누적 상승률은 두 공표 지수에 비해 일관되게 큰 폭으로 나타난다. 분석 기간 전체에 걸친 전국 단위의 누적 상승률은 본 연구 추정치 93%, 한국부동산원 45%, KB 34%이며, 서울의 경우 본 연구 138%, 한국부동산원 107%, KB 54%로 그 차이가 더욱 두드러진다. 다만 이러한 격차는 분석 기간 전반에 균일하게 발생하기보다는 가격 상승 국면에 집중되는 경향이 관측된다. 회복기에는 격차가 비교적 작은 편이며(전국 추정치 +26%, 한국부동산원 +17%, KB +9%), 폭등기에 격차가 가장 크게 벌어진다(전국 추정치 +73%, 한국부동산원 +44%, KB +37%; 서울 추정치 +103%, 한국부동산원 +94%, KB +60%). 반면 조정기에서는 세 지수가 모두 -10% 안팎의 비슷한 폭으로 하락하 되, 두 공표 지수의 하락폭이 본 연구 추정치보다 다소 크게 나타난다(전국 추정치 -11.5%, 한국부동산원 -12.7%, KB -11.0%). 둘째, 시기적 추세의 정합성 면에서는 세 시계열이 매우 높은 일관성을 공유한다. 셋째, 가격 정점 시점은 본 연구 추정치(전국·서울 모두 2021년 4분기)와 한국부동산원(2021년 3분기)이 거의 동일한 반면, KB지수는 2022년 2분기로 다소 늦게 형성된다. 이는 KB지수의 평할 산출 구조와 일관되는 결과이다.



(A) 2012/2분기~2017/1분기 누적변화율



(B) 2017/2분기~2022/1분기 누적변화율



(C) 2022/2분기~2024/1분기 누적변화율

주: 각 구간의 시작 시점 대비 끝 시점의 누적 변화율(%)을 계산하였다. 세종시는 1구간에서 자료 결측으로 제외되었다. 부동산원은 아파트 실거래가 매매가격지수를, KB는 주택가격동향의 월간 아파트 매매가격 지수를 사용하였다.

〈그림 3〉 추정치와 기존 가격지수의 비교

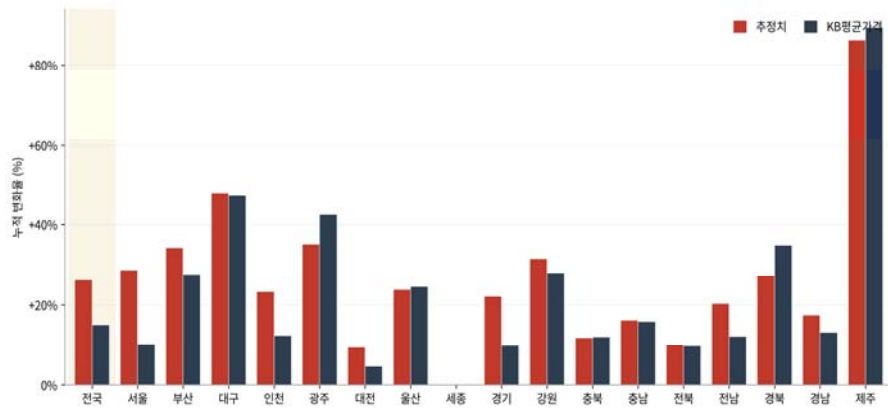
비교 결과를 요약하면, 본 연구의 추정치는 시기적 추세의 방향과 시점에 있어서 기존 공표 지표와 매우 일관된 시장 신호를 공유하면서도, 동일 자산의 가격 변화 폭은 보다 분명하게 드러내는 결과를 보인다. 이러한 차이는 부분적으로는 본 연구가 분석 기간 종료 시점 기준 고정 재고 표본의 가격을 동일 자산 바스켓에 대해 시점별로 추정하는 데에 비해, 두 공표 지표는 표본의 갱신·가중·평활화 과정을 거치는 산출 구조를 가지는 데에서 비롯한 것으로 해석될 수도 있다. 다만 가격 상승기와 조정기 사이에 관측되는 격차의 비대칭적 양상은 이러한 산출 구조의 차이만으로 온전히 설명되지 않는 요소의 가능성에 대해 검토의 여지가 있을 것으로 보인다.

본 연구의 추정치를 보다 입체적으로 평가하기 위해, KB부동산의 m^2 당 평균 매매가격(전용면적 기준) 자료와도 추가로 비교하였다. KB의 평균 가격 자료는 본 연구의 추정치와 마찬가지로 가격의 절대적 수준 정보를 담고 있다는 공통점을 가지나, 본 연구가 측정하는 면적 기준(공급면적)과 KB가 측정하는 면적 기준(전용면적)이 다르므로 절대 수준의 직접 비교는 적절하지 않다. 이에 두 시계열을 모두 2012년 1분기 = 100으로 재지수화한 후, 앞서와 동일한 세 구간 구분에 따라 각 구간의 누적 변화율을 비교하였다. <그림 4>는 그 결과를 18개 지역에 대해 보여준다.

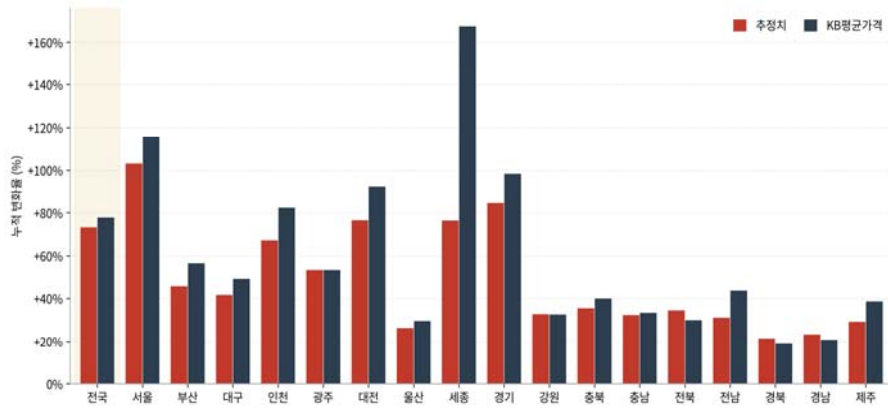
평균가격과의 누적 상승률 비교 결과는 앞서의 KB지수와와의 비교에서 관측된 양상과는 다소 다른 모습을 보인다. KB지수와와의 비교에서는 본 연구의 추정치가 일관되게 더 큰 누적 상승률을 보였으나, KB평균가격과의 비교에서는 18개 지역 중 15개 지역에서 오히려 KB평균가격의 누적 상승률이 본 연구보다 더 크게 나타난다. 특히 세종(추정치 115.6% vs KB평균 209.2%), 제주(131.7% vs 165.2%), 광주(89.3% vs 114.7%) 등에서 그 차이가 두드러진다. 반면 서울(138.0% vs 122.8%), 경기(95.5% vs 89.9%), 전국(93.4% vs 84.0%)에서는 본 연구의 추정치가 여전히 더 큰 누적 상승률을 보인다. 이러한 차이를 구간별로 살펴보면, KB평균가격의 우세는 주로 폭등기에 집중되어 나타나 는 반면(전국 추정치 +73% vs KB평균 +78%; 세종 추정치 +77% vs KB평균 +168%; 인천 추정치 +67% vs KB평균 +82%), 조정기에서는 두 시계열이 전국 단위에서 -11.5%로 거의 동일한 폭으로 하락하며 일부 지역에서는 오히려 본 연구 추정치의 하락폭이 다소 크게 관측된다. 즉, KB평균가격이 본 연구 추정치를 상회하는 양상은 가격 상승 국면에서 명확히 드러나지만, 가격 조정 국면에서는 그러한 우세가 관측되지 않는다.

이러한 누적 상승률 패턴의 비대칭성은 두 시계열의 산출 구조 차이에 기인하는 것으로 보인다. 본 연구는 재고 표본을 고정된 가운데 동일 자산 바스켓의 가격 변화를 측정하는 데에 비해, KB평균가격은 시점별로 다른 표본의 평균 가격을 산출하므로 주로 거래되는 표본의 변화(신축 단지의 추가, 거래 활성화 지역으로의 표본 비중 이동 등)가 가격 변화에 함께 반영된다. 즉, KB평균가격이 본 연구 추정치보다 큰 누적 상승률을 보이는 지역에서는 분석 기간 중 신규 공급된 단지의 가격 수준이

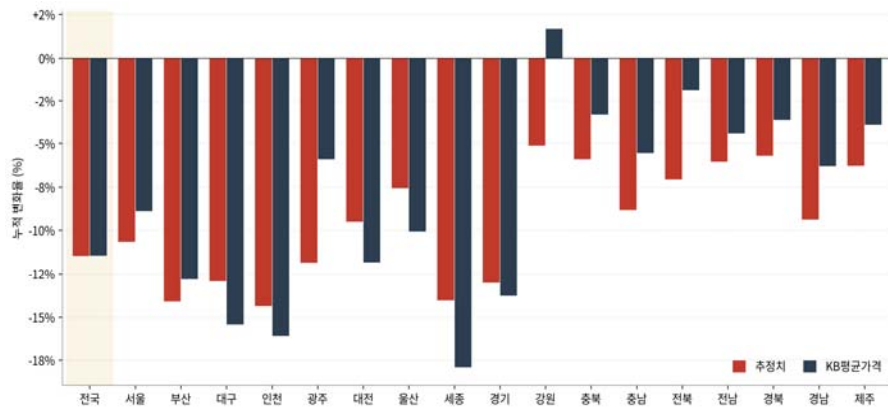
트리 앙상블 기반 자동산정모형을 이용한 재고 주택가격 분포 관측: 전국 아파트 단지 전수 자료의 시계열 분석



(A) 2012/2분기~2017/1분기 누적변화율



(B) 2017/2분기~2022/1분기 누적변화율



(C) 2022/2분기~2024/1분기 누적변화율

주: 각 구간의 시작 시점 대비 끝 시점의 누적 변화율(%)을 계산하였다. 세종시는 1구간에서 자료 결측으로 제외되었다. KB평균가격은 m²당 매매평균가격을 의미한다.

〈그림 4〉 추정치와 평균 거래가격(지수화)의 비교

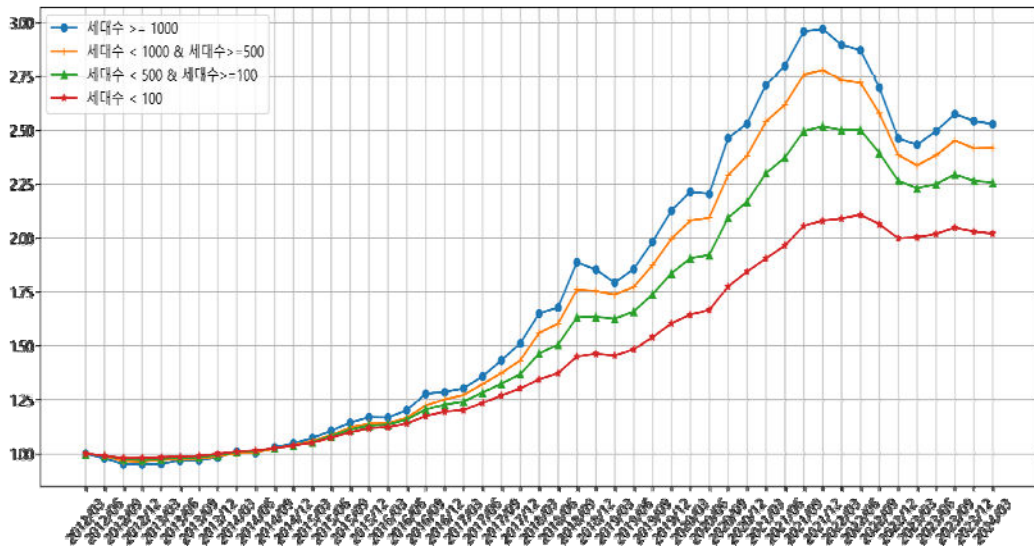
기존 단지에 비해 상대적으로 높아 표본 변화 효과가 평균을 끌어올렸을 가능성이, 반대로 본 연구 추정치가 더 큰 누적 상승률을 보이는 지역에서는 동일 자산 바스켓 자체의 가격 변동 폭이 표본 갱신 효과를 상회할 만큼 컸을 가능성이 시사된다. 이는 두 추정치가 서로 다른 질문에 답하는 도구라는 점에서 상보적인 정보를 제공함을 시사한다.

정리하면, 본 연구의 추정치는 기존 주택가격지수와 비교에서는 평활화로 인한 변동성 축소에서 적시성을, KB 평균가격과의 비교에서는 표본 구성 변화로부터 자유로운 동일 자산 바스켓 기반의 일관성을 각각 확인시켜 준다. 이러한 양 측면의 비교는 본 연구의 추정 결과가 기존 자료들이 갖는 산출 구조상의 두 가지 한계인 평활화로 인한 적시성 저하 및 표본 구성 변화로 인한 일관성 저하를 동시에 통제된 분석 도구로서 기능함을 정량적으로 뒷받침한다. 다만 두 비교에서 공통적으로 관측되는, 가격 상승 국면에서 격차가 두드러지고 조정 국면에서는 그러한 격차가 악화되는 비대칭적 양상은 산출 구조의 차이만으로 전부 설명되지 않을 가능성을 시사하는 측면이 있으며, 이에 대한 보다 면밀한 검토는 향후 과제로 남는다.

본 연구에서 제시된 방식의 장점은 분석 목적에 따라 하위 집단의 가격지수를 자유롭게 편성할 수 있다는 점이다. 전수적 자료에 기반하여 개별 주택 단위까지 가격이 추정되어 있으므로, 지역, 면적, 준공연도, 세대수 등 임의의 기준에 따라 부분 표본을 구성하고 그에 대응하는 가격지수를 즉시 도출할 수 있다. 이는 기존의 공표 가격지수가 제공하지 못하는 세분화된 시장 진단을 가능하게 한다는 점에서 본 방식의 활용 가치를 잘 보여준다.

그 예시로 단지 규모에 따른 가격 변동의 차이를 살펴보았다. 구체적으로, 분석 기점인 2012년 3월을 기준 시점(=1)으로 하여 서울 지역 아파트를 단지 세대수에 따라 100세대 이하, 100~500세대, 500~1,000세대, 1,000세대 이상의 네 집단으로 구분한 뒤 각 집단별 가격 변화 추이를 비교 관찰하였다. <그림 5>에서 확인할 수 있듯이, 분석 기간 동안 서울 지역의 아파트 가격이 전반적으로 상승하는 가운데서도 특히 단지 규모가 큰 아파트일수록 가격 상승 폭이 더욱 가파르게 나타났다. 즉, 세대수가 큰 집단일수록 누적 상승률이 높았으며, 집단 간 격차는 시간이 경과함에 따라 점진적으로 확대되는 양상을 보였다.

일반적으로 단지의 세대수가 크다는 것은 자산으로서의 유동성, 즉 향후 거래 성사 가능성이 높고 거래 비용이 상대적으로 낮음을 의미한다. 따라서 <그림 5>가 보여주는 결과는, 주택가격 상승기에 들어서면서 시장이 대규모 단지가 갖는 높은 유동성과 그로부터 함의되는 자산적 가치를 점차 더 높게 평가하는 방향으로 변화해 왔음을 시사한다. 이는 동일한 시기, 동일한 지역 내의 아파트라 하더라도 단지 규모라는 미시적 속성에 따라 가격 형성 메커니즘이 차별적으로 작동할 수 있음을 보여주는 결과이며, 본 연구에서 구축한 전수적 가격추정은 이러한 세분화된 시장 동학을 포착하는데 활용될 수 있음을 동시에 보여준다.



주: 각 지수는 2012/1분기를 기준으로 지수화되었다.

〈그림 5〉 단지 크기별 아파트 가격 추이(서울)

2. 가격 분포의 추이

이번에는 분위값을 통해 가격의 분포를 관찰해보자. 〈표 11〉은 전국의 아파트 전체를 대상으로 그 가격의 분포 추이를 관측한 결과이다. 먼저 확인할 수 있는 것은, 해당기간 동안 평균가격 뿐 아니라 가격의 분포 구조 자체가 변화해 왔다는 것이다.

2012년 초 기준으로 가격 분포의 하단부인 10% 분위수는 m^2 당 116만 원 내외 수준에 형성되어 있었으며, 중앙값(50% 분위수)은 208만 원/ m^2 , 90% 분위수는 396만 원/ m^2 에 머물렀다. 이후 2014에서 2019년 사이에는 완만한 상승세가 이어졌으며, 전국 각 분위수 모두에서 가격 수준이 점진적으로 상승하였다. 그러나 눈에 띄는 변화는 2020년을 전후로 급격하게 나타났다. 이 시기부터 특히 상위 분위수의 상승 폭이 하위 분위수보다 뚜렷하게 크다는 점이 확인되었다. 예를 들어, 90% 분위수의 경우 2020년 중반 이후 빠르게 1,000만 원/ m^2 을 상회하였고, 이후 2021년에는 1,082만 원/ m^2 에 육박하는 수준으로 치솟았다. 하지만 같은 기간 하위 10% 분위수와 25% 분위수는 다소 더디게 상승하여, 분포의 상하위 간 격차가 점차 확대되는 구조가 나타났다. 다시 말해, 특정 고가 지역(예: 서울, 세종, 일부 수도권 신도시 등)의 가격 급등이 전체 가격 분포 상위 분위수의 빠른 상승을 견인하였으며, 상대적으로 저가 주택이 분포한 지역은 상승폭이 제한되었음을 시사한다. 2022년 이후의 분포 변화

〈표 11〉 아파트 가격 분포: 전국(단위: 만 원/m²)

시점	5% 분위수	10% 분위수	25% 분위수	중앙값(50%)	75% 분위수	90% 분위수	95% 분위수
2012/03	96	116	155	208	289	396	501
2012/06	98	117	155	207	283	390	494
2012/09	98	118	156	206	280	382	482
2012/12	98	118	156	206	278	380	483
2013/03	100	120	159	208	279	382	484
2013/06	101	121	159	209	282	384	488
2013/09	103	123	162	212	285	388	492
2013/12	103	124	164	215	289	393	495
2014/03	105	125	167	219	294	399	504
2014/06	105	126	168	219	295	402	504
2014/09	106	127	171	224	299	408	518
2014/12	107	129	174	227	305	416	524
2015/03	109	131	176	231	310	426	532
2015/06	109	133	181	237	320	438	550
2015/09	111	136	185	244	329	453	568
2015/12	111	136	186	245	332	460	575
2016/03	112	137	186	246	333	462	574
2016/06	112	138	186	248	337	471	587
2016/09	112	139	189	252	346	490	616
2016/12	113	140	190	256	354	501	625
2017/03	113	140	191	257	357	505	632
2017/06	113	140	192	258	361	520	657
2017/09	113	141	192	261	370	542	691
2017/12	112	140	192	261	371	553	715
2018/03	111	138	191	261	376	584	789
2018/06	108	137	188	261	382	604	812
2018/09	107	134	187	262	396	670	889
2018/12	105	133	187	263	406	670	883
2019/03	103	130	184	263	408	660	865
2019/06	101	127	183	263	414	673	884
2019/09	100	127	183	266	419	702	939
2019/12	100	127	185	272	436	747	1,001
2020/03	100	127	187	280	477	795	1,053
2020/06	100	128	190	290	491	805	1,059

〈표 11〉 계속

시점	5% 분위수	10% 분위수	25% 분위수	중앙값(50%)	75% 분위수	90% 분위수	95% 분위수
2020/09	101	129	194	302	533	891	1,155
2020/12	102	132	201	325	569	931	1,206
2021/03	105	136	212	356	620	1,005	1,292
2021/06	107	140	220	375	645	1,033	1,330
2021/09	111	148	234	406	688	1,088	1,408
2021/12	114	153	244	415	689	1,082	1,398
2022/03	116	154	245	410	667	1,059	1,369
2022/06	118	157	247	405	657	1,049	1,355
2022/09	119	157	243	386	611	982	1,269
2022/12	117	153	232	358	569	907	1,176
2023/03	116	151	226	347	559	899	1,162
2023/06	116	151	226	349	571	913	1,189
2023/09	116	151	226	354	583	933	1,223
2023/12	116	150	225	352	575	919	1,203
2024/03	115	148	222	350	572	918	1,205

주: 소수점 이하는 반올림되었다.

에서도 또 하나의 중요한 특징이 나타난다. 전체 가격 수준이 다소 조정 국면에 들어간 것으로 보이지만, 분포 상단(90%, 95%)은 여전히 높은 수준을 유지하고 있는 반면, 하단 분위수에서는 가격이 다소 하락하거나 정체되는 모습을 보이고 있다. 이는 가격 조정기에도 고가 지역 아파트의 절대 가격 수준이 유지되는 경향이 있으며, 시장 내에서 가격 하방의 탄력성이 낮음을 의미할 수 있다. 종합적으로 이러한 변화는 전국적으로 일정부분 공통적인 가격 변화 추세를 공유하는 가운데에서도, 중상위 이상 가격대 아파트의 상대적 상승률 및 하방경직성이 더 크다는 것을 보여주는 결과라 할 수 있다. 물론 주의할 점은, 전국 단위의 가격 분포는 지역 간 격차를 그대로 반영한다는 것이다. 그러므로 위와 같은 비대칭성의 확대는 지역간 격차의 확대의 재확인이라고도 볼 수 있을 것이다.

시도간 격차 확대를 제외한 아파트 가격 분포 추이를 관측하려면, 지역 내 가격 분포를 살펴봐야 할 것이다. 다만 모든 지역을 상세히 기술하면 지면을 지나치게 차지하게 되므로, 본 연구에서는 서울의 경우에 대해서 자세히 기술하려 한다. 〈표 12〉는 서울 내 아파트 가격 분포를 보여준다. 서울은 전국에서 가장 높은 가격 수준을 유지하고 있는 지역으로, 가격 분포의 상단이 특히 빠르게 팽창하는 특징이 있다. 2012년 초 기준으로, 서울 아파트 가격의 하단부(10% 분위수)는 303만 원/m², 중앙값은 410만 원/m², 90% 분위수는 765만 원/m² 수준이었다. 2014년부터 2019년까지의 기간에는

〈표 12〉 아파트 가격 분포: 서울(단위: 만 원/m²)

시점	5% 분위수	10% 분위수	25% 분위수	중앙값(50%)	75% 분위수	90% 분위수	95% 분위수
2012/03	283	303	342	410	534	765	879
2012/06	278	297	335	403	524	756	860
2012/09	273	291	332	398	516	730	830
2012/12	271	290	331	395	517	726	830
2013/03	271	291	332	396	516	738	838
2013/06	272	292	333	399	523	740	844
2013/09	274	294	335	401	523	739	849
2013/12	275	296	339	405	530	748	852
2014/03	279	301	342	410	539	765	879
2014/06	278	299	342	413	539	766	885
2014/09	283	303	347	418	549	790	902
2014/12	286	308	352	425	557	787	916
2015/03	292	312	357	433	565	811	953
2015/06	297	322	366	442	584	836	984
2015/09	306	330	379	460	603	849	1,006
2015/12	311	335	384	468	612	877	1,031
2016/03	311	336	386	470	613	870	1,022
2016/06	316	341	392	479	624	893	1,061
2016/09	326	352	406	500	661	967	1,154
2016/12	329	355	414	512	667	972	1,154
2017/03	331	360	416	515	677	981	1,182
2017/06	337	368	426	532	702	1,035	1,238
2017/09	345	378	443	558	732	1,086	1,319
2017/12	350	384	449	572	776	1,161	1,428
2018/03	355	392	464	606	848	1,308	1,577
2018/06	360	399	479	624	880	1,303	1,590
2018/09	379	430	523	696	967	1,435	1,767
2018/12	395	443	539	701	955	1,379	1,676
2019/03	393	444	537	692	942	1,351	1,613
2019/06	398	450	544	710	967	1,405	1,709
2019/09	409	461	560	743	1,033	1,509	1,866
2019/12	420	477	586	779	1,105	1,654	2,017
2020/03	442	506	624	827	1,161	1,674	1,997
2020/06	447	516	637	839	1,156	1,658	1,968
2020/09	508	581	713	918	1,265	1,824	2,163
2020/12	529	602	741	960	1,311	1,854	2,189

〈표 12〉 계속

시점	5% 분위수	10% 분위수	25% 분위수	중앙값(50%)	75% 분위수	90% 분위수	95% 분위수
2021/03	560	642	794	1,026	1,389	1,984	2,325
2021/06	578	666	821	1,062	1,432	2,028	2,393
2021/09	608	705	878	1,112	1,490	2,154	2,508
2021/12	612	709	878	1,112	1,498	2,168	2,575
2022/03	611	700	863	1,090	1,461	2,146	2,559
2022/06	618	698	857	1,082	1,447	2,108	2,566
2022/09	590	665	803	1,016	1,374	2,013	2,457
2022/12	564	631	751	938	1,271	1,869	2,237
2023/03	553	616	732	929	1,252	1,865	2,246
2023/06	556	618	737	937	1,276	1,939	2,322
2023/09	559	622	745	959	1,319	2,010	2,402
2023/12	556	616	736	947	1,289	1,989	2,371
2024/03	554	612	731	940	1,289	1,976	2,362

주: 소수점 이하는 반올림되었다.

하단부부터 상단부에 이르기까지 전 구간에 걸쳐 뚜렷한 상승이 지속되었다. 예를 들어, 중앙값은 2012년 410만 원에서 2019년 말에는 640만 원 수준까지 약 56% 상승했으며, 90% 분위수도 같은 기간 약 750만 원에서 1,127만 원/m²로 대폭 상승하였다. 그러나 가장 가파른 상승은 2020년에서 2021년 사이에 집중적으로 발생하였다. 서울의 중앙값은 2020년 초 680만 원에서 2021년 말에는 922만 원을 돌파하였고, 90% 분위수는 같은 기간 동안 1,200만 원을 넘어서며 고가 아파트의 폭발적인 상승세가 있었음을 확인할 수 있다.

눈여겨볼 점은, 서울의 경우 가격 상승이 단지 평균 가격의 상승에 그치지 않고, 분포의 상단이 상대적으로 더 빠르게 위로 늘어나는 상단 팽창형 구조를 나타낸다는 것이다. 이는 특히 95% 분위수의 변화를 통해 더욱 뚜렷하게 확인된다. 2012년 약 880만 원/m²이었던 95% 분위수는 2022년에는 약 1,600만 원/m²에 도달하였으며, 이는 서울 내에서도 상위 5% 아파트의 가격 상승률이 다른 계층 대비 월등히 높았음을 의미한다. 중위값 역시 2020년 초 827만 원에서 2021년 말까지 1,490만 원으로 크게 팽창하였다. 그러나 같은 기간 하위 10% 분위수는 약 300만 원에서 450만 원 수준으로의 상승에 그쳤다. 요약하면, 전반적인 가격 상승의 기초 하에서도 중위값 이상의 고가 아파트의 가격 상승이 더 컸다는 것을 보여준다. 또한, 2022년 이후의 조정 국면에서도 서울의 가격 분포는 일정한 특징을 유지하고 있다. 전체 분위수 중 하위 분위수(예: 10%, 25%)에서는 미미한 조정 또는 정체 양상이 나타난 반면, 상위 분위수(90%, 95%)에서는 조정폭이 제한적이며, 여전히 고점에 근접한 가격 수준을 유지하고 있다. 이는 전국 분포에서

관찰된 바와 마찬가지로 고가 아파트의 하방 경직성이 더 강하게 나타나고 있음을 보여준다.

흥미로운 것은 이러한 경향이 전국적으로 공통적으로 관찰된다는 것이다. <표 13>은 각 지역별 가격 분포를 토대로, 각 지역별 가격의 75분위수와 25분위수의 비율이 어떻게 변하는 지를 보여주고 있다. 결과는, 비록 지역마다 정도의 차이는 있으나, 상위 분위 가격 아파트와 하위 또는 중앙값의 비율이 2018년 즈음을 기점으로 2021년까지 거의 모든 지역에서 벌어졌다는 것을 보여준다. 예를 들어, 부산의 경우 2012년 3월에는 75분위수와 25분위수의 비율은 1.37배 수준이었으나 2021년 9월을 기준으로 2.16배까지 상승한 것을 확인할 수 있다. 이러한 경향은 대부분의 다른 지역에서도 나타난다. 2022년 이후 분위수간 격차는 다소 진정되는 추세이나, 대부분의 지역에서 2018년 이전 수준을 회복하지는 못하고 있다. 이는 아파트 가격의 양극화가 지역 간에만 발생하는 것이 아니라 지역 내에서도 심화되어 왔으며, 전국에서 공통적으로 나타났음을 보여준다.

<표 13> 아파트 가격 분위수 비율 75~25(단위: 배)

시점	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2012/03	1.56	1.37	1.37	1.30	1.33	1.45	1.45	1.28	1.51	1.55	1.67	1.53	1.60	1.50	1.59	1.49	1.34
2012/06	1.56	1.37	1.36	1.31	1.32	1.44	1.42	1.23	1.47	1.52	1.63	1.54	1.57	1.48	1.58	1.47	1.39
2012/09	1.56	1.37	1.36	1.33	1.33	1.44	1.38	1.19	1.47	1.52	1.61	1.53	1.57	1.49	1.55	1.45	1.36
2012/12	1.56	1.37	1.38	1.29	1.34	1.45	1.40	1.12	1.48	1.56	1.61	1.54	1.59	1.46	1.57	1.46	1.38
2013/03	1.55	1.38	1.34	1.31	1.33	1.46	1.40	1.15	1.46	1.53	1.59	1.51	1.57	1.45	1.54	1.44	1.47
2013/06	1.57	1.40	1.37	1.27	1.35	1.45	1.41	1.14	1.49	1.51	1.58	1.52	1.58	1.44	1.54	1.49	1.42
2013/09	1.56	1.38	1.33	1.33	1.34	1.44	1.40	1.11	1.49	1.49	1.56	1.50	1.61	1.44	1.55	1.47	1.45
2013/12	1.56	1.39	1.31	1.35	1.35	1.46	1.41	1.11	1.49	1.51	1.55	1.50	1.61	1.45	1.57	1.49	1.39
2014/03	1.58	1.40	1.32	1.31	1.33	1.44	1.39	1.13	1.50	1.50	1.54	1.51	1.61	1.48	1.58	1.51	1.46
2014/06	1.57	1.40	1.32	1.36	1.34	1.43	1.39	1.14	1.52	1.49	1.61	1.53	1.60	1.49	1.57	1.52	1.42
2014/09	1.58	1.41	1.31	1.31	1.33	1.40	1.38	1.14	1.52	1.48	1.60	1.53	1.61	1.49	1.60	1.55	1.38
2014/12	1.58	1.42	1.31	1.34	1.36	1.44	1.38	1.14	1.53	1.50	1.65	1.51	1.62	1.49	1.62	1.54	1.40

〈표 13〉 계속

시점	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2015/03	1.58	1.42	1.33	1.32	1.37	1.43	1.38	1.10	1.54	1.47	1.62	1.53	1.65	1.50	1.64	1.53	1.40
2015/06	1.60	1.45	1.37	1.30	1.35	1.45	1.40	1.12	1.54	1.49	1.64	1.52	1.65	1.51	1.65	1.52	1.42
2015/09	1.59	1.46	1.39	1.31	1.37	1.41	1.41	1.12	1.52	1.50	1.58	1.55	1.68	1.51	1.67	1.51	1.43
2015/12	1.59	1.48	1.39	1.32	1.37	1.41	1.42	1.14	1.51	1.51	1.61	1.56	1.69	1.52	1.68	1.52	1.43
2016/03	1.59	1.49	1.35	1.33	1.36	1.45	1.42	1.21	1.50	1.47	1.61	1.55	1.72	1.53	1.64	1.52	1.40
2016/06	1.59	1.47	1.36	1.29	1.36	1.44	1.44	1.22	1.52	1.48	1.58	1.54	1.72	1.54	1.67	1.50	1.34
2016/09	1.63	1.51	1.37	1.30	1.38	1.47	1.43	1.27	1.51	1.49	1.61	1.57	1.72	1.57	1.69	1.51	1.33
2016/12	1.61	1.55	1.42	1.32	1.37	1.44	1.45	1.27	1.53	1.53	1.58	1.55	1.73	1.62	1.70	1.51	1.33
2017/03	1.63	1.54	1.42	1.33	1.40	1.48	1.45	1.25	1.53	1.54	1.53	1.55	1.75	1.58	1.74	1.52	1.36
2017/06	1.65	1.55	1.42	1.32	1.42	1.48	1.44	1.26	1.54	1.54	1.60	1.56	1.74	1.57	1.78	1.52	1.39
2017/09	1.65	1.55	1.45	1.34	1.42	1.46	1.45	1.28	1.56	1.56	1.63	1.56	1.75	1.61	1.82	1.52	1.40
2017/12	1.73	1.55	1.46	1.32	1.41	1.49	1.47	1.31	1.57	1.58	1.64	1.58	1.75	1.64	1.83	1.53	1.41
2018/03	1.83	1.54	1.47	1.32	1.46	1.50	1.51	1.30	1.61	1.57	1.66	1.62	1.78	1.60	1.91	1.54	1.44
2018/06	1.84	1.57	1.52	1.34	1.52	1.53	1.52	1.34	1.66	1.59	1.67	1.63	1.76	1.64	1.92	1.59	1.46
2018/09	1.85	1.57	1.55	1.34	1.51	1.55	1.54	1.43	1.73	1.62	1.74	1.67	1.75	1.65	1.97	1.61	1.42
2018/12	1.77	1.57	1.56	1.34	1.58	1.60	1.56	1.42	1.79	1.63	1.76	1.72	1.81	1.65	1.97	1.61	1.44
2019/03	1.75	1.59	1.55	1.39	1.65	1.66	1.61	1.44	1.80	1.67	1.79	1.71	1.78	1.67	2.02	1.66	1.45
2019/06	1.78	1.61	1.57	1.37	1.65	1.67	1.63	1.49	1.81	1.71	1.80	1.76	1.80	1.64	2.07	1.69	1.47
2019/09	1.85	1.63	1.56	1.40	1.67	1.78	1.66	1.55	1.85	1.72	1.82	1.78	1.81	1.74	2.10	1.72	1.47

〈표 13〉 계속

시점	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2019 /12	1.89	1.69	1.54	1.43	1.68	1.84	1.71	1.62	1.90	1.72	1.81	1.80	1.86	1.70	2.11	1.77	1.47
2020 /03	1.86	1.74	1.58	1.50	1.68	1.88	1.74	1.58	2.03	1.77	1.85	1.84	1.86	1.70	2.12	1.81	1.47
2020 /06	1.81	1.75	1.57	1.53	1.70	1.85	1.75	1.49	2.00	1.77	1.95	1.84	1.87	1.71	2.19	1.81	1.47
2020 /09	1.77	1.81	1.59	1.54	1.71	1.87	1.83	1.40	2.03	1.80	1.99	1.93	1.88	1.78	2.21	1.84	1.47
2020 /12	1.77	2.07	1.68	1.59	1.80	1.82	2.03	1.42	2.02	1.81	2.04	1.97	1.91	1.78	2.21	1.91	1.49
2021 /03	1.75	2.12	1.64	1.60	1.82	1.93	2.09	1.45	1.98	1.85	2.12	2.02	2.00	1.81	2.24	1.96	1.57
2021 /06	1.74	2.16	1.64	1.55	1.87	1.89	2.02	1.41	1.91	1.88	2.21	2.03	2.02	1.81	2.22	1.97	1.68
2021 /09	1.70	2.16	1.62	1.54	1.92	1.86	2.07	1.36	1.84	1.92	2.19	2.04	2.03	1.85	2.22	2.02	1.77
2021 /12	1.71	2.09	1.62	1.52	1.90	1.87	2.03	1.36	1.80	1.97	2.16	2.01	2.12	1.89	2.20	2.07	1.80
2022 /03	1.69	2.07	1.57	1.46	1.85	1.78	2.04	1.34	1.75	1.96	2.14	2.02	2.04	1.87	2.21	2.05	1.83
2022 /06	1.69	2.04	1.56	1.41	1.83	1.75	2.04	1.32	1.75	1.95	2.09	1.99	2.05	1.91	2.16	2.02	1.86
2022 /09	1.71	1.97	1.56	1.42	1.78	1.67	2.00	1.31	1.72	1.89	2.00	1.94	2.03	1.87	2.13	2.00	1.84
2022 /12	1.69	1.89	1.56	1.39	1.76	1.65	1.94	1.33	1.74	1.86	1.94	1.91	2.00	1.85	2.11	1.92	1.75
2023 /03	1.71	1.88	1.57	1.43	1.75	1.69	1.89	1.41	1.77	1.82	1.90	1.90	1.95	1.84	2.06	1.91	1.76
2023 /06	1.73	1.88	1.61	1.46	1.75	1.69	1.91	1.40	1.81	1.82	1.93	1.90	1.97	1.84	2.09	1.90	1.75
2023 /09	1.77	1.92	1.65	1.50	1.74	1.73	1.95	1.42	1.81	1.85	1.89	1.88	1.97	1.86	2.07	1.92	1.71
2023 /12	1.75	1.91	1.67	1.51	1.72	1.78	1.95	1.40	1.80	1.83	1.92	1.91	2.00	1.84	2.08	1.95	1.71
2024 /03	1.76	1.92	1.66	1.53	1.73	1.75	1.95	1.41	1.82	1.84	1.93	1.87	1.96	1.85	2.07	1.97	1.74

주: 소수점 세자리 이하는 반올림되었다.

3. 분석적 활용 예시 1: 가격 불균등성 지표의 산출

앞에서는 각 지역별 가격 분위수 및 분위수 간의 배율을 확인함으로써 가격 분포의 추이를 확인하였다. 이번에는 직접적인 불균등성 지표인 지니계수를 관측해볼 수 있다. 지니계수는 소득·자산·주택가격처럼 비음수(0 이상)로 측정되는 연속형 변수의 분포가 '완전평등(모든 관측치가 동일)' 상태에서 얼마나 이탈했는지를 0에서 1 사이의 단일 지표로 정량화한 값이다. 값이 0에 가까우면 평등, 1에 가까우면 불평등이 심함을 뜻하며, 관측치 전체를 쌍대 비교해 얻은 절대격차의 평균을 전체 평균으로 정규화한 수준으로 이해할 수 있다. 이때 지니계수의 산식은 아래와 같다.

$$Gini = \frac{1}{2n^2\mu} \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|, \quad \langle \text{식 4} \rangle$$

이때 n, x_i, μ 는 표본수, i 번째 관측값, 평균을 의미한다. 지니계수는 다음과 같은 특징을 갖는다. 첫째, 지니계수는 척도불변성을 가진다. 이는 모든 관측값에 동일한 비율을 곱하더라도 지니계수의 값은 변하지 않는다는 것을 의미하며, 이를 통해 단순한 물가 상승이나 가격 수준의 전반적인 이동과는 무관하게 분포의 불균등성만을 측정할 수 있게 된다. 둘째, 지니계수는 값의 범위가 0에서 1 사이로 제한되기 때문에 시점 간 또는 지역 간의 비교가 직관적으로 가능하다. 값이 0에 가까울수록 완전한 평등 상태에, 1에 가까울수록 극심한 불균등 상태에 해당한다. 마지막으로, 지니계수는 계산 효율성 측면에서도 장점을 가진다. 누적비율(로렌츠 곡선)의 면적을 활용하면 비교적 간단한 정렬과 합산만으로 계산이 가능하므로, 대규모 데이터를 다룰 때에도 실용적으로 사용할 수 있다. 이러한 이유로 지니계수는 불균등의 정도를 하나의 수치로 요약하여 시기별 또는 지역 간 비교 분석을 수행할 수 있는 대표적인 지표로 널리 활용된다. 따라서 앞서 살펴본 분위수 기반 지표가 분포의 위치·폭 변화를 보여줬다면, 지니계수는 전체 분포의 불균등성 수준을 단일 지표로 요약해 지역 간·시점 간 불평등 구조를 비교할 수 있다는 장점이 있다.

본 연구에서는 단지별 아파트 가격의 추정치를 통해 지역별 지니계수의 변화를 산출하였다. 본 연구에서 추정한 지니계수는 아파트 단지·평형 단위의 m^2 당 가격을 관측 단위로 산출한 것으로, 자산(주택)의 단위 가격 분포에 대한 불균등성 지표이다. 이는 가구 단위 소득 또는 가구 단위 자산을 관측 대상으로 하는 통상의 지니계수와 분석 단위가 상이하며, 가구의 주택 보유 수, 거주 가구원 수, 가구 소득 등은 반영하지 않는다는 것을 염두에 두어야 한다. 즉, 본 결과는 "주택의 단위 가격 분포가 얼마나 분산되어 있는가"에 대한 측정으로 이해되어야 하며, "가계 자산의 불균등성"이나 "주거 부담의 불균등성"과는 구별되어야 한다.

〈표 14〉는 각 지역에 대해 아파트 단위 면적 당 가격의 지니계수를 도출한 결과이다. 당연한 결과이지만, 전국을 대상으로 하는 경우 각 지역별 지니계수보다 더 높은 값(즉, 더 불균등한 결과)이 도출된다. 표를 보면, 앞 장에서 얻었던 것과 유사한 합의 결과를 확인할 수 있다. 결과를 요약하면 다음과 같다. 먼저, 지역 간 불균등성의 수준은 전반적으로 크게 상승한다. 전국을 대상으로 지니계수를 추정하는 경우, 2012년 3월 기준 0.283이지만, 2024년 3월에는 0.397까지 상승한다. 둘째, 지역 내 불균등성의 수준도 2018년을 기점으로 2021년까지 뚜렷하게 상승한다. 예를 들어, 경기도의 경우 2012년 3월 기준 지니계수는 0.174에 불과했으나, 2021년 9월이 되면 0.259까지 상승한다. 이러한 경향은, 지역에 따라 정도의 차이는 있으나 전국의 각 지역에서 공통적으로 발견된다. 셋째, 2022년 이후 지역에 따라서 약간의 불균등성 하락이 관측되는 곳이 있다. 하지만 서울의 경우 2022년 이후에도 큰 조정없이 불균등성이 커지는 추세를 보인다. 결론적으로 지니계수의 추정결과 역시 앞의 가격 분포의 관측을 통해 확인할 수 있었던 것과 유사한, 즉 가격의 양극화가 전국적으로도 지역 내에서도 전반적으로 확대되었음을 의미하는 결과를 얻을 수 있다.

본 연구의 가격 지니계수 추정치를 기존의 관련 추정치와 비교해 두는 것은 결과 해석에 도움이 된다. 우선 분석 단위가 다른 통계청의 가구 단위 순자산 지니계수와의 비교가 가능하다. 통계청 가계금융복지조사에 따르면 가구 순자산 지니계수는 2017년 0.584에서 2021년 약 0.603으로 상승하였으며(통계청, 2018, 2022), 이러한 가구 자산 불평등 심화의 주된 원동력은 주택가격 상승으로 보고된다. 본 연구가 추정한 전국 단위 가격 지니계수의 동일 기간 상승 폭(2017년 0.29 내외 → 2021년 0.41 수준)은 가구 단위 지표의 상승 추세와 방향 및 시점에 있어 일관된 신호를 보여, 본 연구의 추정 결과가 거시적 자산 불평등 흐름과 모순되지 않음을 시사한다. 다만 두 지표의 분석 단위(가구 자산과 주택 가격)와 자산 범위(전체 자산과 아파트 단위 가격)가 다르므로 절대 수준의 일대일 비교는 적절하지 않으며, 본 연구의 결과는 가구 단위 지표를 보완하는 자산(주택) 단위 정보로 해석되어야 한다.

〈표 14〉 아파트 가격 분위수 비율 90~50(단위: 배)

시점	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2012/03	1.30	1.17	1.16	1.16	1.15	1.20	1.18	1.09	1.22	1.21	1.28	1.21	1.25	1.20	1.22	1.18	1.16
2012/06	1.30	1.17	1.18	1.15	1.14	1.20	1.17	1.07	1.21	1.20	1.27	1.22	1.26	1.21	1.21	1.18	1.20
2012/09	1.30	1.17	1.17	1.16	1.15	1.20	1.15	1.04	1.21	1.21	1.25	1.23	1.27	1.19	1.21	1.17	1.17
2012/12	1.31	1.17	1.17	1.15	1.16	1.20	1.16	1.04	1.22	1.22	1.24	1.21	1.26	1.18	1.19	1.18	1.17

〈표 14〉 계속

시점	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2013/03	1.30	1.17	1.15	1.15	1.15	1.19	1.15	1.05	1.21	1.22	1.23	1.21	1.26	1.18	1.20	1.18	1.21
2013/06	1.31	1.17	1.17	1.13	1.16	1.19	1.14	1.04	1.22	1.21	1.22	1.22	1.27	1.20	1.20	1.20	1.20
2013/09	1.30	1.17	1.15	1.18	1.17	1.18	1.15	1.06	1.22	1.21	1.22	1.20	1.29	1.18	1.19	1.19	1.20
2013/12	1.31	1.17	1.14	1.18	1.16	1.19	1.17	1.04	1.23	1.22	1.21	1.21	1.27	1.19	1.19	1.20	1.19
2014/03	1.32	1.17	1.15	1.16	1.14	1.18	1.15	1.06	1.23	1.21	1.22	1.19	1.27	1.22	1.20	1.20	1.22
2014/06	1.31	1.16	1.15	1.17	1.15	1.19	1.15	1.05	1.24	1.22	1.24	1.21	1.28	1.22	1.19	1.21	1.18
2014/09	1.31	1.17	1.14	1.15	1.14	1.17	1.15	1.05	1.23	1.20	1.26	1.20	1.27	1.20	1.22	1.21	1.17
2014/12	1.31	1.18	1.15	1.17	1.17	1.18	1.15	1.05	1.24	1.20	1.28	1.18	1.28	1.21	1.23	1.19	1.18
2015/03	1.30	1.17	1.16	1.16	1.16	1.19	1.16	1.04	1.24	1.19	1.27	1.19	1.30	1.21	1.24	1.20	1.19
2015/06	1.32	1.18	1.19	1.16	1.14	1.18	1.17	1.06	1.23	1.19	1.24	1.20	1.31	1.22	1.25	1.19	1.20
2015/09	1.31	1.19	1.19	1.15	1.16	1.17	1.18	1.06	1.22	1.18	1.20	1.21	1.32	1.21	1.25	1.17	1.21
2015/12	1.31	1.20	1.21	1.16	1.16	1.16	1.19	1.07	1.22	1.19	1.24	1.22	1.30	1.21	1.26	1.19	1.19
2016/03	1.30	1.20	1.18	1.15	1.16	1.19	1.19	1.11	1.22	1.18	1.23	1.21	1.30	1.21	1.24	1.19	1.18
2016/06	1.30	1.20	1.19	1.14	1.17	1.18	1.16	1.12	1.22	1.19	1.23	1.22	1.33	1.21	1.25	1.18	1.15
2016/09	1.32	1.21	1.19	1.16	1.18	1.17	1.16	1.10	1.21	1.19	1.21	1.22	1.31	1.22	1.26	1.19	1.14
2016/12	1.30	1.23	1.21	1.16	1.17	1.18	1.16	1.11	1.22	1.22	1.19	1.21	1.33	1.24	1.26	1.19	1.15
2017/03	1.32	1.23	1.21	1.16	1.18	1.19	1.17	1.10	1.22	1.20	1.18	1.22	1.32	1.25	1.25	1.19	1.14
2017/06	1.32	1.24	1.21	1.17	1.19	1.19	1.17	1.12	1.23	1.18	1.21	1.23	1.31	1.24	1.28	1.20	1.17
2017/09	1.31	1.25	1.21	1.17	1.18	1.18	1.18	1.12	1.23	1.18	1.21	1.23	1.30	1.24	1.29	1.21	1.20
2017/12	1.36	1.23	1.23	1.16	1.19	1.20	1.20	1.12	1.24	1.20	1.21	1.21	1.30	1.25	1.29	1.19	1.20

〈표 14〉 계속

시점	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2018/03	1.40	1.24	1.23	1.16	1.21	1.21	1.21	1.11	1.27	1.21	1.22	1.24	1.30	1.23	1.33	1.21	1.20
2018/06	1.41	1.24	1.24	1.17	1.23	1.22	1.21	1.12	1.29	1.22	1.24	1.24	1.26	1.26	1.33	1.24	1.22
2018/09	1.39	1.25	1.24	1.18	1.24	1.23	1.23	1.19	1.32	1.23	1.28	1.26	1.29	1.25	1.35	1.24	1.21
2018/12	1.36	1.24	1.26	1.18	1.29	1.28	1.23	1.18	1.34	1.25	1.28	1.27	1.29	1.24	1.37	1.23	1.20
2019/03	1.36	1.25	1.26	1.19	1.31	1.31	1.25	1.16	1.34	1.27	1.31	1.27	1.29	1.27	1.37	1.24	1.22
2019/06	1.36	1.26	1.26	1.19	1.31	1.32	1.26	1.16	1.35	1.28	1.28	1.27	1.31	1.28	1.38	1.28	1.23
2019/09	1.39	1.26	1.26	1.20	1.32	1.36	1.25	1.17	1.37	1.29	1.29	1.28	1.31	1.30	1.39	1.28	1.24
2019/12	1.42	1.30	1.24	1.20	1.33	1.38	1.26	1.23	1.40	1.30	1.30	1.30	1.33	1.29	1.40	1.30	1.22
2020/03	1.40	1.32	1.26	1.26	1.33	1.37	1.27	1.24	1.43	1.31	1.30	1.32	1.37	1.28	1.40	1.32	1.24
2020/06	1.38	1.32	1.26	1.26	1.34	1.35	1.26	1.20	1.41	1.31	1.36	1.33	1.36	1.29	1.42	1.33	1.23
2020/09	1.38	1.34	1.28	1.26	1.35	1.36	1.32	1.18	1.42	1.33	1.37	1.35	1.36	1.32	1.44	1.34	1.22
2020/12	1.37	1.46	1.32	1.27	1.39	1.34	1.38	1.19	1.39	1.32	1.38	1.36	1.38	1.32	1.44	1.37	1.24
2021/03	1.35	1.46	1.30	1.27	1.40	1.40	1.39	1.17	1.39	1.32	1.40	1.38	1.43	1.34	1.47	1.39	1.27
2021/06	1.35	1.48	1.29	1.24	1.41	1.36	1.37	1.15	1.37	1.32	1.44	1.39	1.43	1.33	1.47	1.37	1.30
2021/09	1.34	1.49	1.29	1.24	1.42	1.35	1.41	1.15	1.34	1.36	1.43	1.39	1.44	1.35	1.46	1.41	1.33
2021/12	1.35	1.46	1.29	1.24	1.40	1.37	1.37	1.16	1.32	1.36	1.38	1.39	1.46	1.38	1.44	1.43	1.32
2022/03	1.34	1.44	1.29	1.21	1.37	1.34	1.37	1.16	1.30	1.37	1.38	1.36	1.44	1.39	1.45	1.43	1.36
2022/06	1.34	1.43	1.28	1.18	1.37	1.32	1.36	1.14	1.30	1.36	1.37	1.39	1.43	1.40	1.44	1.40	1.33
2022/09	1.35	1.40	1.28	1.19	1.37	1.28	1.36	1.12	1.31	1.34	1.34	1.36	1.40	1.41	1.42	1.38	1.33
2022/12	1.35	1.36	1.28	1.18	1.36	1.26	1.32	1.13	1.31	1.31	1.33	1.35	1.41	1.38	1.41	1.36	1.26

〈표 14〉 계속

시점	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2023/03	1.35	1.36	1.30	1.20	1.36	1.28	1.33	1.16	1.34	1.31	1.31	1.36	1.37	1.38	1.39	1.34	1.31
2023/06	1.36	1.37	1.34	1.21	1.35	1.28	1.35	1.16	1.35	1.31	1.32	1.35	1.40	1.37	1.40	1.34	1.28
2023/09	1.38	1.40	1.36	1.24	1.34	1.29	1.36	1.15	1.35	1.32	1.32	1.36	1.40	1.39	1.38	1.35	1.28
2023/12	1.36	1.40	1.36	1.24	1.36	1.31	1.36	1.15	1.34	1.32	1.32	1.37	1.40	1.38	1.41	1.37	1.29
2024/03	1.37	1.38	1.36	1.24	1.36	1.28	1.36	1.15	1.35	1.32	1.35	1.36	1.40	1.37	1.41	1.38	1.33

주: 소수점 세자리 이하는 반올림되었다.

한편, 본 연구의 가격 지니계수 추정치는 동일한 분석 도구를 활용한 선행 연구의 결과와도 그 위치를 비교해 둘 수 있다. 서재원·진장익(2021)은 2011~2019년 국토교통부 실거래가 자료를 활용하여 서울 25개 자치구별 주택가격 지니계수를 산출하고, 이를 계층이동 가능성에 대한 인식과 결합한 분석을 수행하였다. 본 연구의 V.3절은 동일한 도구(주택가격 지니계수)를 (i) 자치구가 아닌 17개 시도 단위로 확장하고, (ii) 분석 기간을 2012~2024년의 13년 시계열로 연장하며, (iii) 거래 표본이 아닌 전수 추정 표본에 적용한다는 점에서 선행 연구의 자치구 단위·단일 시점 분석을 일관된 방식으로 일반화한다. 자치구라는 작은 공간 단위 내부의 가격 분산만을 반영하는 선행 연구의 지니계수와, 시도 안의 모든 단지·평형을 포함하는 본 연구의 지니계수는 측정 대상이 구조적으로 다르므로 절대 수준의 직접 비교는 적절하지 않지만, 두 결과는 모두 분석 기간 동안 주택가격 분포의 불균등성이 확대되어 왔다는 일관된 결론을 공유한다.

4. 분석적 활용 예시 2: 아파트 가격 부담(Price to Income Ratio)의 산출

앞서 확인했듯이, 우리나라 아파트 시장에서는 지난 10여 년간 전반적인 가격 상승과 함께 지역 간 및 지역 내 격차의 확대가 동시에 나타났다. 그러나 주택 가격의 수준을 확인한 것만으로는 그것이 시장의 여력 대비 얼마나 높은 수준인지를 구체적으로 파악하기는 어려울 것이다. 이런 경우, 가계의 소득 대비 주택 가격의 상대적 비율인 PIR은 매우 유용한 정보가 될 수 있다. PIR은 일정 지역의 평균 또는 중위 주택가격을 해당 지역 가구의 연평균 소득으로 나눈 값으로, 가계의 지불 여력 대비 주택가격(즉, 평균적 가구가 자기 지역의 평균적 주택을 구매하기 위해 몇 년치 소득을 모아야 하는지)을 나타낸다.

PIR에 대해 일정 수준 이상 과학적으로 추정하려면 기본적으로 전체 분포를 고려한 평균 주택 가격(또는 중위가격)과 소득 수준의 추정이 필요하므로, 지금까지는 이에 대한 추정이 다소 산발적으로 이루어져 왔다. 하지만 개별 주택에 대한 가격 추정을 하는 경우, PIR을 광범위하게 도출하기 용이할 뿐 아니라 그것이 시계열적으로 어떻게 변화했는지 또는 지역 내에서 PIR의 분포가 어떻게 되었는지 등을 보다 쉽게 관측할 수 있다. 이에 본 절에서는 각 지역의 아파트 시장 PIR의 추이를 먼저 살펴보고, 마지막으로 아파트 가격 분포를 고려한 PIR의 분포를 살펴보려 한다.⁷⁾

PIR의 계산을 위해서는 평균적 가구 크기와 주택 크기, 소득 수준에 대한 가정이 필요하다. 본 연구에서는 평균적 가구 크기에 대한 보정을 위해 2023년 인구 총조사 기준 평균 가구원수(2.2명)를 적용하였으며, 평균적 주택 크기로는 전국 아파트 공급면적의 평균(94.6m²)을 사용하였다. 또한, 1인당 소득 수준은 통계청의 지역별 개인소득을 적용하였다. 본 연구에서는 이를 통해 지역별 평균 가구 소득과 평균 크기의 아파트 가격과의 비율을 도출하였다.⁸⁾ 이때 소득은 평균 기반·세전 정의를 따른다. 즉, 분자에는 본 연구가 추정한 지역별 m²당 평균 가격에 평균 아파트 면적을 곱한 평균 주택가격을, 분모에는 통계청 지역소득의 1인당 개인소득(세전 기준)에 평균 가구원수를 곱한 평균 가구소득을 사용한다. 이러한 산정 방식은 World Bank Housing Indicators Programme의 표준 PIR 정의 및 이창무 외(2012)가 국제비교에 적용한 평균 기반 정의와 일치한다.

〈표 15〉는 전국 및 각 지역의 아파트 가격 PIR 추이를 보여준다. 전국 단위의 PIR은 2015년 6.14에서 시작해, 2019년까지 6.19로 완만한 상승세를 보이다가 2020년과 2021년을 거치면서 9.40으로 급등하였다. 이는 앞서 확인할 수 있었던 전국적인 주택 가격 급등기와 맞물려 있다. 즉, 이 같은 추이는 2020년에서 2021년 사이 주택가격이 소득에 비해 지나치게 빠르게 상승했다는 사실을 명확히 보여준다. 이후 2022년과 2023년에는 각각 8.77, 7.91로 소폭 하락했지만, 이는 2020년 이전에 비하여 여전히 뚜렷이 높은 수준이다. PIR의 절대적 수준에 대한 해석은 제반 경제 환경에 따라 다를 수 있으나 일반적으로 9 이상의 PIR이 대단히 낮은 주택 접근성을 의미한다는 점에서 이는 우리나라의 아파트 가격이 가구소득

7) 본 연구에서 제시하는 PIR은 아파트만을 대상으로 하므로, 시장 전체의 주택 부담 수준을 대표하는 지표로 해석될 수는 없을 것이다. 또한, PIR의 산출은 기준 소득에 따라 달라질 수 있으므로, 본 연구에서 제시되는 결과는 하나의 추정치로 제시된다. 그럼에도 본 절의 결과는 기계학습 기반 전수 추정 자료가 PIR과 같은 시장 부담 지표의 산출로 어떻게 용이하게 확장될 수 있는지에 대한 예시로 의미를 가질 수 있을 것이다.

8) 본 연구는 평균 가구원수와 평균 공급면적에 대해 전국 평균값(각각 2.2명, 94.6m²)을 모든 지역에 동일하게 적용하였다. 이는 본 PIR이 측정하고자 하는 바를 "전국 평균에 해당하는 표준적 가구가 지역별 평균 아파트를 구매할 경우의 부담"으로 명확히 정의하기 위함이다. 만약 가구원수와 평균 면적을 지역별로 달리 적용할 경우, 산출된 PIR의 지역 간 차이에는 가격·소득의 격차에서 비롯되는 부분과 가구·주택 구조의 지역별 차이에서 비롯되는 부분이 동시에 혼재되어 지표의 의미가 모호해지는 측면이 있다. 본 연구는 가격·소득의 차이에 초점을 맞추어 비교를 단순화하기 위해 인구·구조적 변수는 고정하는 방식을 채택하였다.

〈표 15〉 아파트 가격 분위수 비율 95~50(단위: 배)

시점	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2012/03	2.15	1.44	1.47	1.51	1.41	1.45	1.47	1.12	1.69	1.49	1.56	1.49	1.59	1.51	1.51	1.59	1.68
2012/06	2.13	1.45	1.49	1.57	1.40	1.47	1.45	1.10	1.69	1.47	1.54	1.53	1.59	1.50	1.53	1.62	1.67
2012/09	2.09	1.45	1.46	1.56	1.39	1.49	1.44	1.09	1.64	1.44	1.52	1.52	1.60	1.49	1.49	1.61	1.63
2012/12	2.10	1.44	1.46	1.58	1.39	1.48	1.46	1.11	1.66	1.45	1.55	1.49	1.60	1.53	1.45	1.57	1.67
2013/03	2.11	1.46	1.43	1.61	1.39	1.45	1.42	1.13	1.64	1.43	1.52	1.48	1.60	1.49	1.47	1.61	1.65
2013/06	2.11	1.46	1.42	1.59	1.42	1.45	1.43	1.12	1.65	1.44	1.52	1.49	1.62	1.51	1.46	1.63	1.65
2013/09	2.12	1.46	1.39	1.59	1.42	1.45	1.42	1.13	1.66	1.42	1.48	1.47	1.61	1.49	1.46	1.58	1.59
2013/12	2.10	1.46	1.38	1.60	1.41	1.46	1.42	1.13	1.65	1.45	1.48	1.48	1.60	1.52	1.46	1.60	1.56
2014/03	2.14	1.45	1.41	1.56	1.39	1.45	1.43	1.24	1.68	1.46	1.52	1.47	1.62	1.56	1.47	1.58	1.63
2014/06	2.14	1.47	1.43	1.53	1.42	1.48	1.43	1.18	1.71	1.46	1.56	1.48	1.63	1.59	1.47	1.60	1.59
2014/09	2.16	1.48	1.42	1.58	1.40	1.47	1.46	1.17	1.74	1.44	1.60	1.47	1.63	1.57	1.49	1.60	1.60
2014/12	2.16	1.52	1.45	1.60	1.47	1.47	1.43	1.16	1.75	1.44	1.62	1.45	1.64	1.60	1.50	1.58	1.57
2015/03	2.20	1.51	1.52	1.54	1.47	1.52	1.46	1.15	1.78	1.43	1.59	1.44	1.64	1.59	1.50	1.60	1.70
2015/06	2.23	1.55	1.52	1.52	1.46	1.52	1.47	1.20	1.77	1.42	1.48	1.46	1.69	1.65	1.53	1.59	1.79
2015/09	2.19	1.55	1.54	1.50	1.48	1.53	1.48	1.20	1.76	1.43	1.48	1.48	1.68	1.62	1.51	1.55	1.78
2015/12	2.21	1.56	1.57	1.48	1.45	1.52	1.47	1.20	1.76	1.41	1.49	1.45	1.67	1.63	1.53	1.55	1.73
2016/03	2.17	1.59	1.54	1.50	1.47	1.55	1.48	1.25	1.75	1.45	1.47	1.47	1.67	1.65	1.51	1.56	1.77
2016/06	2.22	1.59	1.59	1.48	1.48	1.55	1.45	1.28	1.76	1.47	1.50	1.48	1.68	1.66	1.52	1.54	1.73
2016/09	2.31	1.62	1.60	1.46	1.47	1.57	1.45	1.22	1.77	1.47	1.45	1.49	1.69	1.64	1.53	1.54	1.72

〈표 15〉 계속

시점	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2016/12	2.25	1.68	1.61	1.49	1.48	1.54	1.46	1.22	1.77	1.49	1.47	1.53	1.72	1.64	1.54	1.55	1.72
2017/03	2.30	1.68	1.63	1.49	1.50	1.56	1.47	1.24	1.77	1.46	1.46	1.53	1.72	1.67	1.57	1.51	1.62
2017/06	2.33	1.71	1.62	1.50	1.52	1.58	1.48	1.28	1.83	1.44	1.50	1.53	1.71	1.67	1.60	1.52	1.66
2017/09	2.37	1.73	1.67	1.49	1.51	1.60	1.50	1.29	1.87	1.41	1.53	1.56	1.69	1.79	1.64	1.53	1.67
2017/12	2.49	1.71	1.72	1.49	1.54	1.62	1.50	1.31	1.91	1.44	1.55	1.54	1.70	1.80	1.62	1.51	1.68
2018/03	2.60	1.75	1.77	1.50	1.65	1.64	1.50	1.30	2.10	1.46	1.56	1.59	1.73	1.78	1.67	1.57	1.67
2018/06	2.55	1.76	1.82	1.54	1.74	1.69	1.53	1.36	2.17	1.48	1.57	1.62	1.72	1.77	1.71	1.66	1.73
2018/09	2.54	1.82	1.88	1.54	1.84	1.76	1.59	1.44	2.28	1.54	1.66	1.65	1.78	1.76	1.76	1.68	1.74
2018/12	2.39	1.81	1.89	1.56	1.83	2.01	1.59	1.43	2.22	1.58	1.65	1.68	1.80	1.77	1.77	1.69	1.73
2019/03	2.33	1.84	1.94	1.58	1.85	2.05	1.66	1.39	2.19	1.65	1.73	1.76	1.87	1.84	1.78	1.73	1.73
2019/06	2.41	1.87	1.93	1.62	1.84	2.16	1.64	1.40	2.21	1.69	1.76	1.75	1.94	1.85	1.80	1.76	1.74
2019/09	2.51	1.88	1.91	1.63	1.85	2.00	1.67	1.46	2.30	1.72	1.79	1.77	1.95	1.89	1.82	1.78	1.78
2019/12	2.59	2.06	1.88	1.71	1.88	2.03	1.69	1.55	2.43	1.79	1.86	1.92	1.99	1.88	1.89	1.84	1.77
2020/03	2.41	2.21	1.91	1.86	1.89	2.02	1.81	1.48	2.40	1.82	1.87	1.96	2.23	1.97	1.95	1.88	1.78
2020/06	2.35	2.25	1.90	1.80	1.91	1.92	1.85	1.45	2.29	1.87	2.02	2.01	2.26	2.00	1.95	1.93	1.80
2020/09	2.36	2.43	1.96	1.79	1.94	1.99	1.98	1.38	2.38	1.91	2.09	2.10	2.43	2.15	1.96	2.02	1.78
2020/12	2.28	2.52	2.19	1.73	2.09	2.06	2.19	1.39	2.31	1.98	2.15	2.18	2.61	2.15	2.10	2.19	1.85
2021/03	2.27	2.47	2.06	1.83	2.17	2.07	2.12	1.35	2.29	2.08	2.39	2.36	2.74	2.23	2.21	2.30	1.96
2021/06	2.25	2.48	2.03	1.72	2.13	1.90	2.02	1.36	2.20	2.11	2.27	2.38	2.66	2.18	2.21	2.25	2.02
2021/09	2.26	2.46	2.03	1.79	2.09	1.82	2.08	1.34	2.16	2.18	2.19	2.23	2.65	2.20	2.20	2.29	2.19

〈표 15〉 계속

시점	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2021/12	2.32	2.39	2.04	1.76	2.06	1.81	2.00	1.38	2.08	2.14	2.13	2.17	2.60	2.21	2.18	2.27	2.15
2022/03	2.35	2.30	2.04	1.74	2.02	1.79	1.92	1.37	2.08	2.12	2.06	2.07	2.61	2.20	2.13	2.26	2.08
2022/06	2.37	2.33	1.97	1.75	1.99	1.83	1.87	1.42	2.07	2.06	2.03	2.03	2.58	2.17	2.14	2.20	1.99
2022/09	2.42	2.20	1.94	1.63	1.92	1.78	1.85	1.43	2.06	1.98	1.93	1.96	2.38	2.12	2.03	2.13	1.94
2022/12	2.38	2.21	2.01	1.60	1.88	1.74	1.79	1.42	2.05	1.96	1.89	1.94	2.31	1.99	1.99	2.02	1.89
2023/03	2.42	2.17	2.01	1.68	1.91	1.80	1.83	1.43	2.12	1.95	1.86	1.96	2.25	2.03	1.94	1.97	1.91
2023/06	2.48	2.21	2.05	1.73	1.90	1.83	1.89	1.44	2.16	1.93	1.94	1.98	2.38	2.07	2.01	2.00	1.89
2023/09	2.51	2.23	2.09	1.75	1.94	1.97	1.93	1.42	2.19	1.92	2.01	2.01	2.46	2.14	2.00	2.04	1.86
2023/12	2.50	2.26	2.08	1.72	1.98	1.82	1.89	1.43	2.16	1.95	2.00	2.04	2.45	2.13	2.05	2.05	1.87
2024/03	2.51	2.29	2.10	1.72	1.94	1.80	1.95	1.45	2.17	1.93	2.02	2.05	2.48	2.12	2.09	2.05	1.89

주: 소수점 세자리 이하는 반올림되었다.

에 비해 평균적으로 상당히 높은 수준이라는 것을 의미할 수 있다.⁹⁾ 물론 이는 서울을 포함한 평균 가격으로부터 도출된 것이며 서울과 기타 지역의 격차가 상당하다는 것을 감안해야 할 것이다.

지역별로 보면, 가장 수준이 높고 해당기간 상승을 보여준 지역은 단연 서울이다. 2015년 서울의 PIR은 9.81로 이미 전국 평균을 크게 상회하고 있었는데, 이후 2018년에는 13.36, 2020년 16.64, 그리고 2021년에는 무려 19.53까지 치솟았다. 이는 평균적 가구 소득을 전혀 지출하지 않고 약 20년 치를 모아야 평균 수준의 아파트를 구매할 수 있다는 것을 의미하여, 가계의 구매력 대비 아파트 가격의 괴리가 서울에서 얼마나 심화되었는지를 구체적으로 보여준다. 2022년과 2023년에는 PIR 수치가 각각 18.35, 16.34로 다소 완화되었지만, 절대 수준은 여전히 전국에서 가장 높은 상태이다.

9) PIR의 절대적 수준에 대한 해석은 제반 경제 환경에 따라 다를 수 있으나, Demographia International Housing Affordability 보고서(Cox, 2024)의 분류에 따르면 PIR 9 이상은 "Impossibly Unaffordable"에 해당하는 가장 심각한 주택 접근성 단계로 분류된다. 다만 동 분류 체계는 중위 기반 PIR을 전제로 하므로, 하나의 참고 기준으로 이해되어야 한다.

경기도와 인천 역시 수도권 일원으로 PIR이 함께 상승하는 경향을 보였다. 경기도의 경우 2015년 6.59에서 시작해 2021년에는 10.46까지 상승하였다. 이는 서울보다는 낮지만 전국 평균보다는 높은 수치이며, 수도권 내 비서울 지역 역시 주택가격 부담이 심화되었음을 시사한다. 인천 또한 같은 시기 6.12에서 8.24로 상승한 후 최근에는 6.80으로 다소 조정된 모습을 보인다. 지방 대도시의 PIR은 수도권에 비해서는 상대적으로 낮은 수준에서 유지되었으나, 정점인 2021년에는 부산(8.02), 대구(7.31), 세종(7.58), 제주(7.58)와 같이 상당한 수준에 이르기도 했다. 한편, 충청, 호남, 영남권 대부분의 비수도권 지역은 수도권, 그리고 지방 대도시에 비해서도 PIR 수준이 확연히 낮았음을 확인할 수 있다. 예컨대 충북의 PIR은 2021년에도 5.80, 전남은 5.37, 경북은 5.06 등으로 유지되었다. 그러므로 PIR은 아파트 가격 자체에 비해서도 더 큰 지역간 차이를 보여준다.

이를 동일 시기에 대해 다른 산정방식을 택한 PIR 추정결과와 비교할 수 있다. 국토교통부 주거실태 조사는 자가가구의 중위가격을 중위가구소득으로 나누는 방식으로 PIR을 산출하며, 2022년 기준 전국 6.3배, 수도권 9.3배, 광역시 등 6.8배, 도지역 4.3배로 보고된다(국토교통부, 2023). KB국민은행은 주택가격과 가계소득을 각각 5분위로 분할한 5×5 PIR 매트릭스를 산출하며, 그 중 3분위(중위) 가격을 3분위(중위) 소득으로 나눈 대표 수치로서 2020년 3월 기준 전국 5.22배를 보고하였다(KB국민은행, 2020). 한국주택금융공사는 자사 주택담보대출 이용자를 대상으로 하여 2020년 1분기 기준 전국 4.45배, 서울 5.22배의 PIR을 산출하였다(한국주택금융공사, 2020).

이러한 수준 격차는 본 연구의 추정 자체의 편향이 아니라 산정방식의 구조적 차이로 설명된다. 첫째, 본 연구는 평균 기반 PIR을 채택하였는데, 평균은 분포 상단의 고가 자산에 의해 끌어올려지므로 동일 시장에 대해서도 중위 기반 PIR보다 큰 값으로 산출된다. 이는 Cox(2024)가 평균 기반 PIR이 분포 상단에 의해 왜곡되기 쉽기 때문에 정책 평가용으로는 중위 기반 median multiple을 권장한다고 지적한 점과도 부합한다. 둘째, 본 연구는 분석 대상을 아파트로 한정하였는데, 우리나라에서 아파트는 비아파트 주택 유형(다세대·연립·단독)에 비해 상대적으로 가격대가 높으므로 주거실태 조사의 전체 자가주택 기준 PIR보다 큰 값이 나타난다. 셋째, 본 연구의 서울 PIR이 특히 높게 나타나는 것은 서울시 시도 단위 비교에서 가장 가격대가 높을 뿐 아니라 분포 상단이 빠르게 팽창해 평균이 더 큰 폭으로 끌어올려지는 상단 팽창형 분포를 갖고 있기 때문이다.

이러한 차이를 통제하고 시기적 추세의 정합성을 살펴보면, 본 연구의 PIR과 공표 PIR은 매우 일관된 함의를 공유한다. 주거실태조사 기준 수도권 PIR은 2021년 9.5배에서 2022년 9.3배, 2023년 8.5배로 조정 패턴을 보이며(국토교통부, 2024), 이는 본 연구가 추정한 수도권 PIR의 정점 시점(2021년) 및 이후 조정 추이와 일치한다. 즉 본 연구의 PIR은 절대 수준의 측면에서는 산정방식의 차이로 인해 공표 수치보다 높게 산출되지만, 변동의 방향·시점·지역 간 순서에 있어서는 공표 수치와 일관된 결과를 제공한다(〈표 16〉, 〈표 17〉).

〈표 16〉 단위면적당 가격의 지니계수(단지·평형 기준, 세대수 가중)

시점	전국	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2012/03	0.282	0.204	0.137	0.141	0.119	0.125	0.151	0.148	0.115	0.174	0.168	0.182	0.170	0.182	0.158	0.178	0.174	0.157
2012/06	0.278	0.205	0.135	0.139	0.126	0.123	0.152	0.143	0.112	0.172	0.165	0.180	0.169	0.177	0.153	0.177	0.174	0.168
2012/09	0.272	0.201	0.136	0.134	0.128	0.122	0.153	0.140	0.112	0.167	0.163	0.178	0.167	0.174	0.154	0.173	0.172	0.153
2012/12	0.271	0.202	0.135	0.140	0.126	0.124	0.154	0.144	0.107	0.167	0.166	0.179	0.169	0.179	0.153	0.174	0.169	0.159
2013/03	0.268	0.204	0.137	0.133	0.129	0.123	0.154	0.144	0.117	0.165	0.161	0.174	0.165	0.175	0.151	0.170	0.171	0.162
2013/06	0.270	0.205	0.138	0.127	0.125	0.127	0.152	0.145	0.115	0.170	0.161	0.173	0.165	0.176	0.151	0.170	0.175	0.159
2013/09	0.266	0.203	0.139	0.120	0.129	0.125	0.151	0.143	0.113	0.171	0.159	0.167	0.164	0.178	0.152	0.171	0.172	0.156
2013/12	0.267	0.203	0.141	0.117	0.135	0.126	0.153	0.140	0.119	0.171	0.160	0.169	0.163	0.179	0.152	0.174	0.175	0.149
2014/03	0.268	0.208	0.142	0.120	0.128	0.124	0.150	0.139	0.134	0.173	0.160	0.171	0.164	0.182	0.159	0.175	0.176	0.154
2014/06	0.268	0.208	0.144	0.120	0.127	0.126	0.148	0.141	0.123	0.176	0.160	0.178	0.167	0.183	0.165	0.177	0.178	0.150
2014/09	0.269	0.212	0.145	0.120	0.128	0.126	0.147	0.142	0.126	0.179	0.156	0.180	0.168	0.182	0.166	0.179	0.180	0.150
2014/12	0.269	0.212	0.149	0.124	0.131	0.134	0.149	0.140	0.125	0.181	0.160	0.185	0.166	0.184	0.166	0.181	0.181	0.152
2015/03	0.270	0.214	0.151	0.132	0.127	0.135	0.151	0.142	0.091	0.183	0.156	0.181	0.165	0.186	0.169	0.182	0.181	0.159
2015/06	0.273	0.215	0.157	0.139	0.122	0.135	0.153	0.143	0.099	0.185	0.161	0.177	0.168	0.190	0.173	0.185	0.181	0.167
2015/09	0.274	0.213	0.162	0.144	0.124	0.138	0.154	0.143	0.100	0.184	0.161	0.175	0.169	0.192	0.174	0.186	0.179	0.170
2015/12	0.277	0.216	0.164	0.143	0.121	0.135	0.155	0.143	0.105	0.184	0.162	0.173	0.170	0.193	0.179	0.187	0.178	0.170
2016/03	0.276	0.214	0.167	0.138	0.125	0.134	0.156	0.145	0.117	0.182	0.162	0.173	0.171	0.195	0.182	0.184	0.176	0.175
2016/06	0.281	0.219	0.168	0.141	0.121	0.132	0.157	0.145	0.121	0.184	0.163	0.174	0.171	0.197	0.186	0.186	0.174	0.163
2016/09	0.289	0.226	0.174	0.144	0.120	0.136	0.158	0.145	0.156	0.186	0.166	0.174	0.174	0.199	0.186	0.191	0.175	0.167
2016/12	0.291	0.224	0.183	0.152	0.125	0.134	0.159	0.149	0.158	0.188	0.172	0.173	0.175	0.201	0.191	0.193	0.174	0.164
2017/03	0.293	0.227	0.183	0.155	0.127	0.140	0.160	0.149	0.126	0.188	0.171	0.170	0.176	0.202	0.194	0.199	0.173	0.164

〈표 16〉 계속

시점	전국	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2017/06	0.301	0.234	0.185	0.156	0.125	0.143	0.162	0.150	0.133	0.193	0.171	0.177	0.178	0.204	0.195	0.206	0.172	0.166
2017/09	0.308	0.235	0.183	0.166	0.128	0.146	0.164	0.154	0.137	0.200	0.170	0.183	0.181	0.203	0.202	0.210	0.172	0.170
2017/12	0.318	0.248	0.183	0.173	0.128	0.150	0.168	0.157	0.167	0.204	0.175	0.185	0.183	0.208	0.207	0.214	0.173	0.167
2018/03	0.338	0.267	0.185	0.181	0.128	0.161	0.174	0.163	0.147	0.221	0.174	0.188	0.188	0.211	0.205	0.222	0.180	0.173
2018/06	0.345	0.263	0.188	0.190	0.134	0.176	0.179	0.170	0.161	0.230	0.176	0.190	0.194	0.213	0.205	0.228	0.189	0.172
2018/09	0.367	0.266	0.192	0.199	0.136	0.189	0.191	0.175	0.169	0.247	0.183	0.198	0.198	0.217	0.207	0.234	0.192	0.174
2018/12	0.366	0.252	0.194	0.201	0.137	0.195	0.212	0.180	0.169	0.249	0.188	0.203	0.210	0.221	0.208	0.237	0.196	0.175
2019/03	0.365	0.247	0.199	0.207	0.143	0.201	0.219	0.193	0.175	0.247	0.195	0.209	0.216	0.225	0.213	0.242	0.203	0.175
2019/06	0.372	0.253	0.205	0.208	0.145	0.200	0.224	0.194	0.177	0.250	0.201	0.217	0.221	0.230	0.215	0.248	0.210	0.179
2019/09	0.385	0.263	0.209	0.206	0.149	0.201	0.228	0.199	0.187	0.261	0.203	0.221	0.226	0.236	0.220	0.253	0.215	0.182
2019/12	0.397	0.271	0.228	0.206	0.158	0.206	0.239	0.208	0.200	0.273	0.212	0.226	0.240	0.242	0.220	0.258	0.225	0.183
2020/03	0.404	0.260	0.242	0.209	0.179	0.207	0.245	0.222	0.205	0.282	0.218	0.235	0.252	0.258	0.228	0.266	0.234	0.188
2020/06	0.400	0.254	0.248	0.207	0.181	0.210	0.238	0.225	0.202	0.276	0.226	0.253	0.251	0.261	0.233	0.271	0.239	0.188
2020/09	0.415	0.249	0.265	0.216	0.181	0.213	0.248	0.242	0.188	0.286	0.232	0.258	0.267	0.272	0.245	0.272	0.251	0.187
2020/12	0.413	0.244	0.292	0.236	0.183	0.230	0.251	0.274	0.187	0.282	0.236	0.271	0.276	0.292	0.251	0.286	0.276	0.198
2021/03	0.414	0.243	0.294	0.224	0.193	0.240	0.255	0.272	0.177	0.280	0.253	0.291	0.292	0.297	0.257	0.293	0.283	0.220
2021/06	0.412	0.242	0.295	0.221	0.183	0.242	0.248	0.266	0.175	0.268	0.260	0.290	0.289	0.297	0.260	0.291	0.282	0.238
2021/09	0.410	0.241	0.295	0.217	0.187	0.245	0.242	0.269	0.166	0.259	0.265	0.288	0.285	0.299	0.257	0.292	0.287	0.253
2021/12	0.404	0.244	0.289	0.216	0.181	0.242	0.237	0.262	0.160	0.249	0.266	0.282	0.280	0.296	0.261	0.288	0.288	0.252
2022/03	0.400	0.244	0.282	0.210	0.170	0.237	0.227	0.256	0.156	0.245	0.267	0.275	0.274	0.292	0.258	0.285	0.283	0.250
2022/06	0.398	0.246	0.279	0.206	0.164	0.233	0.224	0.251	0.154	0.244	0.262	0.271	0.267	0.289	0.258	0.279	0.277	0.249
2022/09	0.391	0.248	0.267	0.203	0.157	0.223	0.216	0.244	0.153	0.237	0.253	0.258	0.257	0.280	0.250	0.272	0.271	0.234

〈표 16〉 계속

시점	전국	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
2022/12	0.385	0.245	0.262	0.206	0.150	0.218	0.211	0.236	0.159	0.238	0.247	0.247	0.249	0.273	0.240	0.264	0.255	0.230
2023/03	0.388	0.248	0.257	0.209	0.158	0.216	0.213	0.235	0.167	0.244	0.242	0.245	0.249	0.265	0.244	0.259	0.254	0.225
2023/06	0.393	0.254	0.260	0.215	0.167	0.219	0.214	0.241	0.169	0.250	0.241	0.249	0.251	0.270	0.245	0.263	0.256	0.225
2023/09	0.398	0.260	0.264	0.222	0.173	0.221	0.225	0.249	0.172	0.255	0.243	0.254	0.254	0.275	0.249	0.267	0.262	0.222
2023/12	0.396	0.260	0.264	0.222	0.174	0.220	0.224	0.249	0.170	0.251	0.242	0.254	0.254	0.277	0.251	0.264	0.263	0.215
2024/03	0.397	0.261	0.265	0.225	0.176	0.220	0.221	0.251	0.172	0.253	0.240	0.257	0.257	0.280	0.250	0.265	0.263	0.219

주: 소수점 네자리 이하는 반올림되었다.

〈표 17〉 지역별 PIR 추이(단위: 배)

	2015	2016	2017	2018	2019	2020	2021	2022	2023
전국	6.14	6.37	6.60	6.88	6.91	7.82	9.40	8.77	7.91
서울	9.81	10.41	11.38	13.36	14.02	16.64	19.53	18.35	16.34
부산	5.64	6.21	6.57	6.19	5.89	6.51	8.02	7.44	6.64
대구	6.08	5.85	5.92	6.10	6.08	6.40	7.31	6.34	5.65
인천	6.12	6.30	6.34	6.03	5.84	6.54	8.24	7.74	6.80
광주	4.46	4.59	4.59	4.73	4.77	4.77	5.33	5.28	4.88
대전	4.65	4.65	4.68	4.62	4.88	5.80	6.86	6.16	5.65
울산	4.53	4.79	4.72	4.37	4.07	4.52	5.24	4.92	4.46
세종	4.81	5.42	6.15	6.63	6.90	9.43	10.46	8.88	8.19
경기	6.59	6.79	6.92	7.08	7.09	8.24	10.48	9.69	8.67
강원	3.62	3.84	4.02	3.82	3.49	3.47	4.02	4.26	4.08
충북	4.35	4.22	4.00	3.70	3.36	3.55	4.15	4.10	3.83
충남	4.35	4.32	4.20	3.89	3.74	3.83	4.38	4.30	3.95
전북	3.58	3.67	3.69	3.51	3.30	3.32	3.81	3.79	3.50
전남	3.33	3.43	3.58	3.61	3.53	3.55	3.71	3.52	3.26
경북	4.07	3.93	3.77	3.51	3.27	3.28	3.71	3.62	3.41
경남	4.99	4.95	4.76	4.34	4.06	4.18	4.83	4.78	4.43
제주	5.68	6.87	7.21	6.99	6.63	6.48	7.58	7.76	7.36

주: 1) 평균적 크기의 아파트(공급면적 기준 94.6m²)와 가구원수(2.2명)가 가정되었다.

2) 소수점 2자리 이하는 반올림되었다.

PIR, price to income ratio.

VI. 결론

본 연구는 대표적 의사결정나무 기반 알고리즘인 XGBoost를 활용하여, 전국 아파트 단지(단지 내 개별 평형별)의 가격을 전수적으로 추정하고, 이를 통해 시장 전체의 가격 수준과 분포를 직접 관측하는 방법을 제시하였다(부록 표 1)). 본 연구는 기계학습 기반 AVM이 어떻게 전통적 가격지수와 다른 차원의 시장 정보(절대적 가격 수준, 분포 구조, 하위 집단별 통계량 등)에 직접 접근하는 도구로 활용될 수 있는지를 보여준다.

본 연구의 추정 결과를 요약하면 크게 다음과 같다. 첫째, 2024년 기준 전국 아파트의 평균 가격은 470만 원/m²이며 서울의 경우 1,121만 원 수준인 것으로 나타났다. 둘째, 전국의 아파트 가격은 2018년부터 2021년까지 큰 폭으로 상승하였다가 2022년 이후 약한 조정 상태로 전환하였다. 이때 조정폭은 서울 및 수도권에 비해 지방도시가 더 큰 것으로 나타났다. 셋째, 서울과 지역 간의 가격 격차는 해당 기간 동안 크게 상승하였다. 넷째, 지역 간 가격 격차 뿐 아니라 각 지역 내에서의 가격 격차 역시 전반적으로 상승하였다. 다섯째, PIR 역시 전반적으로 상승하였으며, 특히 서울 지역에서 두드러지게 높아졌음을(2021년 기준 19.53) 확인할 수 있다.

본 연구의 기여는 최근 빠르게 발전하는 기계학습 기법의 활용을 통해 시장 전체의 가격 분포를 직접 관측하는 경로를 제안하여 최초로 시연한 데에 있다. 이 접근은 거래 표본 의존성에서 자유로우며, 가격의 절대적 수준을 제공하고, 연구자가 원하는 기준에 따라 표본을 분할하여 다양한 하위 집단별 통계량을 자유롭게 생성할 수 있다는 점에서 기존의 거래 표본 기반 정보 체계와 보완적이다.

또한, 본 연구에서 제시된 추정 기반 재고 주택가격 관측 기법은 평균이나 분포, 관련 지표의 생성 뿐 아니라 패널 분석, 이층차분 설계, 표본 균형을 고려한 선형회귀, 편집 가능한 주택가격지수 등 다양한 분석으로 확장될 여지가 많다는 것을 고려하면, 추후 풍부한 후속 연구들이 이루어질 것을 기대할 수 있을 것으로 보인다. 예를 들어, 아파트 거래 자료를 통해 헤도닉 모형을 추정하는 경우, 크기가 크고 거래가 자주 일어나는 단지 또는 지역의 거래 표본이 더 많이 포함되어 있기 때문에 표본 불균형으로 인한 편향적 추정의 가능성에 노출되어 있다. 그러나, 전체 거래 자료가 아닌 분석하고자 하는 지역의 개별 단지에 대한 신뢰할 만한 추정치를 표본으로 사용하는 경우 그러한 표본 불균형의 가능성을 줄일 수 있을 것이다.

본 연구의 한계는 다음과 같다. 첫째, 본 연구는 특정한 하나의 가설보다는 시장의 전반적 상태를 관찰하고 재고 기반 주택가격 관측의 가능성을 제시하는 데에 초점이 맞춰져 있으므로, 그 자체로서는 분석적 해석을 제공하지 않는다는 한계가 있을 수 있다. 본 연구에서 관측된 가격 격차의 확대, PIR의 상승 등에 대해 본 연구는 해석의 단서를 제공할 뿐 인과적 결론을 제시하지는 않는다.

둘째, 본 연구에서는 전수적 추정을 통한 가격지수 산정 방법론 자체를 심도 깊게 다루지 않았다. 가격지수의 산정에는 시기별 재고 주택의 변화(멸실 및 신규 입주)에 대한 보정과 가중 방식 선택의 정당화라는 별도의 논의가 요구된다. 본 연구의 단지 자료는 분석 시점 기준의 단일 시점 정보로 고정되어 사용되었으므로, 분석 기간 중에 발생한 입지 속성의 변화(지하철 노선의 개통, 학교의 신설·폐교, 재건축 추진단계의 변화 등) 자체가 시장 가격에 미친 효과를 분리하여 관측하는 데에는 본 연구의 설계가 적합하지 않다. 향후 연구에서는 본 연구의 전수 추정치 위에 적절한 가중·보정 체계를 결합하여 정식 가격지수를 구성하는 방향으로 확장될 수 있다.

셋째, 본 연구가 활용한 기계학습 기반 AVIM 자체에도 잠재적 한계가 있을 수 있다. 거래 빈도가 잦은 주택에 대한 과적합 가능성, 급격한 가격 변동 시기의 예측력 저하, 학습 자료가 희소한 지역에서의 정확도 감소 등이 그 예이다. 본 연구의 IV장 평가 결과는 모든 시·도에서 모형이 평균적으로 거의 불편향에 가까운 정확성을 보임을 확인하였으나, 특정 시점·지역에서의 국지적 편향 가능성을 완전히 배제하지는 못한다.

넷째, 본 연구의 IV장 평가에서 관측된 17개 시·도 간 모형 성능의 격차는 추정 결과의 지역별 해석에 함의를 가지므로 별도의 논의가 필요하다. $\pm 10\%$ 허용 오차 기준의 적중률을 보면 수도권 3개 시·도와 대구·대전은 90% 이상에 위치하는 반면, 강원·충북·전북·전남 등 비수도권 도지역은 81%~85% 수준, 경북과 제주는 80%를 다소 하회하는 수준에 머문다. 이러한 지역 간 격차는 (i) 거래 빈도의 차이로 인한 학습 표본 밀도의 차이, (ii) 비수도권 일부 지역의 단지 규모 분포가 소규모 단지에 치우쳐 단지 간 이질성이 상대적으로 큰 점, (iii) 해당 지역의 주택 재고에서 아파트가 차지하는 비중이 상대적으로 작아 아파트 시장의 가격 형성 구조가 수도권·광역시 대비 표준화 정도가 낮을 가능성 등이 복합적으로 작용한 결과로 추정 가능하다. 다만 IV.2절에서 확인한 바와 같이 모든 지역에서 IAAO(2017)의 동질적 시장 권고 기준($COD \leq 10$)을 충족하며 평균백분오차(MPE)가 제주를 제외한 모든 지역에서 $\pm 1\%$ 이내에서 관리되는 점은, 지역별 평균이나 분포 등 집단 단위 통계량 산출에 대해서는 모든 지역에서 일정 수준 이상의 신뢰성이 확보됨을 의미한다. 따라서 본 연구의 추정 결과를 정책적 목적으로 활용할 때에는 수도권 및 광역시에 대해서는 비교적 정밀한 정량적 해석이 가능한 반면, 비수도권 도지역, 특히 경북·제주 등에 대해서는 결과 통계량의 방향성과 시계열적 추이는 신뢰할 수 있되 절대 수준의 미세한 차이에 대한 해석에는 보다 신중한 접근이 요구된다.

이상의 한계에도 불구하고, 본 연구는 기계학습 기반 AVIM의 활용 영역을 단순 가격 예측을 넘어 시장 전체의 가격 분포·불균등성·접근성에 대한 직접 관측으로 확장하였다는 점에서 의의를 갖는다. 이러한 확장은 전통적 가격지수가 제공하지 못하던 정보 차원을 학술 및 정책 영역에 새롭게 더해 줄 수 있다는 점에서 후속 연구의 토대로 활용될 수 있을 것이다. 향후에는 본 연구의 전수 추정치

위에 적절한 가중·보정 체계를 결합하여 정식 가격지수를 구성하는 방법론적 발전, 비아파트 주택 유형 및 보다 미시적 공간 단위로의 확장, 그리고 본 연구가 산출한 분포·격차·PIR 지표를 활용한 정책·인과 분석 등 다양한 방향의 후속 연구가 이어질 수 있을 것으로 기대된다.

참고문헌

- 국토교통부. (2023). *2022년도 주거실태조사 결과*. 세종: 국토교통부.
- 국토교통부. (2024). *2023년도 주거실태조사 결과*. 세종: 국토교통부.
- 권진웅, 김남정 (2020). 서울시 자치구별 아파트 Price-to-Income Ratio(PIR) 변동 특성에 관한 연구. *주택도시금융연구*, 5(2), 55-72.
- 김세울, 유선종. (2022). 교육환경과 아파트 전월세가격 상관성에 관한 실증연구: 서울시 송파구를 중심으로. *주택도시연구*, 12(1), 1-18.
- 김이환, 김형준, 류두진, 조훈. (2022). 기계학습 방법론을 활용한 아파트 매매가격지수 연구. *부동산 분석*, 8(3), 1-29.
- 김지원, 마강래. (2021). 거주지에 따른 자산 격차에 관한 연구: 주택 보유여부를 중심으로. *대한부동산학회지*, 39(2), 5-28.
- 배성완, 유정석. (2018). 머신 러닝 방법과 시계열 분석 모형을 이용한 부동산 가격지수 예측. *주택연구*, 26(1), 107-133.
- 서재원, 진장익. (2021). 주택가격의 불평등은 계층이동 가능성의 기대감을 낮추는가? *국토계획*, 56(7), 141-156.
- 송영호, 마강래. (2025). 거주지역과 자산 축적 간의 관계에 관한 연구: 주택 보유와 자산 분위에 따른 차이를 중심으로. *국토계획*, 60(1), 113-126.
- 이소영, 김정민. (2025). 머신러닝을 활용한 다세대주택 매매가격지수 산정에 관한 연구: 서울시 소지역 단위를 중심으로. *주택도시금융연구*, 16(2), 93-127.
- 이인호, 이경섭. (2020). 트리 기반 앙상블 방법을 활용한 자동 평가 모형 개발 및 평가: 서울특별시 주거용 아파트를 사례로. *한국데이터정보과학회지*, 31(2), 375-389.
- 이창무, 김현아, 조만. (2012). 소득대비 주택가격 비율(PIR)의 산정방식 및 그 수준에 대한 국제비교. *주택연구*, 20(4), 5-25.
- 정의철, 김진욱, 하두나. (2009). 부동산소득이 소득불평등에 미치는 영향 분석. *주택연구*, 17(2), 5-28.
- 통계청. (2018). *2017년 가계금융복지조사 결과*. 대전: 통계청.
- 통계청. (2022). *2022년 가계금융복지조사 결과*. 대전: 통계청.
- 한국주택금융공사. (2020). *2020년 2분기 주택금융 및 유동화증권 통계: 2020년 1분기 기준*. 부산: 한국주택금융공사.
- 허원제, 오경수, 신미정, 유동영. (2020). 수도권·비수도권 간의 주택가격 격차 요인 분해에 관한

- 연구. *한국지방세연구원 기본연구보고서*, 2019(6), 1-91.
- 홍정의. (2020). 기계학습 알고리즘을 이용한 주택가격감정 시스템의 구축 및 평가: XGBoost, LightGBM, CatBoost 알고리즘에 기반하여. *주택금융연구*, 4, 33-64.
- 홍정의. (2021). 랜덤 포레스트 알고리즘을 통한 주택 대량평가모형 연구. *부동산분석*, 7(1), 1-28.
- KB국민은행. (2020). *KB주택가격동향 보고서: 2020년 5월*. 서울: KB국민은행.
- Antipov, E. A., & 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.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Cox, W. (2024). *Demographia international housing affordability: 2024 edition*. Orange County, CA: Newgeography.
- Ekeland, I., Heckman, J. J., & Nesheim, L. (2004). Identification and estimation of hedonic models. *Journal of Political Economy*, 112(S1), S60-S109.
- Gallin, J., Molloy, R., Nielsen, E., Smith, P., & Sommer, K. (2021). Measuring aggregate housing wealth: New insights from machine learning. *Journal of Housing Economics*, 51, 101734.
- Gatzlaff, D. H., & Haurin, D. R. (1997). Sample selection bias and repeat-sales index estimates. *The Journal of Real Estate Finance and Economics*, 14, 33-50.
- Guliker, E., Folmer, E., & van Sinderen, M. (2022). Spatial determinants of real estate appraisals in the Netherlands: A machine learning approach. *ISPRS International Journal of Geo-Information*, 11(2), 125.
- Hill, R. J. (2013). *Hedonic price indexes for housing* (OECD Statistics Working Paper No. 2011/01). Paris, France: Organization for Economic Co-operation and Development.
- Ho, W. K. O., Tang, B. S., & Wong, S. W. (2021). Predicting property prices with machine learning algorithms. *Journal of Property Research*, 38(1), 48-70.
- Hong, J., Choi, H., & Kim, W. (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.
- Hong, J., & Kim, W. (2022). Combination of machine learning-based automatic valuation models for residential properties in South Korea. *International Journal*

- of Strategic Property Management*, 26(5), 362-384.
- International Association of Assessing Officers (IAAO). (2017). *Standard on ratio studies*. Kansas City, MO: IAAO.
- International Association of Assessing Officers (IAAO). (2018). *Standard on mass appraisal of real property*. Kansas City, MO: IAAO.
- Kang, S., & Koo, J. H. (2023). Exploring social capital level in regions with large and increasing wealth inequality: Lesson from Seoul, South Korea. *Social Indicators Research*, 168(1), 165-183.
- Kim, W., & Hong, J. (2024). Stacked ensemble model for the automatic valuation of residential properties in South Korea: A case study on Jeju Island. *Land*, 13(9), 1436.
- Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74(2), 132-157.
- Malpezzi, S. (2003). Hedonic pricing models: A selective and applied review. In T. O'Sullivan, & K. Gibb (Eds.), *Housing economics and public policy* (pp. 67-89). Oxford, UK: Blackwell Science.
- Melser, D. (2023). Selection bias in housing price indexes: The characteristics repeat sales approach. *Oxford Bulletin of Economics and Statistics*, 85(3), 623-637.
- Ming, Y., Zhang, J., Qi, J., Liao, T., Wang, M., & Zhang, L. (2020). Prediction and analysis of Chengdu housing rent based on XGBoost algorithm. In: *ICBDT 20: Proceedings of the 3rd International Conference on Big Data Technologies* (pp. 1-5). New York, NY: Association for Computing Machinery.
- Pace, R. K., & Hayunga, D. (2020). Examining the information content of residuals from hedonic and spatial models using trees and forests. *Journal of Real Estate Finance and Economics*, 60(1), 170-180.
- Park, B., & 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.
- Rico-Juan, J. R., & de La Paz, P. T. (2021). Machine learning with explainability or spatial hedonics tools? An analysis of the asking prices in the housing market in Alicante, Spain. *Expert Systems with Applications*, 171, 114590.

- Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1), 34-55.
- Sheppard, S. (1999). Hedonic analysis of housing markets. In P. Cheshire, & E. S. Mills (Eds.), *Handbook of regional and urban economics* (pp. 1595-1635). Amsterdam, The Netherlands: Elsevier.
- Tekouabou, S. C. K., Gherghina, Ș. C., Kameni, E. D., Filali, Y., & Idrissi Gartoumi, K. (2024). AI-based on machine learning methods for urban real estate prediction: A systematic survey. *Archives of Computational Methods in Engineering*, 31(2), 1079-1095.
- Yilmazer, S., & Kocaman, S. (2020). A mass appraisal assessment study using machine learning based on multiple regression and random forest. *Land Use Policy*, 99, 104889.

(논문 접수일: 2026.03.16. 수정논문 접수일: 2026.05.27. 논문 채택일: 2026.06.05.)

A Tree Ensemble Approach to Estimating Census-Level House Price Distributions

Jengei Hong*

Abstract

The sample includes 6,118,730 apartment transactions from March 2012 to April 2024 and complex- and floor-plan-level information on 12,517,131 housing units nationwide (99.1% of all apartment units in the Population and Housing Census). The study estimates quarterly prices for virtually the entire apartment stock in Korea. From these estimates, regional average prices, price distributions, and the price-to-income ratio (PIR) are derived. Our estimates show that average apartment prices rose more rapidly than the official housing price index during the 2017–2022 price boom, and that, over the analysis period, both cross-regional price gaps and within-region gaps across price quantiles widened simultaneously.

Keywords : XGBoost, Machine Learning, Automated Valuation Model (AVM), Housing Price Distribution, Price to Income Ratio (PIR)

* Jengei Hong, Corresponding author, Associate Professor, School of Management and Economics, Handong Global University, hwgh024@handong.edu

© Copyright 2026 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

〈부록 표 1〉 XGBoost 모형 하이퍼 파라미터 및 난수 설정

구분	파라미터	값	설명
하이퍼파라미터	n_estimators	400	부스팅 라운드 수(트리 개수)
	learning_rate	0.06	학습률(각 트리의 기여도)
	max_depth	20	개별 트리의 최대 깊이
	gamma	10	추가 분할에 필요한 최소 손실 감소
	reg_alpha	1.5	가중치에 대한 L1 정규화 계수
	reg_lambda	1.2	가중치에 대한 L2 정규화 계수
	colsample_bytree	0.5	각 트리 생성 시 사용하는 변수의 비율
난수 시드	seed	10	재현성 확보를 위한 난수 시드

주: 위에 명시되지 않은 파라미터는 XGBoost의 기본값(default)을 사용함.