creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

도시계획학 석사학위논문

부동산 비기초가격과 부동산 신문담론의 관계에 대한 시계열 분석

-서울 아파트 매매가를 중심으로-

Bridging the Gap: Using News Articles on Real Estate to Explain and Forecast Non-Fundamental Real Estate Prices

2023년 08월

서울대학교 대학원

환경계획학과 도시 및 지역계획학 전공

홍 지 수

부동산 비기초가격과 부동산 신문담론의 관계에 대한 시계열 분석

-서울 아파트 매매가를 중심으로-

지도교수 박 인 권

이 논문을 도시계획학 석사학위논문으로 제출함

2023년 06월

서울대학교 대학원

환경계획학과 도시 및 지역계획학 전공

홍 지 수

홍지수의 도시계획학 석사학위논문을 인준함

2023년 08월

위 원 장 김 경 민 (인)

부위원장 송 재 민 (인)

위 원 박 인 권 (인)

국 문 초 록

이론적으로, 시장가격은 해당 자산을 보유할 때 장래에 기대되는 수익의 현재가치, 즉 시장 기초가치(market fundamental value)로 결정되어야 한다. 그러나 미래의 기대 현금흐름에 대한 현재가치만으로 설명하기에는, 현실적으로 우리가 목도하는 시장가격은 상당히 변동적이고 확률적이다. 그 때문에 부동산 시장에서 객관적 요인만으로 설명되지 않는 주관적이고 심리적인 영역이 존재한다는 인식이 형성되어 있다.

이러한 맥락에서, 거시경제변수만을 이용해 부동산 시장을 분석하는 것의 한계를 극복하려는 노력의 일환으로, 신문기사나 방송뉴스와 같은 비정형 텍스트 데이터를 통해 부동산 시장의 변화를 분석하고 예측하는 연구가 시행되고 있다. 부동산 시장에서 심리적 요인의 중요성이 부상하면서, 부동산 시장 참여자의 심리와 경향에 관한 실증적 연구가 시행되고 있는 것이다. 그러나 부동산 시장 가격, 특히 부동산 기초가격(fundamental price)은 기존의 부동산 임대가격 변수, 거시경제변수 등으로 설명하는 것이 오히려 합리적이고 효율적일 수도 있다. 새롭게 비정형 텍스트 데이터를 발굴하고 구축하는 수고로움이 충분히 가치 있기 위해서는, 기존의 정형 데이터로는 설명이 어려웠던 영역이 새롭게 설명되는 더욱 본질적인 차별성을 지녀야 한다.

이에 본 연구는 정성적이고 변동적인 요소와 얽혀있어 분석과 예측이 어렵다고 알려진 부동산 비기초가격(non-fundamental price)을 대상으로 한다. 대중적 심리가 부동산 시장을 변동시키는 주요한 요인 중의 하나며, 언론매체가 그러한 집단적 인식에 상당한 영향을 준다는 이론적 논의에 기초하여, 부동산 신문기사 데이터를 활용해 부동산 비기초가격을 분석하고 예측한다. 이를 위해 서울을 배경으로 부동산 비기초가격 변수를 FM-OLS 모형을 토대로 추정하고 부동산 신문기사 변수를 Text Mining을 통해 구축한다.

부동산 비기초가격과 부동산 신문담론의 관계를 분석하기 위해 Granger 인과분석, VECM 모형 기반 충격반응분석 및 분산분해분석을 실시했다. 분석 결과, 부동산 비기초가격과 부동산 정책 관련 신문기사와는 부정의 관계가 가졌으며, 부동

산 광고와 생활환경에 관한 신문기사와는 긍정의 관계를 맺었다. 부동산에 대한 정책적 논의가 전개되면 부동산에 대한 과신이나 기대가 일부 수그러드는 것으로, 부동산의 투자가치나 사용 가치에 관한 정보가 확산되면 부동산에 대한 심리적 기대감이 형성되는 것으로 나타났다. 또한, 부동산 비기초가격은 수도권 지역에 관한 신문기사와 부정의 관계를 맺었다. 서울 인근의 수도권에 위치한 부동산 대장주격의 지역을 배경으로 하는 기사가 활발히 보도되면서 서울의 부동산이 지니는 상징성 내지 위치성이 일시적으로 위축될 가능성이 포착됐다. 이와 같은 결과는 부동산 비기초가격과 부동산 신문담론, 나아가 부동산 심리시장과 부동산 신문담론 사이에 유의미한 관계가 존재함을 내포한다.

다음으로, 부동산 비기초가격에 대한 부동산 신문기사의 예측 유용성을 실증하기 위해, ARIMA 분석을 시행했다. 부동산 신문기사 논조지수를 구축하고, 부동산 신문기사 논조지수가 부동산 비기초가격 예측에 유용한지 확인했다. 분석 결과, 부동산 신문기사 논조지수를 외생변수로 추가할 경우 부동산 비기초가격의 예측 오차가 10%가량 감소했다. 이를 통해 부동산 신문기사 논조지수가 부동산 비기초가격 변화 예측의 정확도를 향상함을 확인했다.

본 연구는 부동산 가격에 대한 이해도를 높이고자 부동산 비기초가격과 부동산 신문담론의 관계를 분석했다. 부동산 가격은 부동산 기초가격과 부동산 비기초가격으로 이루어져 있으며 부동산 비기초가격은 심리적 요인으로 설명할 수 있다는 문제의식에 착안하여, 부동산 신문기사를 활용해 부동산 비기초가격을 설명했다. 본 연구는 부동산 신문기사 빅데이터를 활용해 정책이나 정보와 같이 정량화하기 어려운 요인을 변수로 구축할 방안을 제시했다. 더불어, 부동산 신문기사를 활용하여 부동산 가격을 더욱 섬세하게 분석하고 예측할 가능성을 제시했다. 앞으로 새로운 데이터를 활용해 부동산 시장을 분석하는 연구가 계속해서 축적된다면, 부동산 가격에서 일종의 미지 영역으로 간주하던 부분이 점차 축소될 수 있을 것이다.

주요어 : 부동산 가격, 부동산 비기초가격, 신문기사, 부동산 신문기사, 빅데이터

학 번 : 2021-29070

목 차

제 1 장 서 론 ····· 1

제 1 절 연구의 배경 및 목적 ····· 1

제 2 절 연구의 범위 및 방법 ····· 3

제 3 절 연구의 구성 ····· 4

제 2 장 이론적 배경 및 선행연구 고찰 ····· 6

제 1 절 부동산 비기초가격 ····· 7

1) 부동산 기초가격과 비기초가격 ····· 7

2) 부동산 비기초가격의 추정 ····· 11

제 2 절 부동산 비기초가격과 부동산 언론보도 ····· 15

1) 부동산 언론보도 ····· 15

2) 부동산 언론보도의 논조 ····· 17

제 3 절 연구의 차별성 ····· 23

제 3 장 분석 자료 및 분석 방법 ····· 25

제 1 절 자료 구축 ····· 25

1) 부동산 비기초가격 ····· 25

2) 부동산 신문기사의 주제 ····· 29

3) 부동산 신문기사의 논조 ····· 34

제 2 절 분석 방법 ····· 37

1) Granger 인과관계 분석 ····· 37

2) VECM 모형 추정 ····· 38

3) ARIMA 예측 유용성 분석 ····· 40

제 4 장 부동산 신문담론의 형성과 변환 ····· 43

제 1 절 부동산 신문담론의 형성 ····· 43

1) 부동산 가격 ····· 44

2) 부동산 정책 ····· 45

3) 부동산 공급 ····· 48

4) 부동산 투자 ····· 48

5) 부동산 광고 ····· 49

6) 생활환경 ····· 49

제 2 절 부동산 신문담론의 변환 ····· 50

1) 자산 증식 수단이자 적극적 투기 대상으로서 부동산 ····· 52

2) 통제의 대상이자 사회적 병폐로서 부동산 ····· 53

제 5 장 부동산 비기초가격과 부동산 신문담론의 관계 ····· 54

제 1 절 분석 자료의 안정성 검정 ····· 54

제 2 절 Granger 인과관계 검정 ····· 55

제 3 절 공적분 검정 및 VECM 모형 추정 ····· 58

제 4 절 충격반응분석 및 분산분해분석 ····· 60
1) 충격반응분석 ····· 60
2) 분산분해분석 ····· 62
제 5 절 소결 ····· 63

제 6 장 부동산 신문논조를 통한 부동산 비기초가격 예측 ····· 66
제 1 절 분석 자료의 안정성 검정 ····· 66
제 2 절 Granger 인과관계 검정 ····· 67
제 3 절 예측 유용성 분석 ····· 68
1) ARIMA 분석 ····· 68
2) ARIMAX 분석 ····· 70
제 4 절 소결 ····· 72

제 7 장 결 론 ····· 74

참 고 문 헌 ····· 77
Abstract ····· 85

표 목 차

〈표 1〉 부동산 비기초가격 관련 주요 선행연구 ····· 14
〈표 2〉 부동산 신문기사 관련 주요 선행연구 ····· 22
〈표 3〉 FM-OLS 모형 변수 ····· 26
〈표 4〉 FM-OLS 모형 변수별 기초통계 ····· 26
〈표 5〉 Johansen 공적분 검정 결과 ····· 27
〈표 6〉 FM-OLS 회귀식 추정 결과 ····· 28
〈표 7〉 부동산 신문기사 자료 수집 방법 ····· 30
〈표 8〉 LDA 토픽 분석 결과 ····· 33
〈표 9〉 부동산 신문기사 자료의 영역별 기초통계 ····· 33
〈표 10〉 부동산 신문기사 감성사전 예시 ····· 35
〈표 11〉 부동산 신문기사 감성분석 결과 ····· 36
〈표 12〉 부동산 신문기사 논조지수와 관련 정형데이터 간 상관관계 분석 결과 ··· 37
〈표 13〉 부동산 신문기사 관심 순위 지도 ····· 51
〈표 14〉 ADF 단위근 검정 결과 ····· 55
〈표 15〉 Granger 인과관계 검정 결과 ····· 57
〈표 16〉 VAR 시차 선택 검정 결과 ····· 59
〈표 17〉 Johansen 공적분 검정 결과 ····· 60
〈표 18〉 부동산 비기초가격의 분산분해분석 결과 ····· 62
〈표 19〉 ADF 단위근 검정 결과 ····· 67

〈표 20〉 Granger 인과관계 검정 결과 ····· 68
〈표 21〉 부동산 비기초가격의 ARIMA(p,d,q) 분석 결과 ····· 69
〈표 22〉 ARIMA(1,1,3) 모형의 예측오차 ····· 70
〈표 23〉 부동산 비기초가격의 ARIMAX(p,d,q) 분석 결과 ····· 71
〈표 24〉 ARIMAX(2,1,3) Model모형의 예측오차 ····· 71
〈표 25〉 ARIMA(1,1,3) 모형과 ARIMAX(2,1,3) 모형의 예측오차 비교 ··· 72

그 림 목 차

[그림 1] 연구의 흐름도 ····· 5
[그림 2] 부동산 비기초가격(잔자항) 추정 결과 ····· 29
[그림 3] 부동산 신문기사 자료 수집 결과 ····· 30
[그림 4] LDA 토픽 모델링 알고리즘 구조 ····· 32
[그림 5] 부동산 신문기사 논조지수 추정 결과 ····· 37
[그림 6] 부동산 신문기사 영역별 추이 ····· 43
[그림 7] 부동산 신문기사 영역별·주제별 추이 ····· 46
[그림 8] 부동산 신문기사 주요 주제의 연간 변화량 ····· 52
[그림 9] 부동산 비기초가격과 부동산 신문기사 간 충격반응분석 결과 ····· 61

제 1 장 서 론

제 1 절 연구의 배경 및 목적

부동산 가격이 '고점'에 가까워졌다는 상투론, 부동산 가격이 '저점'이라는 바닥론은 오랫동안 우리나라 부동산 시장에서 반복적으로 회자돼 온 주제이다. 이러한 부동산 시장의 향방에 관한 여론은 맥락에 따라 부동산 가격에 대한 과열적 기대나 자기실현적 예언이 형성하기도 하며, 부동산 시장의 급격한 침체를 장기화하기도 한다. 다시 말해, 대중적 심리는 부동산 가격이 내재된 가치를 상회 혹은 하회하도록 유도하는 요인으로 작용한다(Shiller, 2000; Mercille, 2014; Bucchianeri, 2011).

이와 같은 현실적인 배경에서, 부동산 가격에는 현시된 효용 및 미래 기대수익과 같은 본질적 요소로 구성된 영역, 즉 기초가격(fundamental price) 영역으로 설명할 수 없는 부분이 존재한다는 인식이 형성되고 있다. 주류경제학적 가정에 입각해 부동산 시장가격을 기초가격만으로 추정할 경우 설명되지 않는 등락이 실존한다는 것이다. 이에 이성뿐 아니라 감성도 일정 정도 시장에 개입한다는 행동경제학적 관점 아래, 부동산 시장가격에서 기초가격으로 설명되지 않는 부분을 부동산 비기초가격(non-fundamental price) 영역이라 간주하고(Stiglitz, 1990) 이를 추정하는 연구가 시행되고 있다(Blanchard and Watson, 1982; Summers, 1986; Abraham and Hendershott, 1995; Levin and Wright, 1997; Muellbauer and Murphy. 1997; Shiller, 2000; Roche, 2001; Bucchianeri, 2011).

그러나, 부동산 비기초가격 영역의 실존 여부에 관한 연구가 꾸준히 진행되고 있음에도 불구하고, 부동산 비기초가격이 구체적으로 어떠한 심리적 요인에 의해 변동하는지 실증적으로 분석한 연구는 거의 부재하다. 이는 상식상 부동산 비기초가격과 유의미한 연관을 맺는다고 알려진 요인은 대부분 비정량적이기 때문이다.

본질적으로 부동산 정책, 심리, 정보 등의 요인을 측정하고 진단하는 것은 상당히 어렵기에, 부동산 비기초가격과 관련한 요인을 분석하는 연구가 더디게 진행되고 있다는 것이다.

그럼에도 불구하고, 부동산 비기초가격을 해부하고 독해하려는 시도는 중요한 학술적 의미를 갖는다. 부동산 비기초가격에 관한 연구가 더욱 고도화되기 위해서는 관련 설명변수를 모색하는 연구가 적극적으로 시행되어야 하기 때문이다. 지금과 같이 부동산 비기초가격과 연관을 맺는 요인에 대한 연구가 미진한 경우, 부동산 비기초가격에 대한 설명은 대부분 이론적 판단에 근거할 수밖에 없다. 그뿐만 아니라, 부동산 비기초가격에 대한 추정은 부동산 기초가격에 대한 가정에 종속적이게 된다. 예컨대, 부동산 비기초가격에 대한 대리변수(proxy)가 부재하면 비기초가격은 시장가격에서 기초가격을 제외한 값으로 정의될 수밖에 없다. 따라서 부동산 비기초가격에 대한 연관된 요인을 분석한다면 발전된 추정 모형을 구축할 수 있을 것이다.

부동산 비기초가격에 대한 구체화된 접근은 정책적으로도 상당한 함의를 가진다. 부동산 가격이 기초가격에 의해 변동하는지 혹은 비기초가격에 의해 등락하는지에 따라 관련한 정책의 성격이 달라질 수 있다(홍정의 외, 2022). 단순화하자면, 예를 들어 실질적(기초가격) 상승은 공급의 부족 차원에서, 심리적(비기초가격) 상승은 투기적 수요 차원에서 이해해야 할 수 있을 것이다. 더 나아가, 부동산 비기초가격이 어떤 요인에 의해 변화하는지 파악하여 수요자의 인식 구조를 보다 정확히 이해한다면 더욱 효과적이고 종합적인 부동산 정책이 가능할 것이다(권순구·강명구, 2011).

이에 본 연구는 언론매체가 대중적 심리를 일정 부분 주도하며, 이렇게 형성된 집단적 여론이 가격에 영향을 미친다는 점에 착안하여, 부동산 신문기사 빅데이터를 활용해 부동산 비기초가격을 분석한다. 분석 결과를 토대로 부동산 비기초가격과 부동산 신문담론의 관계, 나아가 부동산 비기초가격과 부동산 심리시장의 관계를 고찰한다. 이 연구를 통해, 부동산 비기초가격에 대한 추정 방안을 모색하는 연구에 기초자료를 제공하고, 부동산 비기초가격에 대한 이해도를 재고하여 부동산 가격 안정화 정책에 대한 함의를 도출하고자 한다.

제 2 절 연구의 범위 및 방법

본 연구는 공간적 분석 범위를 국내 부동산 시장논의의 중심지인 서울시 전체로 설정했으며, 분석의 대상은 서울 아파트 매매가로 한정했다. 시간적 분석 범위는 2014년 1월부터 2021년 12월까지 96개월로 하며, 분석의 시간 단위를 월 단위로 특정했다. 2008년 글로벌 금융위기로 인한 구조변동 가능성을 고려하여 2008년부터 2013년까지 이어진 침체기와 안정기를 제외하고, 서울 아파트 매매가격이 상승세로 전환되기 시작한 2014년부터 상승 추이가 더욱 강화된 2021년까지를 분석 기간으로 삼았다(윤성훈·한성원, 2018). 시장 과열 및 가격 거품에 대한 논의가 본격적으로 재확산됐던 2010년대 후반부터 2020년대 초반까지를 연구의 범위에 포함해 부동산 비기초가격을 추정하고 분석했다.

부동산 비기초가격은 부동산 시장가격에서 부동산 기초가격을 제외한 값이라 가정했다. 부동산 시장가격에서 임대가격이나 경기 동향 등 경제적 요인으로 설명되는 기초가격(펀더멘탈) 영역과 정책이나 심리 등 비경제적 변수로 이해되는 비기초가격(비펀더멘탈) 영역이 있다고 상정한 것이다. 부동산 시장가격은 한국부동산원의 아파트매매가격지수를 기본으로 하며, 기초가격 대리변수는 선행연구를 바탕으로 아파트전세가격지수, 소비자물가지수, 주택담보대출 가중평균금리. 전산업생산지수, 경기동행지수 등을 활용했다. 부동산 시장가격에 대한 FM-OLS(fully modified least squares) 회귀식을 구축하여 기초가격으로 해석되는 영역과 그렇지 않은 영역(잔차항)을 분해하고, 잔차항을 부동산 비기초가격으로 가정해 별도의 변수로 삼았다.

부동산 신문기사는 중앙일보(www.joongang.co.kr), 동아일보(www.donga.com), 경향신문(www.khan.co.kr), 한겨레신문(www.hani.co.kr), 매일경제신문(www.mk.co.kr), 한국경제신문(www.hankyung.com)을 대상으로 했다. 국내에서 영향력이 큰 신문사를 선정하고, 자료의 균형을 위해 보수, 진보, 경제지를 모두 포함했다. 한국언론재단의 기사 검색 사이트인 KIND(www.bigkinds.or.kr)를 통해 해당 신문사가 분석 기간 내에 발행한 서울 부동산 관련 기사를 크롤링했다. 수집한 기사는

토픽 모델링을 통해 정량화했다. 산출된 주요 주제가 부동산 신문담론의 골자를 이룬다고 간주하고, 이를 분석의 변수로 활용했다.

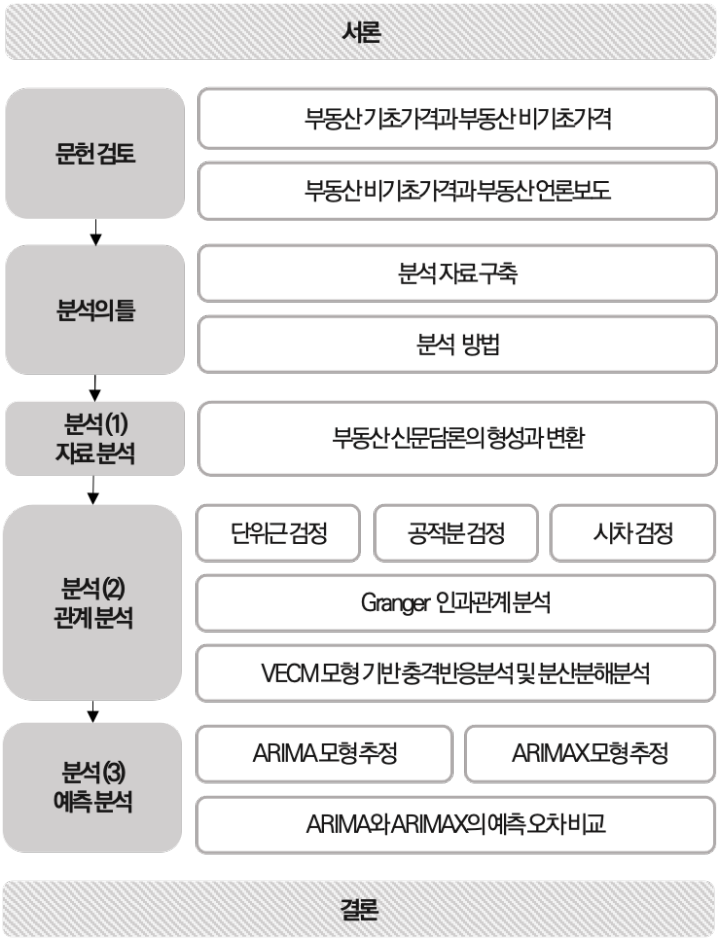
부동산 비기초가격과 부동산 신문담론의 관계를 살펴보기 위해, 다음과 같은 과정을 거쳤다. 분석 자료의 안정성을 검토하기 위해 단위근 검정을 시행하고, 안정적인 시계열 자료로 변환하여 Granger 인과관계 분석을 실시했다. Granger 인과관계 분석 결과를 토대로 변수 간의 관계를 파악하고, 추후의 모형에 포함될 변수를 선별했다. 해당 변수를 대상으로 공적분 검정을 시행하여 분석의 모형을 결정했다. 앞선 검정 결과를 기준으로 VECM 모형을 추정하고, 충격반응분석과 분산분해분석을 실행했다.

더 나아가, 부동산 비기초가격에 대한 부동산 신문기사의 예측 유용성을 분석하기 위해 부동산 신문기사 논조지수를 구축했다. 토픽 분석과 감성분석 결과를 토대로 부동산 신문기사 논조지수를 만들고, 부동산 비기초가격에 대한 ARIMA(Auto Regressive Integrated Moving Average) 모형과 부동산 신문기사 논조지수를 포함한 부동산 비기초가격 ARIMAX(Auto Regressive Integrated Moving Average with eXogenous variables) 모형을 비교했다. ARIMA 모형과 ARIMAX 모형의 예측오차인 RMSE(Root Mean Square Error), MAE(Mean Absolute Error)를 비교하여, 부동산 신문기사 논조지수가 부동산 비기초가격의 예측 정확성을 높이는지 실증했다.

제 3 절 연구의 구성

본 연구는 다음과 같이 구성된다. 제 1 장에서는 연구의 배경 및 목적, 연구의 범위 및 방법, 연구의 구성을 설명한다. 제 2 장에서는 부동산 비기초가격의 개념, 부동산 비기초가격과 부동산 신문담론의 관계에 대해 살펴보고 관련 선행연구를 검토한다. 제 3 장에서는 분석의 틀을 설명한다. 분석 자료의 구축 과정과 분석 방법에 대해 설명한다. 제 4 장에서는 분석 자료의 형성과 변환 과정을 당시

의 사회적·경제적·정치적 맥락에 맞춰 설명한다. 제 5 장에서는 Granger 인과관계 분석과 VECM 모형 기반의 충격반응분석 및 분산분해분석 결과를 토대로 부동산 비기초가격과 부동산 신문담론의 관계에 대해 살펴본다. 제 6 장에서는 ARIMA 모형을 통해 부동산 신문기사의 부동산 비기초가격 예측유용성을 검증한다. 마지막으로, 제 7 장에서는 분석 결과의 학술적 함의와 정책적 시사점을 제안하고, 연구의 의의와 한계에 대해 서술한다.



[그림 1] 연구의 흐름도

제 2 장 이론적 배경 및 선행연구 고찰

행동경제학(behavioral economics)에 따르면, 인간의 합리성은 제한적이다. 사람들은 반드시 이용가능한 모든 정보를 충분히 활용해 최적의 의사결정을 내리지 않는다. 오히려 사람들은 다양한 요인의 영향을 종합적으로 고려하는 합리적 기대(rational expectation) 대신, 과거의 정보만을 제한적으로 활용하는 적응적 기대(adaptive expectation)에 따라 판단할 가능성이 높다(Cagan, 1956; Simon, 1959). 이는 인간은 주어진 조건과 상황을 모두 고려할 수 없을 뿐 아니라, 이들의 판단 과정에 이성 외에 감성도 일정 정도 개입하게 되기 때문이다(Tversky and Kahneman, 1991, 1992; Thaler, 2009; 권순구·강명구, 2011).

이러한 관점에서, 부동산 시장 또한 객관적 요인뿐 아니라 주관적 심리에 영향을 받는다(Tversky and Kahneman, 1991). 부동산에 대해 시장 참여자는 자신의 직관과 기대를 형성하고, 이들의 판단은 부동산 가격에 영향을 미친다. 다시 말해, 부동산 시장 참여자의 심리가 부동산 시장 가격에 실질적 영향을 끼치게 된다(Kahneman and Tversky, 1984; Weber and Devaney, 1996; Johnson, 2010) 그 때문에, 이론적으로 부동산 시장 가격이 해당 자산을 보유할 경우 장래에 예상되는 수익의 현재가치에 의해 결정되어야 하지만, 현실에서는 시장 참여자의 기대나 과신 등에 영향을 받아 변동한다(김봉한, 2004; 김윤영, 2013). 부동산 가격에는 기초가격(market fundamentals)만으로 설명되지 않는 지점, 즉 시장 가격과 기초가격의 차이인 비기초가격(market non-fundamentals) 영역이 존재하게 된다는 것이다(Blanchard and Watson, 1982; Summers, 1986; Stiglitz, 1990; Roche, 2001).

한편, 부동산 비기초가격을 형성하는 주요 요인인 시장 참여자의 심리는 측정이 어렵기 때문에, 그의 관심도와 영향력에 비해 연구가 더디게 진행되어 왔다. 이에 본문은 기존의 정형 데이터 대신 비정형 텍스트 데이터를 활용한다. 특히, 언론보도에 관한 비정형 텍스트 데이터를 활용해, 언론보도가 대중심리에 영향을

주고 대중심리가 비기초가격의 변동에 영향을 주는 연결 관계에 주목하고자 한다(Shiller, 2000; Case and Shiller, 2003; Bucchianeri, 2011). 비정형 데이터를 이용해, 부동산 연구 영역에서 기존에 분석하기 어려웠던 것을 분석할 수 있게 되고 예측하기 어려웠던 것을 예측할 수 있게 되는, 그 확장 가능성에 관한 탐색적 연구를 전개하고자 한다.

제 1 절 부동산 비기초가격

1) 부동산 기초가격과 비기초가격

이론적으로, 시장가격은 해당 자산을 보유할 때 장래에 기대되는 수익의 현재가치, 즉 시장 기초가치(market fundamental value)로 결정되어야 한다(김봉한, 2004). 그러나 미래의 기대 현금흐름에 대한 현재가치만으로 설명하기에는, 현실적으로 우리가 목도하는 시장가격은 지나치게 변동적이고 확률적이다. 그 때문에 현시된 효용과 합리적 가치만으로 설명되지 않는 영역을 설명하려는 시도가 꾸준히 있었다.

다수의 관련 연구는 시장 요소로 설명할 수 없는 가격 영역이 존재한다는 사실을 이용하고 있다(홍정의 외, 2022). 이들 연구는 내재가치로 설명되는 부분을 기초가격(fundamental price)으로, 기초가격과 시장가격 간에 차이를 비기초가격(non-fundamental price)[1]으로 가정하고 있다. 시장가격(P_t)이 기초가격(P_t^f)과 기초가격에서 이탈된 부분인 비기초가격(P_t^{nf})으로 구성되어 있다고 관점에 기초해 가격을 분리하고 있는 것이다(식1).

$$P_t = P_t^f + P_t^{nf} \tag{1}$$

1) 비기초가격에 대한 표현은 각각의 연구 및 모형에 따라 달라진다. 대표적으로, 비기초가격은 유행(Summers, 1986), 거품(Blanchard and Watson, 1982) 등으로 지칭된다. 다만, 사실상 대부분의 실증연구가 기초가격에서 이탈된 시장가격의 크기를 분석대상으로 하고 있다는 점을 감안하여, 본 연구는 논의상의 편의를 위해 관련 용어를 기초가격(fundamental price)과 비기초가격(non-fundamental price)으로 통칭하여 서술한다.

Roche(2001)에 의하면, 자산 가격을 기초가격과 비기초가격으로 설명하는 대표적인 모형은 Summers(1986)의 유행 모형(fad's model)과 Blanchard and Watson(1982)의 거품 모형(stochastic bubble's model)이다.

먼저, 유행 모형은 기초가격을 불안정(non-stationary) 시계열로(식2), 비기초가격(유행)을 안정(stationary) 시계열로 가정한다(식3).

$$P_t^f = P_{t-1}^f + e_t, \quad e_t \sim iid(0, \delta_e^2) \tag{2}$$

$$P_t^{nf} = \rho P_{t-1}^{nf} + \upsilon_t, \quad 0 < \rho < 1, \ \upsilon_t \sim iid(o, \sigma_\upsilon^2) \tag{3}$$

그리고 현실에서는 기초가격만을 독립적으로 관측할 수 없기 때문에 대리변수를 활용해야 하며, 대리변수(P_t^p)는 측정오차(measurement error)를 동반한다(식4).

$$P_t^p = P_t^f + u_t, \quad u_t \sim iid(o, \delta_u^2) \tag{4}$$

식(1)부터 식(4)을 정리하면 식(5)과 같은 회귀식이 산출된다. 이는 $t+1$기의 시장가격은 t기의 시장가격과 기초가격의 조합, 그리고 측정오차로 구성된다는 것을 의미한다[2].

$$P_{t+1} - P_t = \alpha_1 (P_t - P_t^P) + \epsilon_t, \quad \epsilon_t \sim iid(0, \delta_\epsilon^2) \tag{5}$$

다음으로, 거품 모형 또한 비기초가격을 시장가격과 기초가격의 차이로 정의했다. 시장가격이 기초가격과 동일하지 않다면 비기초가격(거품)이 존재한다고 가정하는 것이다(식6), 이때 기초가격과 비기초가격은 각각 식(7)과 식(8)을 만족하며,

2) 유행모형에서는 시장가격과 기초가격의 차이를 유행으로 간주하며, t기의 유행이 일정 비율(α_1) 만큼 $t+1$기의 수익률에 영향을 미친다고 본다.

기초가격은 자산을 소유함으로써 얻게 되는 소득(D_t)과 할인 요소(α)에 의거한다.

$$P_t^{nf} = P_t - P_t^f \tag{6}$$

$$P_t^f = \sum_{k=0}^{\infty} \alpha^{(k+1)} E_t(D_{t+k}) \tag{7}$$

$$P_t^{nf} = \alpha E_t(P_{t+1}^{nf}), \quad 0 < \alpha < 1 \tag{8}$$

Blanchard and Watson(1982)의 거품 모형은 비기초가격(거품)이 계속해서 일정한 확률로 유지 및 성장할 수 있음을 가정했다(식8). 이에 Van Norden(1996), Schaller and Van Norden(1997), Van Norden and Vigfusson(1998)은 비기초가격이 일정한 확률로 소멸하거나 지속된다는 가정은 다소 비현실적이며, 비기초가격의 크기가 증가함에 따라 유지 확률(q)은 감소할 것이라 지적했다(식9).

$$q = q(P_t^{nf}), \quad \frac{\partial q(P_t^{nf})}{\partial |(P_t^{nf})|} < 0 \tag{9}$$

이들은 식(9)과 같은 조건으로 비기초가격은 음의 값과 양의 값을 모두 가질 수 있다고 보았다. 즉, 비기초가격이 소멸하는 경우(C)[3]와 비기초가격이 지속되는 경우(S) 모두를 가정하고 있는 것이다. 두 가지 경우를 반영한 산출식은 식(10)과 같으며, 각각의 상황에 대한 비기초가격의 조건부 기댓값은 식(11), 식(12)과 같다.

$$\begin{aligned} E_t(P_{t+1}^{nf}) = (1 - q(P_t^{nf})) \cdot E_t(P_{t+1}|C) \\ + q(P_t^{nf}) \cdot E_t(P_{t+1}|S) \end{aligned} \tag{10}$$

3) 비기초가격(거품)이 붕괴하는 경우, 정부나 금융기관이 이를 저지하기 위해 개입한다. 따라서 비기초가격은 상당기간에 걸쳐 완만하게 소멸된다고 가정한다. 이때 $g(\cdot)$는 연속적이고 미분 가능한 함수이다.

$$\begin{cases} E_t(P_{t+1}^{nf}|C) = g(P_t^{nf}), \quad 0 \le g \le 1 \quad g(0) = 0 & (11) \\ E_t(P_{t+1}^{nf}|S) = \dfrac{P_t^{nf}}{\alpha q(P_t^{nf})} - (\dfrac{1-q(P_t^{nf})}{q(P_t^{nf})}) \cdot g(P_t^{nf}) & (12) \end{cases}$$

앞선 식을 통해 산출한 자산 보유의 기대수익은 다음과 같다. 비기초가격이 붕괴하는 시기(C)의 기대수익은 비기초가격(P_t^{nf})의 감소함수로(식13), 비기초가격이 유지되는 시기(S)의 기대수익은 비기초가격(P_t^{nf})의 증가함수로 표현된다(식14).

$$\begin{cases} E_t(P_{t+1} - P_t|C_t) = g(P_t^{nf}) - \dfrac{P_t^{nf}}{\alpha} & (13) \\ E_t(P_{t+1} - P_t|S_t) = (\dfrac{1-q(P_t^{nf})}{\alpha q(P_t^{nf})}) \cdot (P_t^{nf} - \alpha g(P_t^{nf})) & (14) \end{cases}$$

Taylor 급수 전개(Taylor series expansion)를 통해 위 식을 1차 근사식으로 변환한 결과는 다음과 같다. $t+1$기의 시장가격은 t기의 시장가격과 비기초가격의 조합으로 표현될 수 있으며(식15-16)[4]. 이는 다시 t기의 시장가격과 기초가격으로 정리될 수 있다(식17-18).

$$\begin{cases} P_{t+1} - P_t = \beta_{S0} + \beta_{S1} P_t^{nf} + \eta_{t+1}, \quad \eta_{t+1} \sim N(0, \delta_S^2) & (15) \\ P_{t+1} - P_t = \beta_{C0} + \beta_{C1} P_t^{nf} + \eta_{t+1}, \quad \eta_{t+1} \sim N(0, \delta_C^2) & (16) \end{cases}$$

$$\begin{cases} P_{t+1} - P_t = \beta_{S0} + \beta_{S1} P_t + \beta_{S2} P_t^{f} + \eta_{t+1}, \quad \eta_{t+1} \sim N(0, \delta_S^2) & (17) \\ P_{t+1} - P_t = \beta_{C0} + \beta_{C1} P_t + \beta_{C2} P_t^{f} + \eta_{t+1}, \quad \eta_{t+1} \sim N(0, \delta_C^2) & (18) \end{cases}$$

앞선 모형들은 모두 애초 금융시장 및 주식시장을 대상으로 설계되었으나, 부동산시장에 관한 연구에도 유용한 이론 틀을 제공한다. 이들은 부동산 시장가격을 종속변수로 하고, 기초가격과 연관된 변수들을 설명변수로 하는 회귀식을 산출하

4) 식(1)부터 식(16)까지 모형의 전개 과정은 Roche(2001), 김봉한(2004), 오근엽 외(2005)에 상당부분 의존하고 있다.

여 잔차항(ε_t)을 구하고, 그 잔차항을 일종의 비기초가격으로 간주하는 일련의 분석에 대한 이론적 토대를 제공하고 있다(식19).

이때 기초가격의 대리변수로 활용되는 자료(x)는 개별 연구에 따라 차이를 보이며, 주로 임대가치, 경기순환, 금리 등과 관련된 변수가 활용되고 있다. 필연적으로 대리변수를 사용해야 하는 만큼, 완전한 객관적 모형을 구축할 수 있는 것은 아니다. 때문에, 유사한 시점 및 배경하에 시행된 선행연구를 면밀히 검토해 참조할 필요가 있다. 이에 국내 관련 실증연구에서 활용되고 있는 분석 자료 및 분석 방법에 대해 이어 후술하기로 한다.

$$P_{t+1} = \alpha_0 + \alpha_1 P_t + \alpha_2 P_t^p + \varepsilon_t \tag{19}$$

$$\alpha_1 P_t^p = [\beta_1 x_1 \cdots \beta_n x_n] \tag{20}$$

2) 부동산 비기초가격의 추정

부동산 비기초가격의 존재 및 추이를 추정하는 방법은 크게 내재가치접근법, 공적분 검정법 2가지로 구분된다(〈표1〉). 내재가치접근법은 주택가격이 미래 기대수익과 현재 가치의 합으로, 다시 말해 해당 자산의 내재가치(instinct value)로 결정된다고 가정한다. 그리고 이러한 내재가치를 기초가격(fundamental price)으로, 시장가격과 내재가치 간의 차이를 비기초가격(non-fundamental price)으로 간주한다(Stiglitz, 1990). 한편, 공적분 검정법은 가격과 임대료의 관계에 기반한 방법으로, 비기초가격이 없다면 주택가격과 임대료의 움직임이 유사할 것이라 가정한다. 즉, 공적분 검정법은 주택가격과 기대수익 사이에 공적분(cointegration) 관계가 존재하면 비기초가격이 없고, 공적분 관계가 존재하지 않으면 비기초가격이 있다고 설명하는 모형이다(조무상·남주하, 2019).

먼저, 내재가치접근법을 이용하여 비기초가격을 추정한 국내 연구로는 김봉한(2004), 오근엽 외(2005), 김원희·강원철(2012), 박헌수·이창원(2012), 김윤영(2013), 신진호·황수성(2015), 홍정의 외(2022) 등이 있다. 김봉한(2004)은

FM-OLS 모형과 상태전환회귀식을 사용해 한국 주택가격에 비기초가격이 존재하는지를 분석했다. 주택을 보유할 경우 장래에 기대되는 수익의 현재 시장가치를 기초가격으로, 기초가격에서 이탈된 부분을 비기초가격으로 가정하여 분석했다. 해당 연구는 아파트매매가격지수를 종속변수로 하고, 산업생산지수(전산업), 이자율(회사채수익률), 소비자물가지수를 설명변수로 하는 회귀식을 추정해 잔차항을 구하고, 잔차항을 비기초가격이라 가정했다. 해당 잔차항에 대한 분석 결과를 바탕으로, 주택가격에 비기초가격 추세가 존재함을 보였다.

김원희와 강원철(2012)은 주택의 비기초가격을 시장가격과 수익의 현재가치 차이로 규정하여 주택시장의 비기초가격을 계산했다. 아파트전세가격지수와 수익 흐름을 아파트매매가격지수로 환산하고, 현금흐름 할인모형을 이용해 거품의 비율을 추정했다. 분석 결과를 바탕으로 서울과 인천 지역에는 비기초가격이 주택 전체 가격에 상당한 비중을 차지하고 있다고 주장했다.

이와 유사하게, 박헌수와 이창원(2012)은 아파트전세가격지수와 회사채수익률을 바탕으로 임대료 변수를 구축해 주택의 내재적 가치를 분석하였다. 이를 통해, 주택 가격변화가 기초가치의 변화에 따른 것인지 혹은 비기초가격에 의한 것인지 파악하였다. 그 결과, 수도권 지역에 20% 이상의 비기초가격이, 비수도권 지역에는 4개 광역시에서 15% 내외의 비기초가격이 존재했다.

김윤영(2013)은 VAR 모형을 활용해 한국 주택가격에 비기초가격이 존재하는지 검정했다. 주택가격의 추세는 거시경제변수에 의한 구조적인 추세인 기초가격과, 이와 독립적인 추세인 비기초가격으로 구분된다고 가정했다. 동 연구는 주택가격지수, 이자율(CD91일금리), 산업생산지수(전산업), 통화량(M1, 평잔), 환율(대미달러, 종가 평균), 주가(KOSPI, 월평균) 등을 변수로 하여 주택의 기초가격 VAR 모형을 구축했으며, 해당 모형에 대한 t검정을 통해 기초가격과 무관한 독립적 추세가 존재함을 밝혔다. 또한, 충격반응분석과 분산분해분석을 통해, 한국 주택가격은 기초가격보다 시장 심리나 정부 정책 등에 기인한 비기초가격의 충격에 크게 반응해 왔음을 보였다.

홍정의 외(2022)는 주택 비기초가격이 존재한다면 주택 가격 상승 수준이 소

득이나 금리 등 시장가격을 결정하는 기본적 요소의 변화로 설명 가능한 정도를 뚜렷이 상회할 것이라는 가정하에, 주택 가격변화 추정치를 분석했다. 해당 연구는 미국과 비교했을 때 서울 아파트 시장의 비기초가격은 미미한 수준이라고 결론지었다. 더불어, 주택가격의 변화가 실질적 변화에 기인하는지 혹은 낙관적 기대에 기인하는지 판단하는 것이 주택 정책 방향을 설계하는 데 있어 중요하다고 역설하며, 주택 비기초가격 연구의 필요성을 강조했다.

한편, 이용만(2000)은 공적분 검정법을 활용하여 부동산 실제 가격과 기초가격에 괴리가 존재하는지 검정하였다. 주택 전세값과 이자율(회사채수익률) 자료를 활용해 주택 임대료를 구하고, 주택매매가격과 임대료 간의 공적분 검정을 수행했다. 분석 결과, 서울 주택 가격과 임대료 사이에 공적분 관계가 있는 것으로 나타났다. 해당 연구는 이와 같은 결과를 바탕으로 부동산 시장가격이 기초가격을 크게 벗어나지 않으며, 뚜렷한 부동산 비기초가격 추세가 존재하지 않는다고 주장했다. 마찬가지로, 최차순(2010)은 아파트전세가격지수에 평당 평균아파트전세가격과 월세전환율을 곱하여 임대료 자료를 구했다. 임대료를 기초가격의 대리변수로 사용하여, 실제 가격과 기초가격의 관계를 실증 분석했다. 그 결과, 실제 가격의 월별 변화율이 기초가격의 월별 변화율을 크게 상회했으며, 이용만(2000)과 반대로 실제가격과 기초가격 간에 공적분 관계가 존재하지 않았다. 해당 연구는 공적분 검정 결과를 토대로 비기초가격의 존재 가능성을 주장했다(최차순, 2010).

계량 방법과 분석 기간에 따라 일부 상이하나, 대부분의 선행연구는 한국 부동산 시장에 부동산 비기초가격이 실존함을 보였다. 이들은 비기초가격이 부동산 가격에 높은 비중을 차지하며, 가격 변동을 상당 부분 주도한다고 지적했다. 이상의 연구 결과는 한국 부동산 시장에 정책, 심리, 정보 등과 같은 부동산 비기초가격 관련 요인이 실질적인 영향을 미친다는 점을 함의하며, 부동산 비기초가격에 대한 연구의 필요성을 내포한다.

〈표 1〉 부동산 비기초가격 관련 주요 선행연구

구분	저자	모형	분석기간	배경지역	분석자료	추정결과
내재가치 접근법	김봉한 (2004)	FM-OLS 모형, 상태전환회귀식	1987.01~ 2002.12	전국	아파트매매가격지수, 산업생산지수, 이자율(회사채수익률), 소비자물가지수	'89~'94: 비기초가격 추세 존재 '01: 비기초가격 추세 존재
	오근엽 외 (2005)	FM-OLS 모형, 상태전환회귀식	1987.01~ 2004.07	서울, 부산, 대구, 인천, 광주, 대전, 울산	아파트매매가격지수, 산업생산지수, 이자율(회사채수익률), 소비자물가지수	시기적으로 외환위기를 제외한 대부분의 시기에 비기초가격 추세가 존재 도시별로 인천지역과 대전지역에서 가장 현저
	김원희·강원철 (2012)	현금흐름 할인모형	2001.09~ 2012.03	서울, 부산, 대구, 인천, 광주, 대전, 울산	아파트매매가격지수, 아파트매매가격 대비 전세비율, 주택담보대출금리, 이자율(3년 만기 국고채수익률)	'12.03 기준 지역별 버블율: 강북(73.53%), 강남(90.13%), 부산(40.36%), 대구(30.80%), 인천(82.61%), 광주(23.49%), 대전(43.78%). 울산(30.24%)
	박헌수·이창원 (2012)	상태공간모형	2002.01~ 2012.10	서울, 부산, 대구, 인천, 광주, 대전, 울산	아파트매매가격지수, 아파트전세가격지수, 이자율(3년 만기 국고채수익률)	'02~'12 기준 지역별 버블율: 강북(22.1%), 강남(28.4%), 부산(12.6%), 대구(10.4%), 인천(23.4%), 광주(5.1%), 대전(15.1%), 울산(12.8%)
	김윤영 (2013)	VAR 모형	1993.01~ 2013.01	전국	주택가격지수, 이자율(CD91일금리), 산업생산지수(전산업), 통화량(M1), 환율, 주가(KOSPI)	'99.03~'13.01: 비기초가격 추세 존재
	신진호·황수성 (2015)	사용자비용모형, 회귀분석	2001.08~ 2014.07	서울, 부산, 대구, 인천, 광주, 대전, 울산	주택매매가격지수, 주택전세가격지수, 전월세 전환율, 이자율(CD91일금리), 산업생산지수(전산업), 통화량(M1), 환율(대미달러, 종가평균),	금융위기 이전: 주택시장에 과신이 형성 금융위기 이후: 과신이 사라지고 투자심리가 위축
	홍정의 외 (2022)	AR 모형	2015.01~ 2020.12	미국, 서울	아파트매매가격지수, 아파트전세가격지수, 실질 국내총생산, 이자율(주택담보대출금리), 소비자물가지수, 경기동행지수	'18 이후 비기초가격 추세 존재 다만, 그 수준은 상대적으로 미미함
공적분 접근법	이용만 (2000)	공적분 검정	1986.01~ 1999.12	서울	주택매매가격지수, 주택전세가격지수, 이자율(회사채수익률)	서울 아파트가격과 서울 아파트 임대료 사이에 공적분 관계가 있음
	최차순 (2010)	과민변동성검정, 공적분 검정	1999.01~ 2009.12	전국, 서울, 강남	아파트매매가격지수, 아파트전세가격지수, 평당아파트전세가격, 월세전환율, 이자율(회사채수익률)	아파트 가격과 아파트 임대료 사이에 공적분 관계가 없음

제 2 절 부동산 비기초가격과 부동산 언론보도

1) 부동산 언론보도

부동산 비기초가격의 추세를 추정하는 데서 나아가, 부동산 비기초가격에 영향을 미치는 요인들을 구체화하는 연구가 일부 시행되었다(Shiller, 2000; Mercille, 2014; Bucchianeri, 2011). 거시경제 변수만으로 설명되지 않는 시장영역에 대한 관심이 증가하면서, 기존의 경제지표 이외의 요인을 통해 부동산 가격을 이해하려는 연구 흐름이 형성된 것이다(박재수·이재수, 2019). 특히, 인간의 심리에 기초하여 시장 변동을 설명하는 행동경제학 이론이 부동산학계에서도 조명되면서, 투자자와 소비자의 심리에 의거하여 부동산 비기초가격의 변화를 야기하는 요인들을 설명하는 시도가 이루어졌다.

시장이란 종국적으로 참여 주체의 이해관계에 의해 형성되는 장이기에, 부동산 시장의 추이 및 변화를 대중적 심리를 중심으로 살펴볼 필요가 있다(김대원·유정석, 2013). 부동산 비기초가격은 공통된 믿음을 전파하고 대중적 심리를 형성하는 사회적 기제가 있는 경우에 존재할 가능성이 높다(Blanchard and Watson, 1982; Camerer, 1989; 김경환·이한식, 2000). 그리고 이러한 통상적 담론(discourse)을 파악하기 위해서는 부동산 시장에 관한 정보를 대중에게 제공하는 언론매체의 역할을 중점적으로 탐색해야 한다(서정석 외, 2021).

심리와 직관은 부동산 시장을 변동시키는 주요한 요인 중의 하나이며(Kahneman and Tversky, 1979, 1984), 언론매체는 그러한 대중의 관점에 상당한 영향을 준다(Shiller, 2000). 부동산 언론보도는 부동산 시장에 대한 전망을 형성해 수요와 공급, 가격과 거래량을 변동시킨다. 예컨대 언론매체가 부동산 시장에 대한 긍정적인 전망을 보도할 경우, 매수자들은 가격 상승을 예상하게 되고 매도자들은 가격 상승을 기대하게 된다(Shiller, 2000; Akerlof and Shiller; 2009). 이에 따라 매수자의 수요는 증가하고, 매도자는 가격을 높이거나 매물을 거두는 경향이 나타난다. 혹은 반대로 언론매체에서 부정적인 전망을 보도하면 매

수자들과 매도자들이 가격 하락을 예상하게 된다. 이에 매수자의 수요는 감소하고, 매도자는 매물을 빨리 팔기 위해 가격을 낮추는 경향이 생겨난다.

나아가, 언론매체는 부동산 시장 전망에 대한 정보를 전달할 뿐만 아니라, 대중의 심리에 변화를 줌으로써 시장 역학(market dynamics)에 영향을 미치는 행위자로서 기능한다. 가령, 반복적 보도로 인해 부동산 정책에 대한 집단적 불신이 조성되거나, 부동산 가격에 대한 낙관적 기대가 형성되는 등 언론매체는 부동산 비기초가격을 형성 및 변화시키는 매개체로서 기능할 수 있다(우윤석·이은정, 2011). 이러한 맥락에서 부동산 비기초가격과 언론매체의 관계, 부동산 심리시장과 언론매체의 관계에 주목하는 연구들이 진행되었다.

대표적으로, Shiller(2000)는 건축 비용, 인구, 금리 등의 거시경제 변수만으로 주택가격을 설명하는 것은 불가능하다고 주장했다. 즉, 기초가격은 주택가격 전체를 설명하기에 불충분하다는 것이다. 대부분의 거시경제 요인은 안정적이고 추세적인 데 반해, 몇몇 지역의 주택가격 변동 패턴은 변칙적이고 일시적이다. 해당 연구는 이와 같은 주택 비기초가격(거품)의 변화 원인을 구조적 요인, 문화적 요인, 심리적 요인 측면에서 입체적으로 분석하고, 문화적 요인으로서 언론매체에 주목했다. 언론은 부동산에 대한 대중적 관심과 사고의 범주를 만들어 내 투기적 가격 운동을 전파하고 비기초가격 변동을 견인하는 역할을 한다고 주장했다.

Mercille(2014)는 주택시장에 비기초가격(거품)이 형성된 이유를 정치경제학적 관점에서 분석하고, 그 원인으로 신문기사 보도 행태를 꼽았다. 해당 연구는 언론매체로 인해 아일랜드 주택가격이 기초가격을 꾸준히 상회했다고 주장했다. 언론매체 자체가 신자유주의적 사고를 갖고 있었고, 정·제계와 다중의 연결고리를 유지했고, 부동산 개발 회사의 광고 수주에 대한 압력을 느끼고 있었고, 엘리트 집단 출신의 소위 부동산 전문가에 전적으로 의존하고 있었기 때문에 이들의 부동산 관련 보도는 선택적이고 때로는 작위적이었다고 보았다. 그리고 이와 같은 보도가 부동산 비기초가격을 형성 및 유지하는 동인이 되었다고 해석했다.

Bucchianeri(2011)은 비기초가격과 언론의 관계에 대한 Shiller(2000), Shiller(2002)의 논의에 창안해, 언론보도가 홍콩 주택시장의 비기초가격(거품)에

미친 영향을 분석했다. 주택경기에 대한 긍정적인 논조의 신문기사 보도는 주택거래회전율과 양(+)의 관계를 가졌다. 이를 통해 주택시장에 대한 언론의 관심과 광고가 투기를 부추겼으며, 주택의 비기초가격과 언론매체 간에 유의미한 연관 관계가 있다고 결론지었다.

한편, 국내에서도 부동산 가격과 언론매체의 관계에 대한 연구가 시행되었다. 국내 부동산 시장의 변칙성, 부동산 언론보도의 화제성을 대변하듯 꽤나 많은 연구가 진행되었다. Park, Min Jeong et al.(2016)은 2009년 1월부터 2013년 12월 사이에 발행된 부동산 신문기사와 부동산 시장 가격의 관계를 분석했다. 토픽분석(topic analysis)을 통해 부동산 신문기사를 10개의 주제(Apartment Lotting-Out, World Economy, Area Development, Estate Development, Government Policy, Rental Housing Market, Bank Finance, Business Industry, Stock Market)로 분류했고, 토픽을 포함한 회귀모형과 포함하지 않은 회귀모형의 R-squared 값을 비교했다. 분석 결과, 토픽을 포함할 경우 모형의 설명력이 15.3% 향상됐다.

이종민 외(2017)는 2011년부터 2016년까지 전세 관련 신문기사의 전세가격변동률 예측력을 분석했다. 먼저 단어 '전세'가 포함된 신문기사를 수집하고 토픽 분석을 시행해, 전체 기사를 9개의 주제(전세. 월세, 대출, 정책, 경기, 거래, 분양, 공급)로 구분했다. 그리고 전세가격변동률에 대한 ARIMA 모형과 전세가격변동률에 토픽이 포함된 ARIMAX 모형을 비교해, 전세 관련 신문기사의 전세가격변동률 예측력 정도를 분석했다. 예측오차인 RMSE(Root Mean Square Error)와 MAE(Mean Absolute Error)를 비교·분석한 결과, 10% 대의 예측력 향상 결과를 보였다. 이를 통해, 이들 연구는 부동산 신문기사가 부동산 시장 가격을 예측하는 보조지표로써 활용할 수 있음을 시사했다.

2) 부동산 언론보도의 논조

더 나아가, 부동산 언론보도의 영향을 구체적으로 파악하기 위해서는 그 논조

를 파악할 필요가 있다. 앞서 언급했듯, 부동산 언론보도가 대중의 심리에 미치는 영향은 논조에 따라 달라진다. 가령, 부동산 시장 전망에 대한 긍정적 논조가 주류를 이룰 경우 거품 현상이 발생할 가능성이 높아질 수 있으며(Shiller, 2000), 부정적 논조가 지배적일 경우 부동산 시장의 침체가 장기화될 수 있다. 이처럼 논조에 따라 부동산 언론보도가 부동산 시장 소비자의 심리에 미치는 영향은 차별적일 것이라는 문제의식하에, 부동산 언론보도의 단순 빈도나 주제뿐만 아니라 논조까지 주목한 연구가 시행됐다.

초기 연구들은 주로 논조와 관련된 특정 단어('상승', '하락' 등)가 포함된 부동산 신문기사 건수를 시계열 자료로 구축하여 부동산 시장과의 관계를 분석했다(김진유, 2006; 우윤석·이은정, 2011; 김대원·유정석, 2016). 최초의 관련 연구인 김진유(2006)는 2001년 1월부터 2005년 12월까지 '투기'라는 단어가 포함된 신문기사를 수집하고, 기사를 논조에 따라 상승·하락·중립 기사로 구분했다. '규제 완화', '집값 상승' 등에 관한 기사를 상승 기사로 구분하고, '세제 강화', '규제 강화', '집값 하락' 등에 대한 기사를 하락 기사로 분류해, 월별 기사 건수를 시계열 자료로 구축했다. 그리고 논조별 신문기사 건수와 지역별 주택가격지수에 대한 Granger 인과관계 검정을 수행했다. 부동산 신문기사 논조와 부동산 가격 추이의 관계를 분석한 결과, 하락 논조 기사 수는 가격에 별다른 영향을 미치지 않았으나, 상승 논조 기사 수는 가격에 유의미한 영향을 미쳤다.

우윤석·이은정(2011)은 2003년 2월부터 2008년 7월까지 부동산 언론보도 건수와 부동산 시장 가격의 관계에 대해 분석했다. '강남', '아파트', '가격', '상승'과 같은 단어가 포함된 신문기사를 수집하고, 보도가 서울(강남·강북) 및 광역시 아파트 매매가와 전세가에 미치는 영향을 파악했다. 연구 결과, 가격 상승에 관한 신문기사 건수는 강남 및 강북 지역의 연립주택 매매가와 전세가, 그리고 광역시 아파트 매매가와 연립주택 전세가와 그레인저 인과(granger cause) 하는 것으로 나타났다. 또한, 강남 지역에 대한 언론보도가 빈번해지면 강남 외 지역의 부동산 가격도 시차를 두고 상승하는 것을 확인했다. 이러한 결과를 토대로, 해당 연구는 부동산 언론보도 건수가 부동산 시장 가격 안정에 불리한 영향을 미칠 수 있다고

결론 내렸다. 즉, 언론보도가 가격 상승을 유도할 가능성이 있으며, 이는 시장 안정에 부정적인 영향을 미친다는 점을 지적했다.

김대원·유정석(2016)은 1990년 1월부터 2014년 6월 전국과 서울을 배경으로 부동산 신문기사의 논조가 아파트 매매가와 전세가에 미치는 영향을 분석했다. 분석을 위해 부동산 신문기사 제목에 나타난 '상승' 또는 '하락' 단어 빈도수를 파악했으며, 신문기사 제목 속의 논조와 아파트 매매가·전세가의 관계를 Cox 비례위험 모형(Cox proportional hazard model)을 토대로 분석했다. 그 결과, 신문 논조가 매매가 가격 국면의 지속 기간 간에는 유의미한 영향을 미치지 않았으나, 논조가 전세가의 지속 의존성에는 일관된 방향으로 영향을 미쳤다. 전세가격과 관련된 신문기사에 '상승'('하락')이라는 단어가 증가할수록 상승(하락) 추세가 오랫동안 지속되었다. 이를 통해, 동 연구는 신문기사 논조가 시장 가격에 유의미한 영향을 미칠 수 있음을 보였다.

반면, 최근 연구는 단순히 특정 단어의 포함 여부로 기사 논조를 파악하는 수준에서 한 단계 더 나아가, 감성분석(sentiment analysis)을 통해 부동산 신문기사의 논조를 체계적으로 분석하여 부동산 시장과의 관계를 파악했다(경정익·이국철, 2016; 박재수·이재수, 2019; 문태헌·김혜림, 2020; 서정석 외, 2021). 대표적으로, 경정익·이국철(2016)은 2013년 1월부터 2015년 8월까지 서울과 수도권 부동산 관련 신문기사를 수집하고, 감성분석을 통해 해당 신문기사의 논조를 정량화했다. 텍스트 마이닝(text mining)을 활용해 부동산 신문기사 감성분석 모형을 개발하여 부동산 신문기사 논조지수를 산출하고, 신문기사 논조지수와 아파트 실거래가의 상관관계를 실증 분석했다. 분석 결과, 신문기사 논조지수는 부동산 가격과 0.6 이상의 상관계수를 보였다. 해당 연구는 높은 상관성을 토대로 부동산에서의 감성분석의 유효성 및 필요성을 논증했다.

박재수·이재수(2019)는 2010년 1월부터 2017년 12월까지 서울을 사례로 부동산 신문기사에서 아파트 매매와 관련된 긍·부정 논조지수를 추출했으며, 논조지수가 아파트 매매가와 시계열적으로 유의미한 영향 관계를 맺는지 분석했다. 우선 토픽 분석을 통해 신문기사 주제를 분류했으며, 아파트 매매와 관련된 단어를 중

심으로 감성분석을 시행했다. 토픽 분석과 감성분석을 통해 산출된 월별 논조지수와 아파트 매매가의 관계를 교차상관분석, 인과관계 분석을 통해 검증했다. 그 결과, 신문기사의 긍정 논조지수는 소형아파트의 매매가격지수에 유의미한 영향을 미쳤다. 해당 연구는 신문기사에 긍정 표현이 늘어나면 아파트 매매가가 상승함을 보였으며, 이러한 선·후행 관계에 착안하여 논조지수 분석을 통한 아파트 매매가의 예측 가능성을 제시했다. 즉, 신문기사와 매매가의 관계를 바탕으로, 논조지수를 활용한 부동산 가격 예측모형의 개발 가능성을 제안했다.

서정석 외(2021)는 2011년 7월부터 2018년 12월까지의 부동산 신문기사에서 자주 등장하는 단어 또는 문장을 추출해 신문 논조의 긍·부정에 대한 감성사전을 개발하고, 감성사전에 기초하여 논조지수를 추출했다. 부동산 신문기사 논조지수와 아파트 매매거래량, 아파트 매매가, 부동산소비심리지수의 시계열 분석을 통해 언론보도의 논조와 주택경기 및 소비심리 간의 영향 관계를 파악했다. 분석 결과, 논조는 단기적으로 아파트 매매거래량에 부정적인 영향을 미쳤고, 아파트 매매가에 긍정적인 영향을 미쳤다. 다만, 장기적으로는 거래량 및 가격에 영향을 미치지 않았다. 또한, 신문기사 논조지수와 소비심리지수는 장기적인 동행 관계를 맺었다. 이를 통해, 해당 연구는 부동산 신문기사 논조지수와 부동산 시장 동향의 관계를 실증적으로 분석했으며, 논조지수가 소비심리지수의 보완재로서 기능할 가능성을 보였다. 다시 말해, 부동산 신문기사 논조지수와 부동산시장 소비심리지수의 영향 관계가 큰 만큼, 월 단위 설문조사를 통해 제공되는 부동산시장 소비심리지수를 보완하는 시장 동향 모니터링 지표로서 주·월간 단위 부동산 신문기사 논조지수가 활용될 수 있을 것이라 역설했다.

앞선 박재수·이재수(2019), 서정석 외(2021) 등의 연구 결과에 기초해, 논조지수를 설명변수로 하여 미래의 부동산 가격을 예측하는 연구가 시행되었다(박재수·이재수, 2021). 박재수·이재수(2021)는 2012년부터 2018년 사이에 보도된 신문기사와 방송뉴스에서 부동산 논조지수를 도출하고, 논조지수가 아파트 매매가격지수 예측에 유용한지 실증했다. 아파트 매매가에 대한 ARIMA 모형과, 논조지수가 추가된 ARIMAX 모형을 추정하여 계수의 통계적 유의성을 검증했다. 나아가, 각

각의 모형의 예측오차인 RMSE(Root Mean Square Error)와 MAE(Mean Absolute Error) 값을 비교해, 논조지수가 아파트 매매가의 예측을 향상하는지 검증했다. 분석 결과, 신문 및 방송 논조지수가 아파트 매매가격지수를 선행했다. 아파트 매매가 상승(하락) 초기에 이와 관련된 긍정(부정)적 내용을 보도되었으며, 보도는 시차를 두고 매매가에 양의 영향을 미쳤다. 또한, 아파트 매매가격지수 ARIMA 모형과 신문 및 방송 논조지수가 외생변수로 포함된 ARIMAX 모형의 예측오차를 비교한 결과, 신문 논조지수 ARIMAX의 RMSE와 MAE는 각각 7.90%, 6.21% 감소했으며 방송 논조지수 ARIMAX의 RMSE와 MAE는 5.05%, 5.42% 감소했다. 이를 통해, 언론매체의 논조지수가 부동산 가격의 예측모형에 유용한 외생변수로 기능할 수 있으며, 방송매체보다 신문기사의 예측 유용성이 더욱 높음을 보였다.

이와 같은 선행연구는 국내의 부동산 시장에서 부동산 언론보도의 내용 혹은 논조가 유의미한 설명변수로 작용함을 보였다. 즉, 이들은 부동산 언론보도 논조를 파악한다면 부동산 시장 가격을 더욱 정밀하게 분석 및 예측할 수 있을 것이라는 함의를 보였다.

〈표 2〉 부동산 신문기사 관련 주요 선행연구

저자	모형	분석기간	배경지역	분석대상	분석자료	신문기사	
						출처	건수(건)
김진유 (2006)	VAR	2001.01~2005.12	서울, 전국	매매가	KB 매매가격지수	일간지 및 경제지 4개	75.620
우윤석·이은정 (2011)	VAR	2003.02~2008.07	서울, 전국	매매가, 전세가	KB 매매가격지수, KB 전세가격지수	일간지 및 경제지 5개	2,757
김대원·유정석 (2016)	Cox 비례위험 모형	1990.01~2014.06	서울, 전국	매매가, 전세가	KB 매매가격지수, KB 전세가격지수	일간지 및 경제지	415,799
경정익·이국철 (2016)	감성분석, 상관분석	2013.01~2015.08	서울	실거래가, 거래량, 주택소비심리	아파트 실거래가, 아파트실거래가격지수, 아파트 매매거래량, 주택소비심리지수	포털사이트	35,082
이종민 외 (2017)	토픽 분석, VAR, ARIMA	2012.01~2016.12	전국	전세가	KB 전세가격지수	일간지 및 경제지 16개	42,272
박재수·이재수 (2019)	토픽 분석, 감성분석, VAR	2010.01~2017.12	서울	아파트 매매가	아파트 실거래가	일간지 및 경제지 6개	184,280
문태헌·김혜림 (2020)	토픽 분석, VECM	2010.06~2019.12	부산	아파트 매매가	아파트매매실거래가격지수	일간지 및 경제지 40개	6,778
박재수·이재수 (2021)	토픽 분석, ARIMA	2012.01~2018.12	전국	아파트 매매가	KB 아파트 매매가격지수	일간지 및 경제지 6개, 지상파 3개	99,427
서정석 외 (2021)	토픽 분석, ARDL-EC	2011.07~2018.12	전국	아파트 매매가, 아파트 거래량, 주택소비심리	아파트 실거래가, 아파트 매매거래량, 주택소비심리지수	포털사이트	640,445

제 3 절 연구의 차별성

부동산 비기초가격은 부동산 시장가격의 변동과 이탈을 이해하기 위해 고려해야 할 핵심 변수이며(김윤영, 2013), 과신(overconfidence)이나 기대(expectation) 같은 심리적인 요인과도 맞닿아 있는 영역이라 알려져 있다. 다만, 비기초가격의 주요 요인으로 제기되고 있는 시장참여자의 심리를 실증적으로 측정하기는 매우 어려운 문제이며(신진호·황수성, 2015), 그 때문에 비기초가격에 영향을 미치는 요인에 관한 연구는 아직 초기의 단계에 머물러있다.

본 연구는 부동산 비기초가격에 대한 언론매체의 힘을 과소평가해서는 안 된다는 관점, 그리고 그 힘을 오히려 이용할 수 있다는 관점에 기초한다. 앞서 기술한 바와 같이, 언론매체가 부동산 가격에 영향을 미친다는 것은 다수의 선행연구에서 확인되었다. 이들은 언론보도가 부동산 시장의 동향과 전망에 대한 정보를 제공함으로써 시장참여자들의 심리와 의사결정에 유의미한 영향을 미친다고 파악하였다. 그러나, 대다수의 관련 연구는 언론보도가 부동산 경기에 투기적 거품을 형성할 수 있다는 관점을 취하면서도, 정보나 심리와 같은 비정량적 요인으로 설명되는 비기초가격과 임대수익과 경기변동 등의 경제적 요인에 의해 설명되는 기초가격을 구분하지 않았다. 그 대신 부동산 가격 전체에 대한 언론매체의 영향을 파악하여, 분석을 통해 설명할 수 있는 영역과 설명할 수 없는 영역이 혼재되어있는 한계를 보였다.

이에 본 연구는 대중적 심리에 영향을 받고 또 영향을 미치는 신문기사 빅데이터를 일종의 설명변수로 활용하여, 부동산 비기초가격과 신문담론의 관계, 부동산 비기초가격과 대중심리의 관계를 탐색적으로 분석한다. 부동산 비기초가격과 부동산 신문기사에서 다루는 주제 및 내용의 관계를 검토하고, 부동산 신문기사를 활용하여 부동산 비기초가격을 분석하고 예측할 방안을 살펴보고자 한다. 이를 통해 부동산 비기초가격에 대한 이해도를 재고하고, 부동산 가격 안정화 정책에 대한 함의를 도출하고자 한다.

본 연구는 부동산 비기초가격을 별도의 변수로 구축하여, 부동산 비기초가격에 영향을 미치는 요인을 분석했다. 따라서 비기초가격의 존재 여부를 판단하는 데 주로 활용되는 공적분 검정법 대신, 비기초가격를 직접 추출하기에 용이한 내재가치접근법을 활용했다. Abraham and Hendershott(1995), Levin and Wright(1997), Muellbauer and Murphy(1997), Roche(2001), Bucchianeri(2011) 등의 기존 주택가격에 대한 연구와 같이, 주택매매가격을 종속변수로 하고 주택임대가격, 물가수준, 금리수준 등을 설명변수로 하는 회귀식을 추정하여 잔차항을 구하고, 해당 잔차항을 주택 비기초가격으로 가정했다. 구체적으로, Phillips and Hansen(1990)의 FM-OLS를 활용했다. 또한, 텍스트 마이닝(text mining)을 통해 부동산 신문기사 빅데이터를 정량화하여 시계열 자료로 구축했다. 토픽 분석(topic modeling)을 활용해 부동산 신문기사 주제 변수를 구축하고, 감성 분석(sentiment analysis)을 토대로 부동산 신문기사 논조를 지수화했다.

부동산 비기초가격과 부동산 신문기사의 영향 관계를 Granger 인과관계 분석, 충격반응분석, 분산분해분석 등을 통해 확인했다. 아울러, 부동산 신문기사 논조지수가 부동산 비기초가격 예측에 유용한지를 부동산 비기초가격 ARIMA 모형과 부동산 신문기사 논조지수를 외생변수로 포함한 부동산 비기초가격 ARIMAX 모형을 비교해 실증했다.

제 3 장 분석 자료 및 분석 방법

제 1 절 자료 구축

1) 부동산 비기초가격

먼저, 비기초가격을 산출하기 위해 기초가격에 영향을 미치는 시장 요소를 선정했다. 관련 연구를 참조하여 시장가격, 임대가격, 물가수준, 금리수준, 경제 성장률, 경기 순환성 등이 주택의 기초가격에 영향을 미친다고 가정했다(〈표3〉). 실질 시장가격은 한국부동산원에서 발표하는 아파트매매가격지수를 소비자물가지수로 나눠 사용했다. 실질임대가격은 주택의 사용 가치를 의미하며, 아파트전세가격지수를 소비자물가지수로 나눈 값을 사용했다. 물가수준은 인플레이션 헤지(hedge) 수단에 대한 선호를 반영하며, 본 연구는 소비자물가지수를 이에 대한 지표로 사용했다(김봉한, 2004; 오근엽 외, 2005; 김동환, 2015; 신진호·황수성, 2015; 임대봉, 2015; 이영수, 2019; 김상엽 외, 2020; 홍정의 외, 2022). 금리수준은 주택의 구매 비용과 적정 수익률에 영향을 미치는 요인으로, 예금기관 가계 주택담보대출금리(신규 취급)를 사용했다(조경준, 2013; 김경민, 2017; 이석원·정재호, 2017; 김경민, 2018; 이영수, 2019; 김상엽 외, 2020; 조한국·김영곤, 2020; 홍정의 외, 2022). 경제 성장률은 전반적인 거시경제의 규모에 해당하며, 산업생산지수(전산업)를 사용했다(김윤영, 2013; 신진호·황수성, 2015; 이석원·정재호, 2017; 이영수, 2019; 조무상·남주하, 2019). 마지막으로, 거시경제의 경기 순환성을 반영하기 위해 경기동행지수를 사용했다(홍정의 외, 2022). 전 자료의 기초통계는 〈표4〉와 같다.

〈표 3〉 FM-OLS 모형 변수

Variable	Code	Definition	Data Source	Period
실질시장가격	RHPI	서울시 아파트매매가격지수를 소비자물가지수로 나눈 값	한국부동산원 부동산통계시스템 (2021)	96개월 (2014.01~ 2021.12)
실질임대가격	RRPI	서울시 아파트전세가격지수를 소비자물가지수로 나눈 값		
물가수준	P	소비자물가지수	한국은행 경제통계시스템 (2021)	
금리수준	r	예금기관 가계 주택담보대출 가중평균금리 (신규 취급 기준)		
경제 성장률	N	농림, 어업을 제외한 전산업생산지수	통계청(2021)	
경기 순환성	Y	경기동행지수		

note) RHPI, RRPI, N, Y은 계절조정함.

〈표 4〉 FM-OLS 모형 변수별 기초통계

	Obs	Mean	std.dev	Max	Min
RHPI	96	1.131	0.272	1.729	0.829
RRPI	96	1.008	0.125	1.239	0.776
P	96	97.94	2.72	104.04	93.73
r	96	3.04	0.372	3.75	2.39
N	96	105.476	4.633	116.06	97.149
Y	96	107.568	6.534	119.345	95.5

앞선 자료를 바탕으로 시장가격에 대한 회귀식을 구축했다(식21). 실질시장가격($RHPI_t$), 실질임대가치($RRPI_t$), 물가수준(P_t), 금리수준(r_t), 경제 성장률(N_t), 경기 순환성(Y_t) 등을 기초가격에 대한 대리변수로 가정하고, 이외 시장가격에서 기초가격을 제외한 잔차항(ε_t)을 비기초가격이라고 가정했다. 이때 실질시장가격, 실질임대가치, 물가수준, 경제 성장률, 경기 순환성 변수는 자연로그를 취해 사용했고, 금리수준 변수는 원자료를 사용했다.

$$\begin{aligned}\ln RHPI_t = \beta_0 + \beta_1 \times \ln RHPI_{t-1} + \beta_2 \times \ln RRPI_t \\ + \beta_3 \times \ln P_t + \beta_4 \times r_t + \beta_5 \times \ln N_t \\ + \beta_6 \times \ln Y_t + \varepsilon_t\end{aligned} \tag{21}$$

위와 같은 모형을 토대로, FM-OLS(fully modified least squares) 방법을 통해 계수를 추정했다. FM-OLS는 변수들의 공적분 관계를 단일 방정식 체계에서 추정하는 분석 기법으로, ADF 검정이나 Johansen 검정 등의 공적분 추정법에 비해 변수 간의 관계를 더욱 정교하게 규명할 수 있다고 알려져 있다(이현재, 2001). FM-OLS 방법을 활용하기 이전에 시계열 자료에 대한 ADF 단위근 검정(Augmented Dickey-Fuller unit root test)을 시행했으며, 검정 결과 모든 자료에 단위근이 존재했다. 이에, 자료들 사이에 장기적 균형 관계인 공적분 관계가 존재하는지 검정하기 위해 Johansen 공적분 검정을 시행했고[5], 그 결과 5%의 유의수준에서 공적분 벡터가 존재하지 않는다는 가설이 기각됐다(〈표5〉). 이와 같은 결과에 기초해, 공적분 관계가 존재하여 가성 회귀의 문제가 없음을 확인했다.

〈표 5〉 Johansen 공적분 검정 결과

r	AIC	SC
0	-12.839*	-9.727*
1	-13.158*	-9.451
2	-13.204*	-9.168
3	-13.161	-8.800

confidence level: *0.05 lower
VAR lag: 1

FM-OLS를 사용한 회귀식 추정 결과는 〈표5〉와 같으며, 회귀계수는 모두 유의하게 나타났다. 실질시장가격($\ln RHPI_t$)은 전기의 실질시장가격($\ln RHPI_{t-1}$)과 실질임대가격($\ln RRPI_t$), 물가수준($\ln P_t$)과 양(+)의 관계를 보였으며, 이자율(r_t)과 음(-)의 관계를 보여 예상에 부합한 결과가 나타났다. 경제 성장률($\ln N_t$)과 경기 순환성과($\ln Y_t$)는 음(-)의 관계를 보였다(김윤영, 2013; 신진호·황수성, 2015; 이석원·정재호, 2017; 조무상·남주하, 2019). 다만, 회귀계수의 부호는 계량 방법이나 분석 대상 및 분석 기간에 따라 변동할 수 있으며, 실제로 선행연구마다 상이한 결괏값을 보

5) 시차 검정(lag test) 결과를 토대로, VAR 시차를 1로 설정했다. SC(Schwarz criterion) 기준 결괏값을 준용했다.

였다는 점을 감안할 필요가 있다.

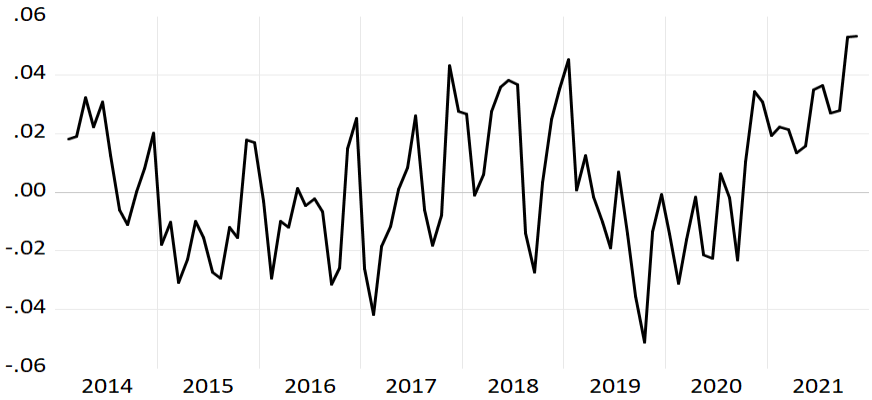
〈표 6〉 FM-OLS 회귀식 추정 결과

Variable	Coefficient	Std.Error	t-statistic	p-value
$\ln RHPI_{t-1}$	0.666	0.046	14.562	0.000*
$\ln RRPI_t$	0.082	0.027	3.075	0.003*
$\ln P_t$	4.613	1.010	-16.118	0.000*
r_t	-0.083	0.004	-18.965	0.000*
$\ln N_t$	-0.527	0.115	-4.565	0.000*
$\ln Y_t$	-0.441	0.197	-2.244	0.027*
상수항	-16.284	1.010	-16.118	0.000*
R-squared			0.995	
Adjusted R-squared			0.994	
S.E. of regression			0.019	

confidence level: *0.05 lower
lag: 1

최종적인 회귀식은 식(22)이며, 그에 따른 회귀식 잔차항은 [그림2]와 같다. 앞선 이론적 가정과 수식에 의거하면, 회귀식 잔차항이 양(+)일 경우 시장가격이 기초가격을 넘어서는 것으로, 음(-)일 경우 시장가격이 기초가격을 밑도는 것으로 이해할 수 있다. 회귀식 잔차항은 전체 시기에 걸쳐 등락을 반복하는 양상을 보였으며, 이는 잔차항이 단기적이고 변동적인 움직임을 보인다는 것을 짐작게 한다. 더불어, 잔자항은 2020년 후반부터 꾸준히 양(+)의 값을 보였는데, 이는 해당 시기부터 시장가격은 계속해서 기초가격을 웃돌았음을 암시한다. 본 연구는 해당 잔차를 비기초가격 변수로 간주해, 추후의 분석에 활용했다.

$$\begin{aligned}\ln\widehat{RHPI}_t = & 0.666 \times lnRHPI_{t-1} + 0.082 \times lnRRPI_t \\ & + 4.613 \times lnP_t - 0.083 \times r_t - 0.527 \times lnN_t \\ & - 0.441 \times lnY_t - 16.284\end{aligned} \tag{22}$$



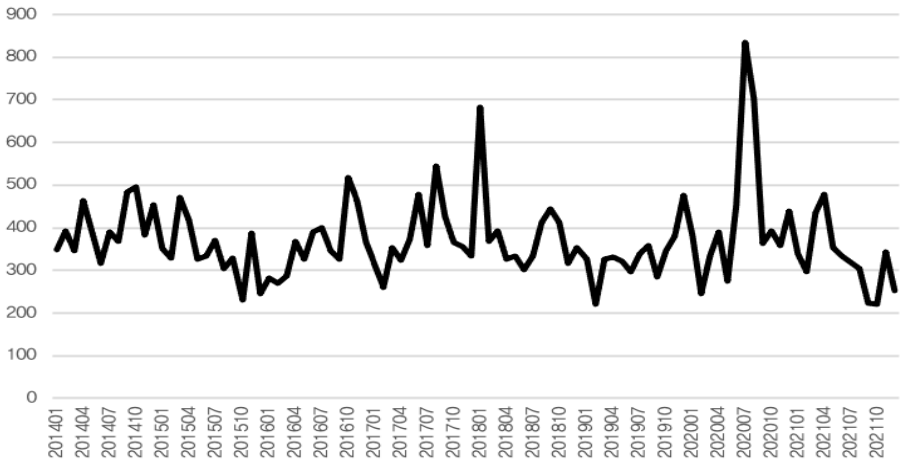
[그림 2] 부동산 비기초가격(잔자항) 추정 결과

2) 부동산 신문기사의 주제

다음으로, 부동산 신문기사 자료를 구축하기 위해 한국언론진흥재단의 뉴스포털사이트 'KINDS(www.kinds.or.kr)'를 이용했다. 2014년 1월부터 2021년 12월까지 서울 부동산에 관한 중앙일보, 동아일보, 경향신문, 한겨레신문, 매일경제신문, 한국경제신문의 신문기사를 크롤링했다(구정우·조성권, 2017). 자료 분석의 균형을 위해 보수, 진보, 경제지를 고루 포함했으며, 경제 신문은 이념적으로 보수신문과 중복될 가능성이 있지만 부동산에 대한 기사량과 보도 비중을 고려해 분석에 포함했다. 단어 '서울'을 포함하면서 동시에 단어 '부동산' '주택' '집값' '아파트' 중 한 단어 이상을 포함한 신문기사를 수집했으며, 중복기사는 제외하고 광고기사는 포함했다. 결과적으로 총 35,542개의 신문기사를 수집했다(〈표7〉).

〈표 7〉 부동산 신문기사 자료 수집 방법

Criteria		Content
검색엔진		한국언론재단 기사검색사이트 KINDS
검색대상신문	종합일간지	중앙일보, 동아일보, 경향신문, 한겨레신문
	경제일간지	매일경제, 한국경제
검색기간		96개월 (2014.01. ~ 2021.12.)
검색어		(부동산 or 주택 or 집값 or 아파트) and 서울
기사 수		35,542



[그림 3] 부동산 신문기사 자료 수집 결과

수집한 다량의 신문기사 자료를 분석하기 위해 빅데이터 분석기법 중 하나인 텍스트 마이닝(text mining) 기법을 사용했다. 텍스트 마이닝이란 여러 문서를 한 번에 처리하는 통계적 텍스트 처리 기법을 의미한다. 텍스트 데이터를 대용량의 정형화된 텍스트 집합(large and structured set of texts)으로, 소위 말뭉치(corpus)로 간주하고, 그 안에서 텍스트 간의 관계를 찾아내는 체계적 분석 도구다(백영민, 2020).

텍스트 분석을 위해 비정형의 온라인 신문기사를 정형화하는 텍스트 사전 처리(preprocessing) 과정을 거쳤다. 먼저, 말뭉치를 형태소 단위로 토큰화(tokenization)하고, 기호나 숫자와 같은 불용어(stopword)를 제거했다. 다음으로 '전셋값'과 '전세금', '종부세'와 '종합부동산세' 등의 표현 방법이 다르지만, 의미

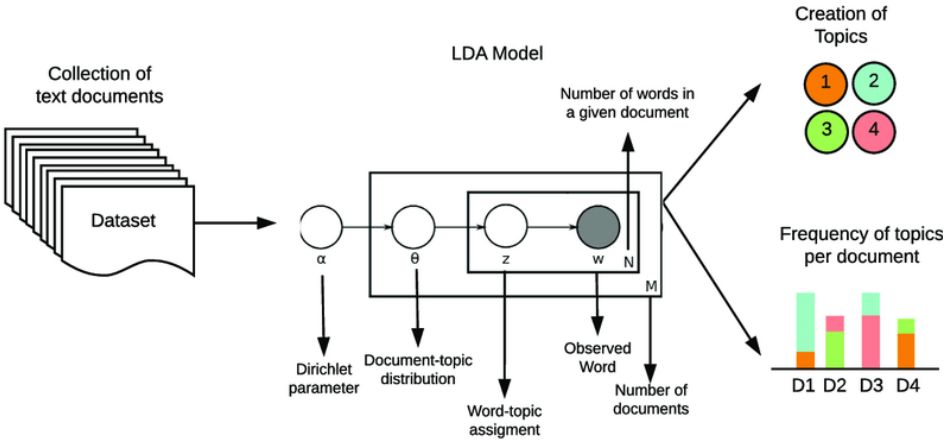
가 같은 단어를 같은 단어로 정규화(normalization)했다. 그리고, 단어 빈도-역문서 빈도(term frequency-inverse document frequency, TF-IDF)를 적용해 모든 문서에 빈번히 나타난 단어를 걸러내고 전체 단어 중 핵심 단어만을 추출했다(정민경·최규완, 2021).

텍스트 마이닝 과정을 통해 추출한 단어 중에서 기사의 전체적인 흐름을 반영하고 있는 단어를 찾아내기 위해 토픽 모델링(topic modeling) 분석을 시행했다. 토픽 모델링이란 문서에 잠재되어 있다고 가정한 토픽의 등장 확률을 추정하는 통계적 텍스트 처리기법이며(Blei, 2012), 토픽은 문서 내에서 함께 등장할 확률이 높은 단어들의 집합을 뜻한다. 즉, 토픽 모델링은 텍스트 데이터 단어들을 동시 사용 패턴을 토대로 묶어내는 통계기법으로, 문서의 의미론적 구조를 반영하고 있는 주제를 찾아내기 위해 활용하는 분석 방법이다(방한솔·문호석, 2019).

본 연구는 확률적 토픽 모델링 가운데 가장 대표적인 모형인 LDA 토픽 모델링(Latent Dirichlet Allocation topic modeling)을 Gibbs sampling 기법을 적용해 사용했다. LDA는 단어들의 집합이 공통된 주제를 포함한다는 가정하에, 단어들이 주제에 포함될 확률을 조건부 확률분포로 계산하여 확률이 높은 단어들의 집합을 추출하는 통계적 모형이다(Griffiths and Steyvers, 2004; Blei, 2012; 조성배 외, 2018; 이효섭·조대연, 2020; 김경아·김선희, 2021). LDA에서 M은 말뭉치를, N은 문서를, w는 단어를, z는 토픽을, k는 토픽의 수, α는 외생적 모수를 의미한다([그림4]). 이때 잠재 토픽 수 k는 연구자가 판단하여 설정하며, 본문은 R의 ldatuning 패키지의 FindTopicsNumber 함수를 통해 적정 토픽 개수를 20개로 결정했다. 또한, 토픽 z을 생성하는 모수 θ는 외생적 모수인 α에 의해 생성되며, 토픽의 수 k만큼의 차원을 갖는 디리클레 랜덤 변수(k-dimensional Dirichlet random variable)이다(백영민, 2020). 모수 θ에 의한 토픽 z_n 생성 확률은 식(23)과 같으며, 토픽 z_n 생성 확률에 따른 단어 w_n 생성 확률은 식(24)과 같다.

$$p(z_1, \cdots, z_n) = \int p(\theta) \prod_{n=1}^{N} p(z_n|\theta)) d\theta \tag{23}$$

$$p(w,z) = \int p(\theta)(\prod_{n=1}^{N} p(z_n|\theta)p(w_n|z_n))d\theta \tag{24}$$



Buenano-Fermandez D. et al., (2020).

[그림 4] LDA 토픽 모델링 알고리즘 구조

LDA를 통해 추출한 토픽은 〈표8〉과 같다. 구체적인 수준에서 신문담론의 구성을 파악하기 위해, 토픽마다 어떤 단어가 가장 빈번하게 등장하는지 확인하고 토픽과 연관된 기사 15~20편의 본문 내용을 검토했다. 분석 결과를 바탕으로 토픽의 제목을 선정했으며, 직관적인 이해를 돕기 위해 20개의 토픽(소주제)을 7개의 영역(대주제)으로 재분류했다. 본 연구는 7개의 영역(부동산 가격, 부동산 정책, 부동산 공급, 부동산 투자, 부동산 광고, 생활환경, 지역시장)이 부동산 신문기사의 주요 골자를 이룬다고 가정했다.

문서의 토픽별 빈도수(frequency of topics per document, [그림4] 참고)를 영역별로 재합산하여 시계열 자료로 구축했으며, 본 연구는 해당 자료를 부동산 신문기사 자료로 사용했다. 부동산 신문기사 자료를 도출한 결과, 부동산 정책 기사와 부동산 가격 기사가 가장 높은 비중을 차지했으며, 부동산 광고 기사, 생활환경 기사, 부동산 공급 기사 순으로 뒤를 이었다(〈표9〉. 부동산 정책 기사의 경우 부동산 가격에 대한 정책의 발표 및 시행 시기에 맞춰 변동하는 양상을 보였으

며, 부동산 공급과 생활환경 기사는 전 시기에 걸쳐 비교적 일관된 추세를 가졌다.

〈표 8〉 LDA 토픽 분석 결과

Topics		Key Words (Top 8)							
부동산 가격	부동산 가격 추세	거래	매물	가격	아파트	부동산	호가	단지	중개
	강남권 아파트 가격	아파트	강남구	지역	가격	시세	정도	부동산	서초구
	수도권 아파트 가격	아파트	상승	가격	지역	강남구	하락	매매	평균
	전세금 동향	전세	월세	전셋값	전세금	주택	입주	아파트	수요
	상가건물 임대	건물	상가	빌딩	임대	보증금	대지	연면적	상권
부동산 정책	부동산 정부 정책	대통령	후보	의원	정부	정책	국민	청와대	장관
	부동산 대책	정부	부동산	시장	집값	대책	규제	주택	정책
	부동산 과세	주택	세금	양도세	종부세	공시	기준	보유	부담
	부동산 대출	대출	금리	주택	규제	경매	DTI	담보	은행
부동산 공급	재개발 재건축	단지	재건축	사업	아파트	조합	조합원	건축	추진
	부동산 청약	분양	분양가	청약	아파트	공급	평균	전매	물량
	택지 개발	주택	공급	지구	지역	계획	사업	토지	지정
부동산 투자	부동산 투자	투자	부동산	자산	수익	주식	펀드	상품	시장
	소형 부동산 투자	오피스텔	분양	수익	부동산	수요	투자	임대	소형
부동산 광고	주거생활형 광고	아파트	단지	예정	이용	가능	세대	지상	위치
	대단지 아파트 분양	가구	분양	아파트	공급	단지	예정	물량	전용
생활환경	교통	도시	개통	예정	지구	동탄	분양	단지	지역
	소비문화	호텔	브랜드	고급	운영	서비스	매장	고객	성공
	오피스타운	기업	사업	회사	회장	센터	금융	지원	관리
지역시장	수도권 지역명	분당	지역	판교	용인	수도권	경기	신도시	일산

〈표 9〉 부동산 신문기사 자료의 영역별 기초통계

	Obs	Mean	std.dev	Max	Min
부동산 가격	96	93.588	24.514	184.975	47.043
부동산 정책	96	71.766	62.729	406.432	10.730
부동산 공급	96	52.483	18.319	118.864	23.429
부동산 투자	96	30.652	12.054	84.400	12.504
부동산 광고	96	61.285	21.097	124.903	20.818
생활환경	96	53.417	16.489	125.495	27.270
지역시장	96	7.039	2.406	15.844	2.937

본 연구는 부동산 신문기사 자료의 양적 증감을 각각의 주제에 대한 관심과 기대심리의 증감과 연관 지어 해석했다. 예컨대, 부동산 정책 기사 변수의 증감을

부동산 정책의 시행으로, 그리고 그에 따른 부동산 정책에 대한 대중적 관심과 심리적 영향의 증감으로 가정했다. 즉, 각각의 주제에 대한 기사가 증가했다는 것을 해당 주제에 대한 프레임(frame)이 적극적으로 조성되었던 것으로, 관련 주제에 관한 언론과 대중의 이목이 집중되었던 것으로 간주했다.

3) 부동산 신문기사의 논조

부동산 신문기사의 논조지수를 산출하기 위해, 감성분석(sentiment analysis)을 시행하였다. 먼저, 감성사전을 작성하였다. 토픽 분석(topic analysis)을 통해 신문기사를 부동산 가격·부동산 정책·부동산 공급·부동산 투자·부동산 광고·생활환경·지역시장 등의 7개 주제로 구분했으며, 키워드 네트워크 분석(keyword network analysis)을 통해 각각의 주제에서 빈번히 등장하는 키워드와 이들의 연결망을 확인했다.

앞선 결과를 토대로, 감성과 관련된 주요 속성(키워드)과 표현을 추출해 감성사전을 작성하였다. 주제별로 긍정과 부정의 감성을 표현하는 속성어를 수집했다. 긍정 속성어와 부정 속성어를 각각 200여 개씩 선정하여 감성사전으로 작성했다(〈표10〉 참고). 감성사전에 포함된 긍정어에 대해서는 +1이라는 극성값($sen_{positive}$)이, 부정어에 대해서는 -1이라는 극성값($sen_{negative}$)이 태깅되도록 설정했다.

〈표 10〉 부동산 신문기사 감성사전 예시

Topic	Positive (+1)	Negative (-1)
부동산 가격	상승, 상승세, 강세, 인상	하락, 하락세, 약세, 인하
	오름, 오름세	내림, 내림세
	금값, 최대치	헐값, 최저치
	급상승, 급등, 폭등, 폭증	급하락, 급락, 폭락, 거품
	회복, 반등, 만회	부진, 주춤, 삐걱
부동산 정책	금리 인하, 금리 하락	금리 인상, 금리 상승
	규제 완화	규제 강화 강력조치
	세제 인하, 세제 완화	세제 인상, 세제 강화
	세금 인하	세금 인상
		제재, 적발, 불허, 단속
부동산 공급	증가, 증가세	감소, 감소세
	활성화, 활기	불안, 위축
부동산 투자	호황, 훈풍, 활발, 열기	빙하기, 냉동, 냉각, 꽁꽁
	회복	둔화
부동산 광고	역대 최대, 대란	코웃음, 시들
	1위	
생활환경	편한, 편리한	불편, 불편한, 불만
	화려, 화려한	낙후, 낙후된

감성사전을 바탕으로, 개별 부동산 신문기사(RN)의 논조를 긍정·부정·중립으로 구분했다. 태깅된 값의 총합이 0보다 크면 긍정($RN_{positive}$)으로, 0보다 작으면 부정($RN_{negative}$)으로, 0이면 중립($RN_{neutral}$)으로 분류되도록 했다(식25). 개별 신문기사를 논조별로 분류한 결과, 14,670개 기사(41.28%)는 긍정기사로, 5,443개 기사(15.31%)는 부정기사로, 그리고 15,429개 기사(43.41%)는 중립기사로 분류되었다(〈표10〉). 분석 시기인 2014년부터 2021년은 전반적으로 주택 가격 상승하던 시기이기 때문에, 긍정기사가 부정기사가 많았던 것으로 추측된다.

$$RN = \begin{cases} RN_{positive} & \text{if } \sum_{i=1}^{n} sen_{positive} + \sum_{i=1}^{n} sen_{negative} > 0 \\ RN_{neutral} & \text{if } \sum_{i=1}^{n} sen_{positive} + \sum_{i=1}^{n} sen_{negative} = 0 \\ RN_{negative} & \text{if } \sum_{i=1}^{n} sen_{positive} + \sum_{i=1}^{n} sen_{negative} < 0 \end{cases} \tag{25}$$

〈표 11〉 부동산 신문기사 감성분석 결과

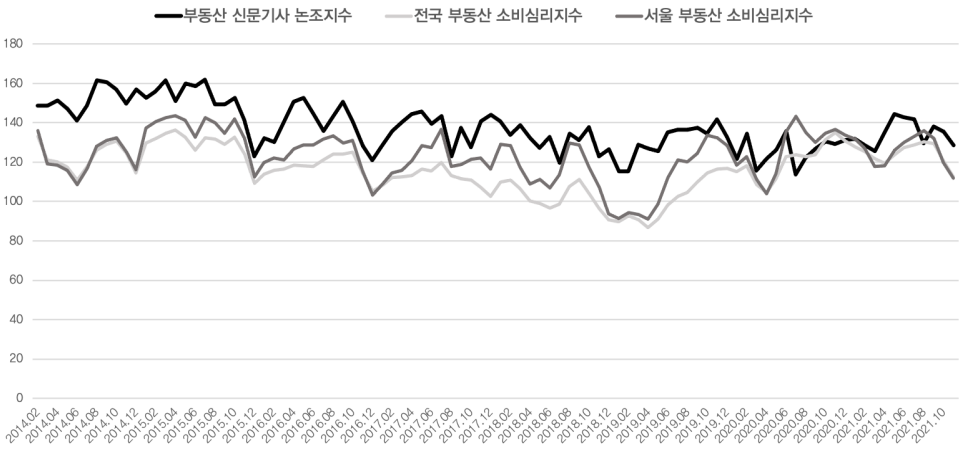
Total		Sentimental		Neutral
		Positive	Negative	
신문기사	35,542 (100%)	14,670 (41.28%)	5,443 (15.31%)	15,429 (43.41%)

개별 신문기사의 분류 결과를 종합하여, 월별 신문기사 논조지수(*RNSI*)를 산출했다(식26). 해당 월에 긍정적인 논조의 신문기사가 많을수록 논조지수가 100보다 크도록, 부정적인 논조의 신문기사가 많을수록 논조지수가 100보다 작도록 표현했다. 부동산 신문기사 논조지수는 0에서 200 사이의 값을 갖도록 설정했다.

$$RNSI_{month} = \left\{ \frac{\left(\sum_{i=1}^{n} RN_{positive}(i) - \sum_{i=1}^{n} RN_{-ative}(i) \right)}{\sum_{i=1}^{n} RN(i)} + 1 \right\} * 100 \qquad (26)$$

산출한 부동산 신문기사 논조지수의 타당성과 실효성을 검증하기 위해, 부동산 소비심리지수와 비교했다. 국토연구원에서 매월 '부동산 시장 소비자 심리조사'를 통해 발표하는 전국 주택시장 소비심리지수·서울 주택시장 소비심리지수와 비교분석했다. 분석 결과, 전반적으로 변곡점, 기울기, 등락 시기 등이 유사하게 나타났다([그림5]).

부동산 신문기사 논조지수와 부동산 소비심리지수에 대한 상관분석(correlation analysis)을 시행한 결과, 모든 상관계수가 0.5~0.6($p<0.01$)로 비교적 높게 나타났다. 이는 비정형 텍스트 데이터로 구축한 부동산 신문기사 논조지수가 정형 데이터인 부동산 소비심리지수와 높은 상관관계를 가지는 유의한 데이터임을 의미한다(〈표12〉).



[그림 5] 부동산 신문기사 논조지수 추정 결과

〈표 12〉 부동산 신문기사 논조지수와 관련 정형데이터 간 상관관계 분석 결과

	전국 주택시장 소비심리지수	서울 주택시장 소비심리지수
부동산 신문기사 논조시수	0.5777*** (0.000)	0.5615*** (0.000)

confidence level: *0.10 lower, **0.05 lower, ***0.01 lower

제 2 절 분석 방법

1) Granger 인과관계 분석

부동산 비기초가격과 부동산 신문기사의 인과관계 영향력 및 방향성을 검정하기 위해 Granger 인과관계 분석을 시행했다. Granger 인과관계 분석을 통해 부동산 비기초가격과 부동산 신문기사 사이에 인과관계가 존재하는지 확인하고, 신문기사 주제별 영향력을 정량화해 파악했다.

Granger 인과관계 분석은 만약 다른 조건이 불변일 때, X의 과거값을 사용할 경우에 그렇지 않을 경우보다 Y값을 더욱 정확하게 예측할 수 있다면 X가 Y에 영향을 미친다고 평가한다. 이를 X가 Y를 Granger 인과(cause) 한다고 표현한

다. Granger 인과관계 분석 결과는 각각의 변수 X와 Y의 예측에 적합한 정보가 단지 변수의 시계열 자료 속에만 포함되어 있다고 가정하고, 변수의 관계는 X가 Y에 영향을 미치거나, Y가 X에 영향을 미치거나, X와 Y가 서로에 영향을 미치거나, X와 Y가 독립적인 4가지 가능성으로 표현된다고 간주한다(박승록, 2020). Granger 인과관계 분석의 일반형식을 표현하면 다음과 같다(식27-28).

$$X_t = \mu_2 + \sum_{i=1}^{n} \lambda_i Y_{t-i} + \sum_{j=1}^{n} \delta_j X_{t-j} + \epsilon_{2t} \tag{27}$$

$$Y_t = \mu_1 + \sum_{i=1}^{n} \alpha_i Y_{t-i} + \sum_{j=1}^{n} \beta_j X_{t-j} + \epsilon_{1t} \tag{28}$$

Granger 인과관계 분석을 위해 먼저 사용변수의 안정성을 검정했다. 본 연구는 변수의 안정성을 ADF(Augmented Dickey-Fuller) 검정을 토대로 확인했다. 불안정한 시계열은 1차 차분 과정을 거쳐 안정적인 시계열로 변환하고, 최장 시차를 5로 설정하여 Granger 인과관계를 분석했다.

2) VECM 모형 추정

부동산 비기초가격과 부동산 신문기사의 동태적 관계에 대한 더욱 구체적인 분석을 위해 Engle and Granger(1987)의 VECM(Vector Error Correction Model) 모형을 활용했다. VECM은 VAR에 오차수정항을 도입한 모형이다. 일반적으로 시계열 변수에 단위근이 있으면 차분하여 안정변수로 이용하나, 변수 사이에 공적분이 존재하는 경우 차분변환이 두 변수 사이의 중요한 장기적 정보를 상실케 할 수 있다는 의식하에 두 비정상 시계열을 선형 결합한 모형이다. 즉, VECM은 시계열 사이에 공적분 관계가 있음에도 불구하고 차분하여 변수들 사이의 장기적 관계에 관한 정보를 잃어버리는 문제를 개선한 모형으로, 개별 변수가 불안정하더라도 변수 간에 공적분 관계가 존재한다면 이들 사이에 장기적 균형관계가 존재한다고 가정하는 모형이다(김동환, 2015). VECM을 활용하면 변수 간에

공적분 벡터를 바탕으로 장기적 정보를 보유한 채 분석할 수 있으며(이우민, 2019), 그 일반형식은 다음과 같다(식29-31),

$$\Delta Y_t = \alpha + \gamma \hat{\epsilon}_{t-1} + \sum_{i=1}^{n} (\delta_i \Delta X_{t-i}) + \sum_{i=1}^{n} (\theta_i \Delta Y_{t-1}) + \upsilon_t \tag{29}$$

$$\hat{\epsilon}_{t-1} = Y_{t-1} - \hat{\beta} X_{t-1} \tag{30}$$

$$\hat{\epsilon}_{t-1} \sim I(0) \tag{31}$$

VECM 모형을 이용하여 부동산 비기초가격과 부동산 신문기사의 관계를 파악하기 위해서는 변수들이 안정적인 시계열인지 단위근 검정을 거쳐 확인해야 하지만, 이는 앞서 Granger 인과관계 분석 과정에서 수행했으므로 생략했다. 또한, VECM 모형 사용 가능 여부를 판별하기 위해 변수 간에 공적분 관계가 존재하는지 검정해야 한다. 만약 시계열 변수가 단위근을 가지고 있어 안정적이지 않지만, 1차 차분 후에 안정화되고 변수 사이에 공적분이 존재한다면 VECM을 적용하여 인과성 여부를 판단할 수 있다(김선재·이영화, 2014). 본 연구는 Johansen 공적분 검정을 활용해 공적분 존재 여부를 확인했으며, 시차(lag) 검정과 Johansen 공적분 검정 과정을 결과를 토대로 VECM 모형을 추정했다.

VECM 모형을 기반으로, 부동산 비기초가격 변수와 부동산 신문기사 변수 간의 연관관계와 파급효과를 분석했다. VECM과 같은 VAR 기반 모형의 특징은 충격반응분석(impulse response function)과 분산분해분석(variance decomposition analysis)을 통해 변수 간의 장기적 관계를 파악할 수 있다는 점이다. 충격반응분석을 통해 특정 변수의 단위 변화(충격)가 모형 내의 다른 내생변수에 미치는 동태적 반응의 방향성과 지속성을 파악할 수 있고(김동환, 2015), 분산분해분석을 통해 시간의 변화에 따른 특정 변수의 변동을 과거 자기의 분산과 다른 내생변수의 분산으로 분해하여 변수 간의 상대적 중요성을 측정할 수 있다. 본 연구는 VECM 모형을 기초로 충격반응분석을 시행해 부동산 비기초가격과 부동산 신문기사의 충격이 서로에 미치는 영향을 관측했으며, 분산분해분석을 거쳐

부동산 비기초가격의 변동에 대한 부동산 신문기사의 상대적 기여도를 파악했다.

3) ARIMA 예측 유용성 분석

AR(Auto Regressive)은 주로 시계열 데이터 예측에 사용되는 분석 방법으로, 이전의 데이터가 현재의 데이터에 영향을 미치는 자기회귀 요소를 고려하는 모형이다. 해당 모형은 이전 데이터들의 가중치(weight)를 조정하여 현재의 데이터를 예측한다. AR(p) 모형에서 p는 모형의 차수를 의미하며, 현재 시점 데이터(Y_t)는 상수항(c), 이전 시점 데이터에 대한 가중치인 자기회귀계수($\phi_1 \sim \phi_p$), 백색잡음(ε_t) 등으로 구성된다(식32).

$$\begin{aligned} Y_t &= c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \cdots + \phi_p y_{t-p} + \varepsilon_t \\ &= \sum_{j=1}^{p} \phi_j y_{t-j} + \varepsilon_t \end{aligned} \tag{32}$$

MA(Moving Average) 모형은 현재의 데이터가 이전 오차항의 가중합(weighted sum)에 영향을 받는다고 가정한다. MA(q) 모형에서 q는 모형 차수를 나타내고, 현재 시점의 데이터(Y_t)는 백색잡음(ε_t), 이동평균계수($\theta_1 \sim \theta_q$)의 합으로 이루어진다(식33).

$$\begin{aligned} Y_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \cdots + \theta_q \varepsilon_{t-q} \\ &= \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \end{aligned} \tag{33}$$

만일 시계열 데이터(Y)가 자기회귀 요소와 이동평균 요소를 동시에 가지고 있다면, 해당 데이터는 ARMA(Auto Regressive Moving Average) 모형을 따른다. ARMA(p,q)는 AR 모형과 MA 모형을 동시에 포함하는 모형으로, 시계열 데이터의 자기 상관성과 이전 오차항의 영향을 모두 고려한다. 곧, ARMA는 과거값과

과거 예측오차를 통해 현재값을 설명하는 모형이다. ARMA(p,q) 모형에서 p와 q는 각각 AR 모형의 차수와 MA 모형의 차수이며, 해당 모형에서 현재 시점 데이터(Y_t)는 상수항(c), 자기회귀계수($\phi_1 \sim \phi_p$), 이동평균계수($\theta_1 \sim \theta_q$), 백색잡음(ε_t) 등으로 구성된다(식34).

$$\begin{aligned} Y_t &= c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \\ &= c + \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \end{aligned} \tag{34}$$

한편, 대부분의 부동산 시계열 데이터에는 추세, 계절성, 주기 등과 같은 성분들이 내재하고 있다. 이처럼 불안정(non-stationary) 시계열 데이터를 분석하는 경우, ARMA 모형에 차분(difference) 과정을 포함한 모형인 ARIMA(Auto Regressive Integrated Moving Average)를 활용하게 된다. ARIMA는 불안정 시계열 자료를 차분하여 안정적(stationary) 시계열 자료로 변환해 사용하는 모형이며(식35), ARIMA(p,d,q) 모형에서 p는 AR 모형의 차수, d는 차분 차수, q는 MA 모형의 차수를 의미한다.

$$\Delta^d Y_t \equiv c + \sum_{j=1}^{p} \phi_j \Delta^d y_{t-j} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{35}$$

본 연구는 부동산 신문기사 논조지수의 예측 유용성을 검증하기 위해 ARIMAX(Auto Regressive Integrated Moving Average with eXogenous variables) 모형을 적용했다. ARIMAX는 ARIMA에 외생변수(exogenous variable)를 추가한 모형으로, 시계열 데이터 외부에서 발생하는 변화를 고려해 보다 정확한 예측을 수행하는 모형이다(이종민 외, 2017). ARIMAX(p,d,q) with k exogenous variables 모형에서 현재 시점 데이터(Y_t)는 상수항(c), 자기회귀계수($\phi_1 \sim \phi_p$), 이동평균계수($\theta_1 \sim \theta_q$), 백색잡음(ε_t), 그리고 외생변수($X_{1_t} \sim X_{k_t}$)로 구성된다.

$$\Delta^d Y_t \equiv c + \sum_{j=1}^{p} \phi_j \Delta^d y_{t-j} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \sum_{m=1}^{k} \beta_m X_{m_t} + \varepsilon_t \tag{36}$$

예측 유용성을 분석하기 위해 ARIMA 모형과 ARIMAX 모형의 RMSE(Root Mean Square Error)와 MAE(Mean Absolute Error) 값을 비교했다. 먼저 부동산 비기초가격에 대한 최적 ARIMA 모형을 추정하고, 해당 모형의 RMSE와 MAE를 구했다. 그리고, 부동산 신문기사 논조지수를 외생변수로 투입한 부동산 비기초가격의 최적 ARIMAX 모형을 재구축하고, 해당 모형의 RMSE와 MAE를 산출한다. ARIMA 모형과 ARIMAX 모형의 예측오차를 비교해, 부동산 신문기사 논조지수의 예측 유용성을 검증했다.

$$\sqrt{\frac{\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2}{n}} \tag{37}$$

$$\frac{1}{n}\sum_{t=1}^{n}|Y_t - \hat{Y}_t| \tag{38}$$

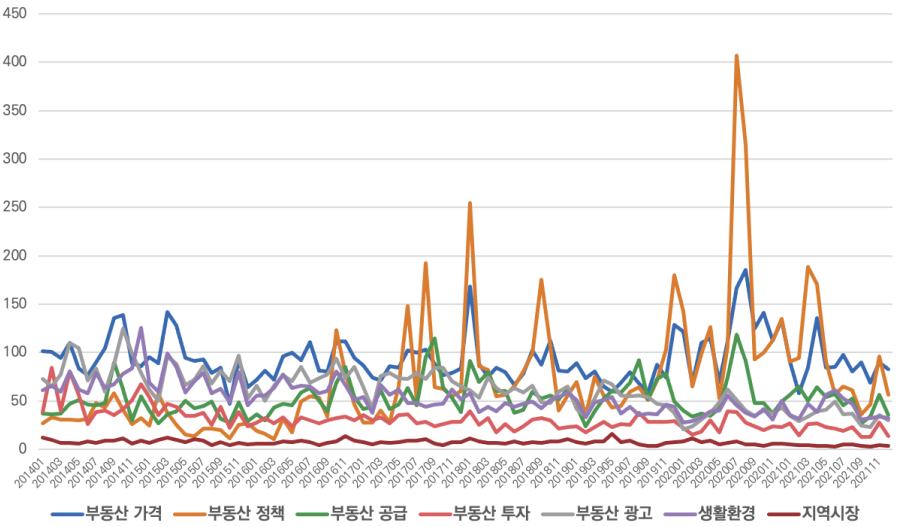
이때 RMSE와 MAE는 모형의 예측값(Y_t)과 실제 관측값($\hat{Y}_t$)의 차이를 측정하는 지표이다. RMSE는 예측값과 관측값의 차이를 제곱하여 평균을 구하고 제곱근을 취한 값이며(식37), MAE는 예측값(Y_t)과 관측값($\hat{Y}_t$)의 차이를 절댓값으로 변환하여 평균한 값이다(식38). 두 지표 모두 작은 값이 더 좋은 결과를 뜻한다(박재수·이재수, 2021).

제 4 장 부동산 신문담론의 형성과 변환

제3장 제1절에서 부동산 신문담론의 전반적인 구성을 간략히 살펴보았다. 이어 본 장에서는 기사 게시 시기의 사회적·경제적·정치적 맥락에 따른 부동산 신문담론의 변화 흐름을 포착했다. 이를 통해 부동산 신문담론 자료에 당시의 사회적 여론이 압축되어 있는지, 당시의 경제적 상황과 정치적 추이가 반영되어 있는지 확인하고자 했다.

제 1 절 부동산 신문담론의 형성

신문담론의 흐름은 [그림6]과 같다. 이때 x축은 신문기사 발행 연월을, y축을 신문기사 부수를 의미하며 y값이 클수록 전체 신문기사에서 해당 영역 및 주제가 차지하는 비중이 크다고 말할 수 있다(구정우·조성권, 2017).



[그림 6] 부동산 신문기사 영역별 추이

서울 부동산 신문담론의 전반적 추이를 살펴본 결과, 첫째 '부동산 가격'은 전 시기에 걸쳐 꾸준히 보도되었던 영역으로 보였다. 둘째, '부동산 정책'과 '부동산 공급'은 부동산 정책이 강화된 시기에 주목받은 것으로 나타났다. 셋째, '부동산 광고'나 '생활환경'은 앞선 부동산 정책 영역과 반대의 흐름을 보였다. 부동산 관련 규제가 강화된 시기에는 정책이나 공급 관련 기사가, 완화된 시기에는 광고성 기사가 추세적 상승을 지속한 것으로 나타났다.

1) 부동산 가격

서울 아파트 매매가격은 2014년부터 상승세로 전환되었고 2017년부터는 상승 속도가 빨라졌다. 한국부동산원에 따르면 서울은 2014년 2.0%, 2015년 6.7%, 2016년 3.3%, 2018년 8.0%, 2019년 1.1%, 2020년 3.0%, 2021년 8.0%, 상승했다. 더불어 2016년부터는 지방의 아파트 매매가격과도 상승률에 있어서 격차가 벌어져, 서울을 중심으로 한 아파트 가격 차별화 현상이 나타났다(윤성훈, 2020).

'부동산 가격' 영역은 이러한 가격 추세에 발맞춰 보도되었다. 가격 상승기에 접어들면서 '수도권 아파트 가격', '부동산 가격 추세'에 관한 기사가 적극 보도되었다. 한국 사회에서 부동산 가격 상승은 경제적 현상이면서 동시에 상당 부분 심리적인 현상이기 때문에(강준만, 2006), 고발성 기사이건 또는 정보성 기사이건 서울 인근 부동산값이 유달리 폭등했다는 사실을 상세하게 보도하면서 이에 대한 위상, 당위, 욕망은 더욱 강화되었을 것으로 추측된다.

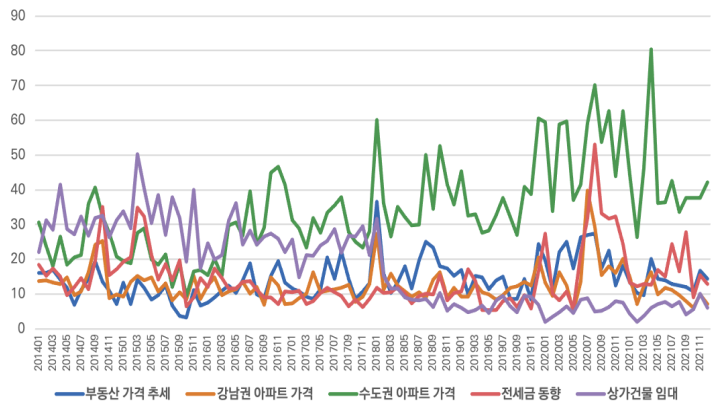
'부동산 가격' 영역은 부동산 정책 기조와도 밀접한 연관을 보였다. 특히 2018년부터 '수도권 아파트 가격', '부동산 가격 추세', '강남권 아파트 가격', '전세금 동향'과 관련한 기사는 '부동산 정책' 영역에 관한 기사와 함께 급증하는 양상을 보였다. 당시의 부동산 정책은 주로 부동산 가격 상승에 후행적으로 대응하는 방식으로 시행되었기에, 부동산 정책이 부동산 가격에 종속되었던 상황이 어느 정도 반영된 것으로 해석된다.

2) 부동산 정책

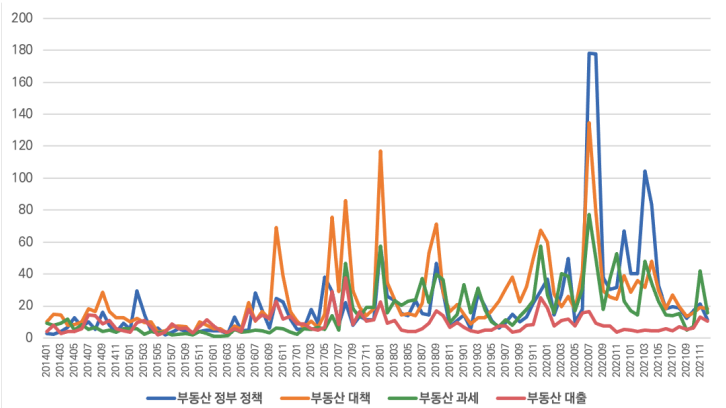
‘부동산 정책’ 영역에 대한 보도는 부동산 정책이 강화되던 시기에 일시적이고 집중적으로 늘어나는 추이를 보였다. 박근혜 정부 시기, 2016년 무렵 가계 부채가 급증하고 시장이 과열되면서 부동산 시장을 억제하는 방향으로 정책을 전환했다. 정부는 LH 택지공급을 조절하고, 중도금 대출 보증을 강화하고, 청약조정대상지역 내에 전매제한을 강화했다(박인권, 2022). 이에 해당 시기에는 ‘부동산 대책’과 관련한 기사가 다량 보도되었다.

이어 문재인 정부는 과열된 부동산 시장에 대응하기 위해 각종 정책을 시행했다. “2017년 8·2 대책”에서는 금융규제 정책과 양도세 중과 정책을 내놓았으며, 재건축 및 재개발 예정지역을 중심으로 투기지역과 투기과열지구를 지정했다(국토교통부, 2017a; 국토교통부, 2017b). 이후 서울 집값이 계속 상승하자 “2018년 8·27 대책”에서 투기지역과 투기과열지구를 지정했고, “2018년 9·13 대책”에서 고가주택과 다주택자 대상으로 대출 규제를 강화하고 종부세를 인상하는 안을 제시했다(국토교통부, 2018a; 국토교통부, 2018b). 그런데도 서울을 비롯한 수도권 지역에 국지적인 상승세가 재현되자(국토교통부, 2019), “2019년 12·16 대책”을 통해 과열 지역을 중심으로 대출 규제를 강화하고 세제를 상향했다(문화체육관광부, 2022). 2020년 7월에는 계약갱신청구권제·전월세상한제·전월세신고제를 골자로 하는 “임대차 3법(주택임대차보호법 및 부동산거래신고법 개정안)“이 의결되었다, 이러한 부동산 정책이 시행될 때마다 정책 내용과 그에 대한 평가를 담은 기사가 쏟아져 나왔다. 부동산 시장에 관한 각종 대책이 발표될 때마다 부동산 정책 영역 전반에 대한 여론의 관심이 거듭 증가하는 양상을 보였다.

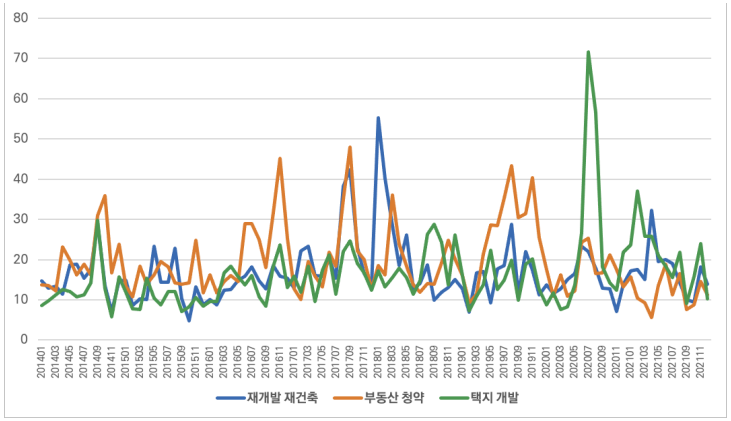
요약하면, 부동산 정책에 대한 기사는 사실상 부동산 “규제” 정책을 전제했다. 규제 강화 정책이 발표됨에 따라 부동산 정책에 관한 기사 수가 폭발적으로 증가했으며, 집값 상승 대책과 금융 및 조세 규제 방안에 대한 내용이 부동산 정책 기사 전반을 주도했다.



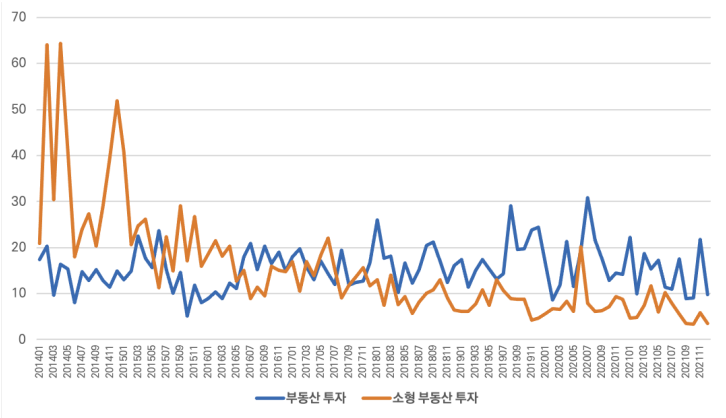
〈부동산 가격〉



〈부동산 정책〉

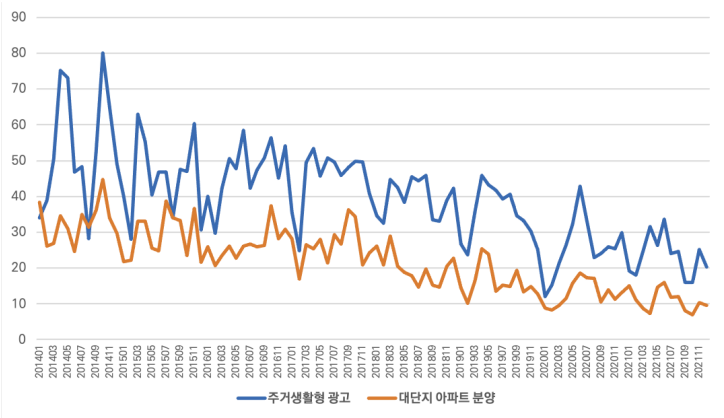


〈부동산 공급〉

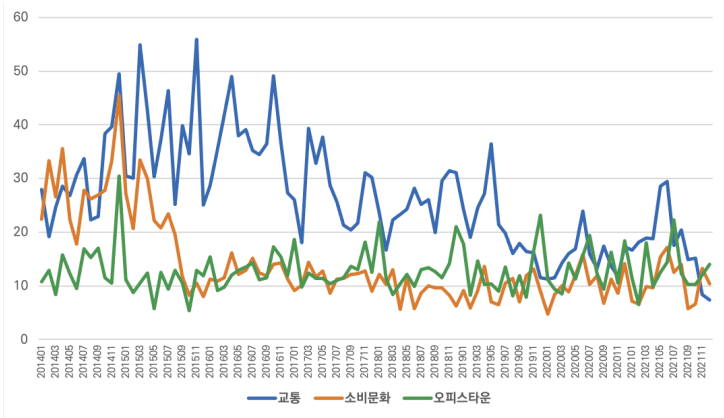


〈부동산 투자〉

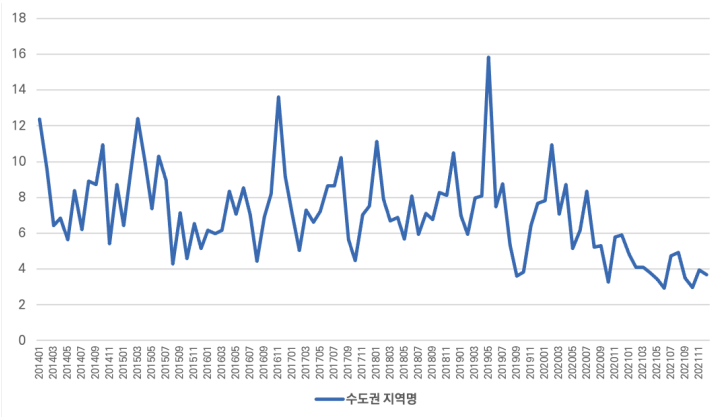
[그림 7] 부동산 신문기사 영역별·주제별 추이



〈부동산 광고〉



〈생활환경〉



〈지역시장〉

[그림 7] 부동산 신문기사 영역별·주제별 추이 (계속)

3) 부동산 공급

‘부동산 공급’ 영역은 당시의 부동산 정책의 내용에 따라 주제가 변화하는 양상을 보였다. 박근혜 정부 시기, “2014년 9·1 대책”에서 재건축 규제를 완화하고 택지개발촉진법 폐지하고 주택 청약 제도를 간소화하겠다고 선언하면서 ‘재개발·재건축’, ‘택지 개발’, ‘부동산 청약’ 주제의 부동산 공급 기사가 활발히 보도되었다. 이후 규제 완화에서 규제 강화로 정책 기조가 바뀌면서, “2016년 11·3 대책”에 행복주택·뉴스테이 공급, 분양권 전매 제한과 청약 자격 강화에 대한 내용 비중이 늘어났다. 이에 따라 “부동산 청약” 주제로 관심이 몰렸다.

또한 문재인 정부 시기 “2017년 8·2 대책”에서 수요 억제책으로 재건축 및 재개발 예정지역의 집값 상승을 견제하는 방안을, 공급 확대 방안으로 임대·분양주택 공급을 확대하는 동시에 실수요자 우선으로 청약제도를 개편해야 하는 방안을 내놓으면서, ‘재개발·재건축’, ‘부동산 청약’ 주제가 대거 보도되었다. 이후에도 정부가 “수도권 30만 호 공급“ 의지를 꾸준히 표명하면서, 2019년 하반기에는 ‘부동산 청약’ 주제에 대한 여론의 관심이 몰렸으며, 2020년 하반기에는 공공택지 분양에 관한 ‘택지 개발’ 기사가 조명되었다.

2010년대 후반기에는 부동산 가격과 부동산 정책이 사회적·경제적·정치적으로 핵심 주제로 자리 잡으면서 ‘부동산 공급’ 영역이 부동산 정책 내용에 종속되는 양상을 보였다. 부동산 가격과 부동산 정책이 부동산에 대한 담론을 전반을 관통하면서, 여타 주제도 하나의 흐름으로 귀착되는 모습을 보인 것이다.

4) 부동산 투자

‘부동산 투자’ 영역은 부동산에 대한 규제가 완화되었던 시기에 활발히 보도되다가 규제가 강화되자 수그러들었다. 박근혜 정부 시기, 2014년 소위 “부동산 3법”을 시행하면서 부동산에 대한 규제가 대폭 완화했다. 누구나 쉽게 청약 1순위 자격을 얻을 수 있게 했으며, 6개월만 지나면 분양받은 아파트에 웃돈을 붙여 전매할 수 있게 했다. 주택 관련 대출 요건도 대폭 완화해, 사실상 ”빚을 내서라도

주택을 사라“라는 기조를 보였다. 사회적으로는 저금리 시대에 부동산을 통해 불로소득을 확보해야 한다는 논의가 일상이 되고, 정책적으로는 부동산 투기에 더 쉽게 뛰어들 수 있도록 대문을 열어주면서(이준구, 2017), 부동산 투기는 합리적 행위로 일상화·상식화되었다. 당시에 강남이 아니면 제2강남이라도, 아파트가 아니면 빌라나 오피스텔이라도 부동산에 투자해야 한다는 담론이 구조화되면서 ‘소형 부동산 투자’에 관한 논의가 활발히 전개되었다.

5) 부동산 광고

부동산 광고는 건설업체의 시장점유 전략이자 중산층의 부동산 욕망의 매개 수단으로서 기능한다. 이들 광고는 상품에 대한 정보를 제공하는 본질적 기능을 실천함과 동시에, 사람들의 욕망을 매개하는 중개자의 역할을 수행한다(신인섭·김병희, 2007; 박철수, 2008). 따라서 ‘부동산 광고’ 영역의 내용과 흐름은 주거에 대해 인식과 욕구가 어떠한 형상을 지니는지, 부동산 투자 열풍이 어떠한 방식으로 확산되었는지를 재현한다.

‘부동산 광고’ 영역은 ‘부동산 투자’ 영역, ‘생활환경’ 영역과 상호 연관된 흐름을 보였다. 박근혜 정부 시기, 2014년 정부가 내 집 마련을 적극 종용하면서부터 부동산에 대한 광고성 기사가 꾸준히 보도되었는데, 아파트를 실속의 수익형 투자처이자 교통, 학교, 편의시설을 고루 갖춘 주거 공간으로 묘사하는 방식의 부동산 광고가 주를 이뤘다.

6) 생활환경

‘생활환경’ 영역은 부동산 광고 영역과 함께 증가했다. 아파트라는 상품에 대한 수요자 요구와 경제적 가치를 적극적으로 반영하는 분양 광고가 보급되면서(김명수, 2021) 교통 접근성, 부대시설, 어메니티 등의 아파트 분양 계약률에 긍정적 영향을 미치는 변수도 덩달아 가시화됐다(고현림 외, 2018). 생활환경 기사가 부동산의 입지적 특성, 잠재적 가치, 상승 기대치를 분석해 주는 역할로서 다뤄진 것이다.

제 2 절 부동산 신문담론의 변환

신문담론의 흐름을 파악한 데 이어, 신문담론의 골격이 변화한 맥락과 과정을 검토했다. 특정 시점에서 상대적으로 어떤 주제가 주목받았는지 비교하기 위해 기사량을 정규화했다. 반기별로 신문기사의 관심 영역 및 주제를 확인하기 위해 기사량 정규화 과정을 거쳤다. 전체 기사 수와 반기별 기사 수를 이용해 시기별 가중치를 산출했으며(식39), 시기별 가중치 값을 신문기사 주제별 부수에 곱해 각각 주제의 점유율을 계산했다.

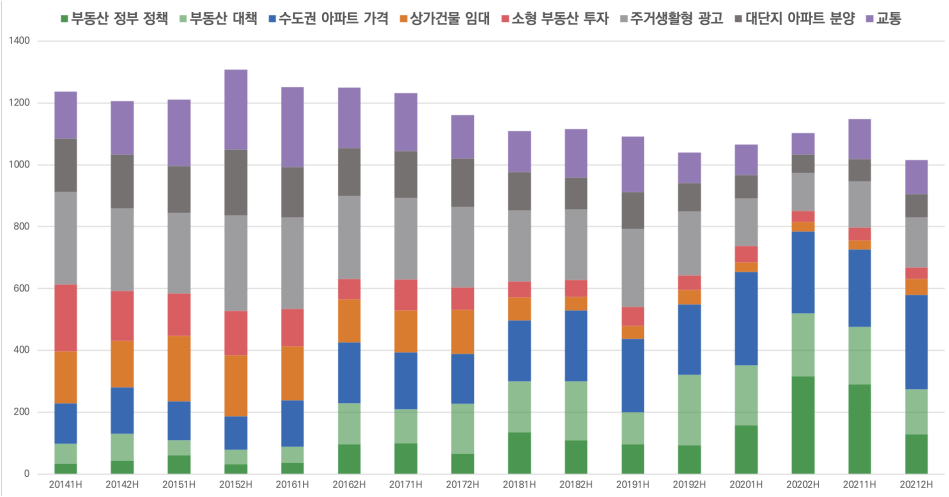
$$\left(\frac{\text{반기별 기사량}}{\text{전체 기사량}} \times 100\right)^{-1} \tag{39}$$

계산 결과를 토대로 반기별 주제 순위를 산출했다. 반기별 신문기사의 관심도 우선순위를 한눈에 이해하기 위해 '관심 순위 지도'를 작성했으며(방한솔·문호석, 2019), 관심 순위 1위부터 5위까지 음영 표시했다(〈표13〉). 관심 순위를 통해 특정 주제가 어느 시점에서 주목받았는지 해석하고, 동시에 특정 시점에 어떤 주제가 중심을 이뤘는지 비교하고자 했다.

관심 순위 분석 결과를 바탕으로 관심을 많이 받은 특정 주제를 선정해 시계열 그래프로 표현했다([그림8]). 시각화를 통해 부동산 신문담론의 핵심 골자가 변화하는 시점과 흐름을 보이고자 했다.

〈표 13〉 부동산 신문기사 관심 순위 지도

	부동산 가격					부동산 정책				부동산 공급			부동산 투자		부동산 광고		생활환경			지역 시장
	부동산 가격 추세	강남권 아파트 가격	수도권 아파트 가격	전세금 동향	상가 건물 임대	부동산 정부 정책	부동산 대책	부동산 과세	부동산 대출	재개발 재건축	부동산 청약	택지 개발	부동산 투자	소형 부동산 투자	주거생활형 광고	대단지 아파트 분양	교통	소비 문화	오피스 타운	수도권 지역명
20141H	12	13	7	10	4	19	15	17	20	9	8	16	11	2	1	3	5	6	14	18
20142H	16	13	7	9	6	18	10	20	17	12	8	14	15	4	1	2	3	5	11	19
20151H	14	12	8	7	3	13	18	20	19	11	10	15	9	6	1	4	2	5	16	17
20152H	16	12	7	8	4	19	17	20	15	10	6	14	13	5	1	3	2	9	11	18
20161H	15	12	5	9	3	19	16	20	18	11	7	8	14	6	1	4	2	10	13	17
20162H	13	18	2	17	6	9	7	20	12	10	5	11	8	16	1	4	3	15	14	19
20171H	14	16	3	18	5	9	7	20	17	6	12	11	8	10	1	4	2	15	13	19
20172H	14	18	3	19	5	16	2	9	11	8	7	10	12	15	1	4	6	17	13	20
20181H	10	13	2	16	14	6	3	5	19	4	9	12	11	18	1	8	7	17	15	20
20182H	7	14	1	15	20	8	3	5	16	13	10	6	11	17	2	9	4	18	12	19
20191H	10	15	2	14	19	9	8	4	20	12	5	11	7	17	1	6	3	18	13	16
20192H	14	13	1	16	19	10	2	6	17	7	3	9	5	18	4	11	8	15	12	20
20201H	6	16	1	11	20	4	2	3	15	8	9	12	10	18	5	13	7	17	14	19
20202H	9	8	2	5	19	1	3	4	18	15	10	7	11	17	6	14	12	16	13	20
20211H	11	15	2	10	18	1	3	6	19	8	12	4	9	17	5	13	7	16	14	20
20212H	11	16	1	4	18	5	3	7	17	9	13	6	12	19	2	15	8	14	10	20



[그림 8] 부동산 신문기사 주요 주제의 연간 변화량

1) 자산 증식 수단이자 적극적 투기 대상으로서 부동산, 2014-2016

2014년 상반기부터 2016년 상반기까지는 '부동산 가격', '부동산 투자', '부동산 광고'와 '생활환경'에 대한 기사가 신문 지면을 차지했다. 상가나 오피스텔 가격은 어느 정도인지, 그에 따른 경제적 이익은 어느 정도일지에 대한 정보가 적극 공유되었다.[6] 더불어, 아파트의 입지적 메리트와 연계 편의성을 강조하거나, 아파트를 단지에서 모든 것이 해결되는 일종의 자기충족적 세계로 묘사하는 식의 광고가 대거 확산되었다.

부동산을 통한 자산 증식과 투기 행위를 합리적인 전략으로써 전파하는 내용이 부동산 신문담론의 상당 부분을 차지했다. 금리나 세금 부담 없이 부동산 투기에 뛰어들기 쉽던 시기이자 정부가 오히려 이를 권장하던 셈이었던 시기였기에(이준

6) 현재의 관점에서는 이러한 내용이 다분히 당연하고 익숙하게 느껴질 수도 있다. 그러나, 2010년대 이전까지만 하더라도 '부동산 투자'는 언론에서 현재만큼 빈번히 다루어지지 않았다. 1990년대 초반에는 부동산 투자를 경제적 전략보다는 국가적 범죄로서 지면에 다뤘고, 2000년대에는 부동산 투자 논의를 특정 계층이나 특정 주택으로 한정했다. 2010년대에 접어들면서 부동산 투자 논의 대상이 땅이나 아파트에서 오피스텔로, 부동산 투자 주체를 특권층에서 대중으로 확장했다. 이러한 신문담론의 변화는 우리 사회에서 부동산 투자·투기가 사회적으로 선택되고 승인되어 온 흐름을 간접적으로 드러낸다(홍지수, 2022).

구, 2017) 환금성과 수익성이 부동산 신문담론의 헤게모니적 위치를 점유했다.

2016년 하반기, 재건축·재개발 아파트와 재건축 후보지 아파트 가격이 상승하기 시작하면서 정부가 부동산 투기 진정안을 발표하기 이르던 때까지 이와 같은 논의가 지속되었다.

2) 통제의 대상이자 사회적 병폐로서 부동산, 2017-2021

2017년 하반기부터 2021년 하반기까지는 '부동산 가격', '부동산 정책'에 대한 신문기사가 연이어 보도되었다. 부동산 가격 상승과 부동산 규제 강화라는 일련의 사이클이 누차 반복되면서, 이들이 부동산 신문담론의 핵심 축을 담당하게 되었다. 부동산 가격 상승을 우려하거나 혹은 부동산 매매 차익을 부각하는 기사가 보도되었으며, 부동산 정책의 내용을 설명하고 성패를 평론하는 기사가 뒤를 이었다.

해당 시기는 서로 대비되는 논조의 기사가 공존했다. 한편으로는, 아파트 가격 상승 추이가 가속화되면서 부동산 불패 신화가 부동산 신문담론의 헤게모니로 다시금 고착되었다. 다른 한편으로는, 부동산을 통제의 대상이자 사회적 병폐로서 전제하는 정부 기조가 주요한 부동산 신문담론 골자가 되었다.

2019년 하반기부터는 기존의 '부동산 광고'나 '생활환경'에 대한 기사가 자취를 감췄고, 빈자리를 '부동산 정책'이 채웠다. 가격은 오르고 정책은 강화되는 양단의 상황이 지속되며, 수도권 부동산만 한 자산이 없다거나 수도권 주거비 부담이 막심하다는 내용의 부동산 가격 기사가, 또는 규제의 난맥상에 대한 부동산 정책 기사가 점점 더 부동산 신문담론을 지배하는 상황에 이르렀다.

종합하면, 부동산 신문담론은 당시의 사회적·경제적·정치적 맥락에 영향을 받고 다시 영향을 주며 일련의 방향으로 전개되었다. 때로는 환상과 신화를 형성하는 방식으로, 또 때로는 불안감과 위기감을 조성하는 방식으로 설파됐다. 부동산 신문담론의 변화만으로 실제 부동산이 어떻게 변화했는지 구체적으로 설명할 수 없으나, 시장 추세와 정부 기조에 종속되는 신문담론의 특성에 말미암아 이들이 부동산의 특정한 추이를 조금 더 강화하고 연장하는 방식으로 작용했을 것이라 추측할 수가 있다.

제 5 장 부동산 비기초가격과 부동산 신문담론의 관계

제 4 장에서 부동산 신문담론이 사회적·경제적·정치적 상황과 밀접한 연관을 맺으며 형성하고 변화해 왔음을 확인했다. 부동산 신문기사 비정형 텍스트 데이터에는 정형 데이터로는 표현하기 어려운 사회적 관념, 경제적 내러티브, 정치적 기조와 같은 요소들이 담겨있음을 보았다. 본 장에서는 담론들이 사람들의 감정이나 행위를 변화시킬 수 있는 형태로 발산된다는 전제하에(Shiller, 2021), 부동산 비기초가격과 부동산 신문담론의 관계를 분석하고자 한다.

제 1 절 분석 자료의 안정성 검정

시계열 분석 자료의 안정성을 검정하기 위해 ADF 단위근 검정을 수행했다. 시계열이 안정적일 경우, 해당 시계열의 평균, 분산, 자기공분산은 일정한 추이를 보인다. 다시 말해, 안정 시계열은 일시적인 변화가 일어날지라도 다시 평균으로 돌아가려 하는 평균 회귀(mean reversion) 성향을 지니며, 일반화를 통한 예측이 가능하다(Gujarati and Poter, 2009). 반면, 시계열이 단위근을 갖는 불안정 시계열일 경우, 충격이 발생하면 그에 따른 효과가 소멸하지 않고 지속되기 때문에 예측이 어렵다. 또한, 불안정 시계열을 사용해 분석할 경우 실제로 관계가 없는 변수들 사이에 유의미한 관계가 존재하는 것처럼 나타나는 가성 회귀(spurious regression) 현상이 발생할 수 있다(성채원, 2005; 이우민, 2019). 이와 같은 문제를 막기 위해, 본 연구는 시계열 변수에 대한 ADF 단위근 검정을 선행했다.

〈표 14〉 ADF 단위근 검정 결과

		Level Variables		Difference Variables	
		ADF	p-value	ADF	p-value
신문기사	부동산 가격	-3.650	0.028*	-10.296	0.000*
	부동산 정책	-3.120	0.104	-9.823	0.000*
	부동산 공급	-3.496	0.043*	-10.479	0.000*
	부동산 투자	-2.081	0.543	-9.167	0.000*
	부동산 광고	-1.881	0.627	-9.191	0.000*
	생활환경	-7.739	0.000*	-5.159	0.000*
	지역시장	-3.221	0.087	-4.042	0.011*
부동산 비기초가격		-2.272	0.278	-8.346	0.000*

confidence level: *0.05 lower

단위근 검정 결과는 〈표14〉와 같다. 먼저 시계열 자료를 수준변수(level)로 하여 단위근 검정을 수행했으며, 1차 차분한 변수에 대한 추가 단위근 검정을 실시했다. 수준변수는 대부분 귀무가설을 기각하지 못하고 불안정했으나, 1차 차분 후에는 귀무가설을 기각하며 안정화되었다($p<0.05$). 이를 통해 모든 1차 차분 자료에 단위근이 존재하지 않음을 확인했으며, 해당 자료를 인과관계 분석에 활용했다.

제 2 절 Granger 인과관계 검정

2014년 1월부터 2021년 12월까지 부동산 비기초가격과 부동산 신문기사의 Granger 인과관계를 분석했다(〈표15〉). 그 결과, 부동산 비기초가격과 부동산 정책 기사, 부동산 광고 기사, 생활환경 기사, 지역시장 기사가 유의미한 인과관계를 가졌다. 부동산 비기초가격이 정책, 심리, 정보 등과 연관을 지닌다는 선행연구의 이론적 판단과 일정 정도 일맥상통한 결과가 도출되었다. 한편, 부동산 비기초가격과 부동산 가격 기사, 부동산 공급 기사, 부동산 투자 기사 간에는 인과관계가 나

타나지 않았다. 부동산 가격에서 경제적 요인으로 설명되는 부동산 기초가격 영역을 제외하고 비정량적 요인으로 설명된다고 알려진 부동산 비기초가격을 별도의 변수로 활용한 만큼, 경제적 부문과 직접적인 연관을 보이는 기사와는 유의미한 인과관계가 나타나지 않은 것으로 예상된다.[7)]

주된 분석 결과를 정리하면 다음과 같다. 첫째, 부동산 정책 기사가 부동산 비기초가격에 유의미한 영향을 미쳤다. 이는 부동산 과세, 부동산 규제 등의 부동산 정책이 부동산 비기초가격과 연관이 있는 것으로 해석된다. 부동산 정책 기사는 부동산 가격 추세에 대한 정부의 정책적 대응을 내용으로 하며, 부동산 규제 정책이 시행된 시점에 증가하는 양상을 보였다. 즉, 부동산 정책 기사의 추이는 부동산 정책의 시행 여부 및 시행 강도 등과 맞닿아 있다. 대다수 수요자는 언론매체를 통해서 부동산 정책 내용을 접하게 되며, 정책 발표 및 시행 이후에는 언론을 통해 부동산 전문가의 평가와 전망에 관한 정보를 연이어 얻게 된다(김성용·오세준, 2016). 때문에 부동산 정책이 시행되고 수요자와 투자자가 언론매체를 통해 부동산 정책의 내용이나 정책에 대한 평가를 접하게 되면서(박재수·이재수, 2019), 심리적 시장에 유의미한 변화가 나타나는 것으로 해석된다.

7) 부동산 기초가격과 부동산 신문기사의 Granger 인과관계를 분석한 결과, 부동산 가격 기사, 부동산 공급 기사, 부동산 투자 기사와 유의미한 인과관계를 가졌다. 부동산 기초가격의 경우 비교적 경제 부문과 연관성이 강한 기사와 유의미한 관계를 보였다.

〈표 15〉 Granger 인과관계 검정 결과

Granger Test	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
	F-value (p-value)	F-value (p-value)	F-value (p-value)	F-value (p-value)	F-value (p-value)
부동산 비기초가격 → '부동산 가격' 기사	0.8445 (0.3594)	0.5958 (0.5523)	0.5834 (0.6267)	0.3998 (0.8086)	0.3536 (0.8794)
'부동산 가격' 기사 → 부동산 비기초가격	0.2171 (0.6418)	1.0037 (0.3686)	1.1015 (0.3502)	1.1276 (0.3454)	0.8344 (0.5271)
부동산 비기초가격 → '부동산 정책' 기사	0.0104 (0.9189)	0.0067 (0.9933)	0.0725 (0.9746)	0.1887 (0.9440)	0.2897 (0.9181)
'부동산 정책' 기사 → 부동산 비기초가격	0.3162 (0.5746)	0.8861 (0.4141)	4.5064*** (0.0054)	3.9773*** (0.0042)	3.6763*** (0.0036)
부동산 비기초가격 → '부동산 공급' 기사	1.1607 (0.2828)	1.5919 (0.2065)	1.0649 (0.3656)	0.7585 (0.5537)	0.5971 (0.7022)
'부동산 공급' 기사 → 부동산 비기초가격	0.0329 (0.8562)	0.2466 (0.7818)	0.6375 (0.5918)	1.1728 (0.3248)	1.0452 (0.3932)
부동산 비기초가격 → '부동산 투자' 기사	0.5857 (0.4451)	0.3180 (0.7280)	0.6225 (0.6014)	0.4954 (0.7392)	0.4009 (0.8477)
'부동산 투자' 기사 → 부동산 비기초가격	1.2205 (0.2707)	1.3384 (0.2650)	0.9075 (0.4388)	1.2115 (0.3080)	1.1710 (0.3259)
부동산 비기초가격 → '부동산 광고' 기사	2.8291* (0.0937)	1.7105 (0.1838)	2.6255* (0.0522)	3.0395** (0.0189)	1.7442 (0.1276)
'부동산 광고' 기사 → 부동산 비기초가격	0.0945 (0.7589)	0.3758 (0.6873)	2.8513** (0.0390)	1.9064 (0.1118)	2.1122* (0.0668)
부동산 비기초가격 → '생활환경' 기사	3.5771* (0.0602)	1.5974 (0.2054)	1.2519 (0.2927)	2.6196** (0.0369)	2.4539** (0.0359)
'생활환경' 기사 → 부동산 비기초가격	1.2791 (0.2596)	1.1973 (0.3045)	4.4627*** (0.0048)	2.9779** (0.0209)	2.6450** (0.0252)
부동산 비기초가격 → '지역시장' 기사	2.4821 (0.1169)	1.1018 (0.3346)	1.1046 (0.3489)	1.2734 (0.2826)	0.9665 (0.4402)
'지역시장' 기사 → 부동산 비기초가격	5.8878*** (0.0162)	4.1662*** (0.0171)	3.8377*** (0.0108)	2.4718** (0.0466)	2.2614* (0.0510)

confidence level: *0.10 lower, **0.05 lower, ***0.01 lower

둘째, 부동산 광고 기사와 부동산 비기초가격 상호 간의 유의미한 인과관계가 나타났다. 광고는 본질적으로 상품에 대한 정보를 제공하는 기능을 수행하며, 사람들의 욕망을 매개하는 중개자로서 기능한다(박철수, 2008). 마찬가지로, 부동산 광

고는 건설업체의 시장점유 전략과 수용자의 부동산 욕망, 즉 부동산이라는 상품에 관한 기업의 필요와 사회적 의미를 포함한다(박철수, 2008; 김명수, 2021; 홍지수, 2022). 더 나아가, 부동산 광고는 이들 간의 매개자 역할을 수행한다. 따라서 부동산에 대한 광고를 빈번히 접하는 환경이 조성되면 수요자 및 투자자의 정서와 행동에 영향을 미쳐, 부동산 심리시장에 변화가 생기는 것으로 추측할 수 있다.

셋째, 생활환경 기사가 부동산 비기초가격과 상호 유의미한 인과관계를 가졌다. 앞서 주로 아파트 분양 일정이나 아파트 입지, 대단지 규모 등을 담고 있는 부동산 광고 기사가 부동산 비기초가격과 인과관계를 가진 것과 유사하게, 부동산이 위치한 지역의 연결성이나 편의성에 관한 정보를 포함하는 생활환경 기사 또한 부동산 비기초가격과 유의미한 인과관계를 가졌다. 이는 개별 아파트나 건설 브랜드의 경쟁력뿐 아니라, 부동산이 위치한 공간의 흡인력이 부동산 심리시장에 영향을 미친다는 것을 의미한다.

넷째, 수도권 지역 부동산, 그중에서도 분당이나 판교와 같이 소위 대장주 격이라 지칭되는 수도권 지역 부동산에 관한 기사가 부동산 비기초가격에 유의미한 영향을 미쳤다. 이는 인근 지역 부동산 시장의 공급량이나 상품성 등이 서울 부동산 시장의 위상에 관한 인식에 유의미한 영향을 준다는 것을 추론케 한다.

제 3 절 공적분 검정 및 VECM 모형 추정

앞선 분석에서 부동산 비기초가격과 부동산 정책 기사, 부동산 광고 기사, 생활환경 기사, 지역시장 기사가 인과관계를 갖는 것으로 나타났다. 인과관계 분석 결과가 유의했던 변수를 중심으로 부동산 비기초가격와 부동산 신문기사의 시계열적 영향 관계를 분석하기 위해, VAR 모형(Vector Autogressive Model)을 수정한 VECM 모형(Vector Error Correction Model)을 추정했다.

VAR은 시계열 변수에 대한 충격반응분석, 분산분해분석을 통해 변수 간의 시계열적 영향 관계를 분석할 수 있다는 장점을 지닌다. VECM은 VAR의 이러한 장

점을 포함하되, 기존에 불안정 시계열 자료를 차분해 사용하면서 시계열 사이의 장기적 관계에 대한 정보를 상실했던 문제를 개선한 모형이다. VECM은 변수가 불안정하더라도 변수들이 장기적인 균형 관계(long-term equilibrium)를 맺을 경우, 다시 말해 변수들 사이에 공적분이 존재할 경우 해당 변수들의 선형결합은 안정적일 수 있다고 가정한다. VECM은 장기 공적분 관계를 기반으로 하는 오차수정항(error correction term)을 도입한 모형으로, 변수들이 공적분 관계를 가진다면 차분하지 않고 원자료 그대로 장기적 정보를 보유한 채 분석할 수 있어 VAR 모형의 단점을 개선한 모형으로 평가받는다.

본 연구는 VECM 모형을 추정하기에 앞서, 변수의 시차 길이를 선정했다. 시차 구조는 인과성 검정의 결과에 민감한 영향을 미치므로, 시차 길이를 자의적으로 결정하면 추정 계수를 왜곡하거나 잘못된 인과성을 추론하는 문제에 이를 수 있다(김선재·이영화. 2014) 이에 본 연구는 적정 시차 검정을 수행했다. AIC는 적정 시차 값을 과도하게 크게 판정하는 경우가 있기 때문에, SC 값과 HQ 값이 최소가 되는 시차인 1을 적정 시차로 간주하였다. VAR 모형의 적정 시차인 p를 1로 결정했으므로, VECM 모형의 적정 시차는 p-1인 0으로 설정했다(이우민 외, 2019).

〈표 16〉 VAR 시차 선택 검정 결과

Lag	AIC	SC	HQ
0	27.300	27.443	27.357
1	25.596	26.452*	25.940*
2	25.881	27.451	26.513
3	25.695	27.978	26.614

*indicates lag order selected by the criterion

또한, VECM이 분석에 적합한지 검토하기 위해, 변수 사이에 장기적 균형 관계 유무 여부를 공적분 검정을 통해 살펴봤다. 공적분 관계가 존재하지 않을 경우 안정 시계열로 차분하여 VAR 모형을 추정해야 하며, 공적분 관계가 존재할 경우

장기적 정보를 보유한 채로 분석하는 VECM 모형을 활용할 수 있다(이우민 외, 2019). 본 연구는 최우추정법을 통한 Johansen 공적분 검정을 시행했으며, 변수 사이에 공적분 관계가 존재하는 것으로 나타났다(〈표17〉). 이에 VECM 모형을 추정하여 부동산 비기초가격와 부동산 신문기사의 관계를 분석했다.

〈표 17〉 Johansen 공적분 검정 결과

r	Eigenvalue	Trace Statistic	5% Critical Value	p-value
0*	0.461	183.701	79.341	0.0000
1*	0.431	126.154	55.246	0.0000
2*	0.314	73.685	35.011	0.0000
3*	0.226	38.610	18.398	0.0000
4*	0.147	14.838	3.841	0.0001

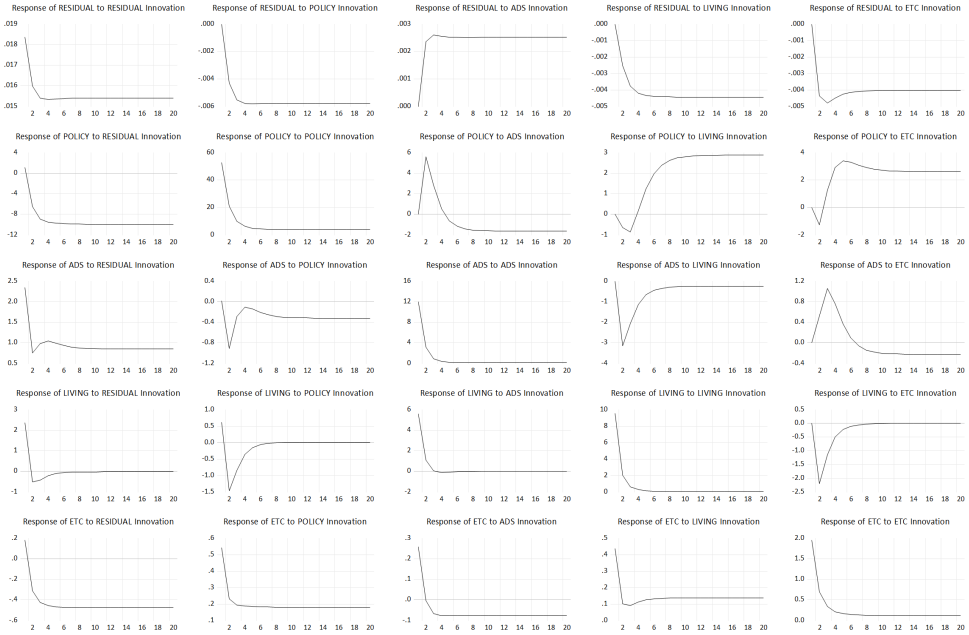
confidence level: *0.05 lower

제 4 절 충격반응분석 및 분산분해분석

1) 충격반응분석

VECM 모형 추정 결과를 기초로 충격반응분석(impulse response function)을 시행했다. 부동산 비기초가격과 부동산 정책 기사, 부동산 광고 기사, 생활환경 기사, 지역시장 기사 등의 부동산 신문기사 시계열 자료로 추정한 VECM 모형을 적용해, 각각의 시계열에 대하여 1단위 표준편차 충격을 주었을 때 여타 시계열이 어떠한 영향력을 받는지 분석했다([그림9]).

우선, 부동산 비기초가격과 부동산 신문기사 변수는 모두 자체적인 충격에 대하여 1기에 큰 영향을 받고, 2기부터 다시 회복되면서 장기적으로 일정 수준을 유지하는 것으로 나타났다. 이는 개별 변수의 자기상관성이 높기 때문이다(김동환, 2015).



[그림 9] 부동산 비기초가격과 부동산 신문기사 간 충격반응분석 결과

다음으로, 부동산 비기초가격의 단위 충격이 부동산 신문기사에 미치는 영향을 살펴보았다. 부동산 비기초가격의 충격은 부동산 정책 기사(POLICY)에 음(-)의 방향으로 영향을 미쳤다. 한편, 부동산 비기초가격은 부동산 광고 기사(ADS)와 생활환경 기사(LIVING)에 일시적인 양(+)의 영향을 미치는 것으로 분석됐다. 부동산 광고 기사와 생활환경 기사가 대략 2기까지 양(+)의 영향을 미치다가, 3기부터는 그 효과가 소멸되는 것으로 나타났다. 부동산에 대한 과열이 형성되면서, 그에 대한 홍보성 기사가 더욱 적극적으로 보도된 것으로 예상할 수 있다. 마지막으로, 부동산 비기초가격의 충격은 지역시장 기사(ETC)에 음(-)의 방향으로 영향을 미치는 것으로 나타났다.

반대로, 부동산 신문기사의 구조적 충격이 부동산 비기초가격에 미치는 영향을 분석했다. 부동산 정책 기사의 충격은 부동산 비기초가격에 음(-)의 방향으로 영향을 미쳤으며, 그 효과가 지속되는 것으로 나타났다. 이는 과세나 규제와 같은 부동산 정책이 부동산 심리시장을 일정 정도 진정시키는 효과를 갖는다는 것을 시사

한다. 반면, 부동산 광고 기사의 충격은 부동산 비기초가격에 계속된 양(+)의 영향을 미쳤다. 이러한 결과는 부동산의 투자가치나 사용 가치에 대한 적극적 홍보가 부동산에 대한 심리적 기대감에 영향을 미치며, 이러한 사회적 경향이 가격에도 반영된다는 것을 나타낸다. 지역시장 기사의 충격에 대해 부동산 비기초가격은 음(-)의 영향을 받았다. 이와 같은 결과를 해당 기사가 주로 서울 인근의 수도권 지역을 배경으로 한다는 점을 염두에 두고 해석할 필요가 있다. 이는 분당, 판교 등 경기권 부동산 지역에 대한 긍정 여론이 팽창하면, 상대적으로 서울 부동산에 대한 심리시장이 다소 위축될 수 있음을 의미한다.

2) 분산분해분석

다음으로, VECM 모형에 대한 분산분해분석(variance decomposition analysis)을 수행했다. 부동산 비기초가격 변수의 예측오차가 부동산 정책 기사, 부동산 광고 기사, 생활환경 기사, 지역시장 기사 등의 부동산 신문기사 변수들에 의해 얼마나 설명되는지 백분율로 나타냈다(〈표18〉). 이를 통해 변수 간의 상대적 중요성을 분석하고, 각각의 변수가 부동산 비기초가격의 예측에 어느 정도의 영향을 주는지 파악했다.

〈표 18〉 부동산 비기초가격의 분산분해분석 결과

Period	S.E.	부동산 비기초가격	신문기사			
			부동산 정책	부동산 광고	생활환경	지역시장
1	0.018	100.000	0.000	0.000	0.000	0.000
2	0.025	92.292	2.876	0.862	1.005	2.966
3	0.031	87.017	5.139	1.288	2.160	4.397
4	0.036	84.069	6.513	1.486	3.022	4.910
5	0.040	82.333	7.362	1.596	3.621	5.088
6	0.044	81.214	7.923	1.666	4.046	5.151
7	0.047	80.432	8.319	1.716	4.357	5.176
8	0.050	79.853	8.613	1.753	4.594	5.186
9	0.053	79.405	8.841	1.782	4.780	5.192
10	0.056	79.047	9.023	1.806	4.929	5.195

부동산 비기초가격의 예측오차에 대한 분산분해분석을 실시한 결과, 부동산 비기초가격의 예측오차는 대부분 자신의 변동에 의하여 설명되었다. 부동산 비기초가격의 경우, 2기 92%에서 꾸준하게 감소하지만, 10기 79%까지 지속적인 영향을 미치는 것으로 나타났다. 부동산 신문기사는 부동산 정책 기사, 지역시장 기사, 생활환경 기사, 부동산 광고 기사 순으로 설명력을 갖는 것으로 파악됐다. 부동산 정책 기사는 2기부터 영향을 미치기 시작하여, 10기에 10%에 가까운 영향력을 갖는 것으로 나타났다. 지역시장 기사와 생활환경 기사는 모두 2기에 1%~3%대의 영향을 미치고, 10기에 5% 수준의 설명력을 갖는 것으로 분석됐다. 반면, 부동산 광고 기사는 3기에 1%대의 미미한 영향력을 미치고, 10기까지 유사한 수준을 이어가는 것으로 나타났다.

제 5 절 소결

본 장에서는 부동산 비기초가격과 부동산 신문담론의 관계를 Granger 인과관계 분석과 VECM 기반의 충격반응분석, 분산분해분석을 통해 살펴보았다. 주요 분석 결과를 요약하면 다음과 같다.

첫째, Granger 인과관계 분석 결과, 부동산 비기초가격은 부동산 신문기사 중 부동산 정책 기사, 부동산 광고 기사, 생활환경 기사, 지역시장 기사 등과 유의미한 인과관계를 가졌다. 반면, 부동산 가격 기사, 부동산 공급 기사, 부동산 투자 기사 등 경제적 영역과 직접 연관된 기사와는 유의미한 인과관계를 보이지 않았다. 이는 부동산 가격을 경제적 요인으로 설명되는 부동산 기초가격과 설명되지 않는 부동산 비기초가격으로 구분하여 부동산 비기초가격을 별도의 변수로 구축한 이 연구의 맥락과 상통하며, 부동산 비기초가격이 정책, 심리, 정보 등과 유관할 것이라는 선행연구의 주장과도 연결된다.

둘째, 충격반응분석 결과, 부동산 신문기사의 구조적 충격이 부동산 비기초가격에 영향을 미쳤다. 부동산 신문담론이 내용의 구성에 따라 부동산 비기초가격을

부풀리거나 누그러뜨리는 요인으로 작용할 수 있음을 확인했다. 부동산 정책 기사의 충격은 부동산 비기초가격에 부정의 영향을 미쳤으며, 그 효과가 일정 기간 지속되었다. 이는 과세나 규제와 같은 부동산 정책이 부동산 심리시장을 어느 정도 진정시키는 효과가 있다는 것을 의미한다. 반면, 부동산 광고 기사와 생활환경 기사의 충격은 부동산 비기초가격에 긍정의 영향을 미쳤다. 이는 부동산의 투자가치나 사용 가치에 대한 정보가 확산되면 부동산에 대한 심리적 기대감을 부풀리고, 이러한 사회적 경향이 부동산 비기초가격에 반영된다는 것을 내포한다. 한편, 지역시장 기사의 충격은 부동산 비기초가격에 부정의 영향을 미쳤다. 이와 같은 결과는 해당 기사가 서울 인근의 수도권에 위치한 부동산 대장주 격의 지역을 배경으로 한다는 점을 인지하고 해석할 필요가 있는데, 주변 지역 부동산이 적극적으로 조명되면서 일시적으로 서울의 부동산이 지니는 상징성 내지 위치성이 위축될 수 있음을 암시한다.

셋째, 분산분해분석 결과, 부동산 비기초가격은 자신의 변동에 의해서 80%~90%대의 강한 영향을 받았으며, 부동산 신문기사는 부동산 정책 기사, 지역시장 기사, 생활환경 기사, 부동산 광고 기사 순으로 영향력을 가졌다. 부동산 정책 기사는 10%에 가까운 큰 영향력을 갖는 것으로 나타났고, 지역시장 기사와 생활환경 기사는 모두 최대 5% 수준의 영향을 미치는 것으로 나타났다. 한편, 부동산 광고 기사의 영향력은 1%대 미미한 수준이었다.

전체적인 결과를 종합해 보면, 부동산 비기초가격과 부동산 정책에 관한 여론 간에 유의미한 영향 관계에 존재한다. 부동산 정책이 시행되어 정책적 논의가 형성되면, 부동산 가격에 대한 과신이나 기대가 일부 수그러드는 것으로 예상된다. 이는 부동산 비기초가격을 언론과 정책으로 설명한 Bucchianeri(2011)의 연구와도 유사한 결과이다. 부동산 비기초가격과 부동산 전망에 관한 정보 사이에도 유의미한 관계가 존재한다. 정보의 비대칭성이 높은 부동산 시장의 특성을 감안하면, 이와 같은 결과는 개인이 한정된 관점의 정보를 불필요하게 반복적으로 접하면서 타인을 참고하고 모방하여 행동하게 되는 것으로 이해된다(Bikhchandani et al., 1992). 더해, 부동산 비기초가격과 인근 수도권 지역에 대한 정보가 유의미한 관

계를 맺는다. 서울의 아파트는 위계적인 한국의 부동산 시장 내에 독보적인 지위와 정체성을 가졌다고 알려진 만큼, 그에 대한 과대평가와 낙관주의가 가격에 반영되어 있고, 주변 새로운 부동산이 조명되면 이러한 심리적 기대가 일부 옮겨가는 것으로 예측할 수 있다.

제 6 장 부동산 신문논조를 통한 부동산 비기초가격 예측

제 5 장에서 부동산 비기초가격과 부동산 신문기사 사이에 유의미한 연관관계가 있음을 확인했다. 본 장에서는 부동산 비기초가격과 부동산 신문기사의 관계를 분석하는 수준에서 한 단계 더 나아가, 부동산 신문기사를 활용해 부동산 비기초가격을 예측할 방안을 고안했다. 본 장에서는 부동산 신문기사의 논조를 지수화해, 해당 부동산 신문기사 논조지수의 예측 유용성 및 활용 가능성을 실증했다.

제 1 절 분석 자료의 안정성 검정

시계열 분석을 위해서는 자료가 안정성(stationarity)을 지니는지를 확인해야 한다. 안정성은 시계열 변수의 통계적 특성이 시간에 따라 일정하게 유지됨을 의미한다. 만일 시계열 변수가 안정성을 가지지 않는다면, 이를 통해 도출한 결과는 부정확할 가능성이 높다.

본문은 시계열 변수의 안정성 여부를 판단하기 위해 ADF(Augmented Dickey Fuller) 단위근 검정을 수행했다. 해당 검정에서 단위근이 존재하면 시계열 데이터가 추세를 가지고 있어 불안정(non-stationary)하다는 것을 의미하며, 단위근이 존재하지 않는다면 시계열이 안정적(stationary)인 상태임을 뜻한다.

시계열 자료를 수준변수(level variable)로 하여 단위근 검정을 수행한 결과, 귀무가설을 기각하지 못했다. 이는 해당 시계열은 단위근을 가지는 불안정 변수임을 의미한다. 이에 수준변수를 차분하여 단위근 검정을 재실시했다. 1차 차분한 시계열 자료에 대한 단위근 검정을 추가로 시행한 결과, 1차 차분 변수는 귀무가설을 기각하여 단위근을 가지지 않고 안정적이었다($p<0.01$)(<표19>). 해당 검정 결과를 토대로 차분 변수를 채택하고, 추후의 분석에 활용했다.

〈표 19〉 ADF 단위근 검정 결과

	Level Variables		Difference Variables	
	ADF	p-value	ADF	p-value
부동산 신문기사 논조지수	-2.801	0.201	-5.824	0.000*
부동산 비기초가격	-2.724	0.278	-8.346	0.000*

confidence level: *0.01 lower

제 2 절 Granger 인과관계 검정

시계열 변수의 인과관계를 검정하고자 할 경우 Granger 인과관계 검정(Granger causality test)을 활용할 수 있다. Granger 인과관계 검정은 두 개의 시계열 변수에 대해, 하나의 시계열 변수가 다른 시계열 변수의 과거 값을 예측하는 데 유의미한 정보를 제공하는지 검정하는 방법이다. Granger 인과관계 결과가 유의미하다면, 한 변수가 다른 변수의 예측에 유의미한 정보를 제공하여, 해당 변수가 다른 변수를 예측하는 데 유용한 변수로 활용될 수 있음을 의미한다. 이러한 검정은 직접적인 인과관계를 추론하는 과정이 아니며, 예측력에 기초하여 간접적으로 인과관계를 파악하는 방법이다.

부동산 신문기사 논조지수와 부동산 비기초가격의 차분변수에 대해 Granger 인과관계 검정을 수행했다. 분석 결과, 부동산 신문기사 논조지수는 시차(lag) 2에서 부동산 비기초가격에 대해 유의미한 인과관계를 가졌다($p<0.05$). 한편, 부동산 비기초가격은 부동산 신문기사 논조지수에 특별한 인과관계를 보이지 않았다(〈표 20〉) 해당 결과를 말미암아, 부동산 신문기사 논조지수가 부동산 비기초가격을 예측하는 데 활용될 수 있는 가능성을 확인했다. 부동산 신문기사 논조지수의 예측 유용성은 이어 추가적인 분석을 통해 실증했다.

〈표 20〉 Granger 인과관계 검정 결과

Granger Test	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
	F-value (p-value)	F-value (p-value)	F-value (p-value)	F-value (p-value)	F-value (p-value)
부동산 신문기사 논조지수 → 부동산 비기초가격	1.3784 (0.2419)	3.0849** (0.0482)	2.0394 (0.1103)	1.9166 (0.1101)	0.17769 (0.1206)
부동산 비기초가격 → 부동산 신문기사 논조지수	1.7074 (0.1930)	1.0873 (0.3394)	1.6044 (0.1903)	1.1595 (0.3307)	1.2049 (0.3094)

confidence level: *0.10 lower, **0.05 lower, ***0.01 lower

제 3 절 예측 유용성 분석

부동산 신문기사 논조지수가 부동산 비기초가격 예측력 향상에 유의미한 영향을 미치는지 검증하기 위해, 부동산 비기초가격에 대한 ARIMA 기본 모형과 부동산 신문기사 논조지수가 추가된 ARIMAX 모형을 구축하여 비교분석했다.

1) ARIMA 분석

1차적으로, 1차 차분한 변수(d=1)에 대한 ACF(Auto Correlation Function)와 PACF(Partial Auto Correlation Function)을 확인해, ARIMA의 AR 모형 p값과 MA 모형 q값을 추론했다. ACF 값은 두 개까지 스파이크를 보였으며, PACF는 세 개의 스파이크까지 유의수준을 넘어섰다. 시각화 결과를 바탕으로, ARIMA(p,d,q) 기준 모형을 ARIMA(3,1,2)로 설정하고, ARIMA 적정 모형을 결정하기 위한 차수를 최대 시차 3까지로 지정했다.

2차적으로, ARIMA 모형들의 LL(log-likelihood) 값과 AIC(Akaike Information Criterion) 값을 비교했다. LL은 모델의 설명력을, AIC는 모델의 적합도를 나타내는 지표이다. 일반적으로 LL 값이 크고 AIC 값이 작을수록 적절한 모형이라 평가한다. 이들 평가 지표를 비교해, ARIMA(1,1,3)을 적정 모형으로

선정했다(〈표21〉).

최종적으로, ARIMA(1,1,3) 모형의 잔차에 대한 독립성 검정을 시행했다. Box-Pierce 검정과 Ljung-Box 검정을 실시한 결과, Q 통계량의 p-value가 유의수준보다 컸으며(p>0.05), 잔차에 자기상관이 존재하지 않는다는 귀무가설을 기각하지 못했다. 이에 잔차의 자기상관이 나타나지 않은 ARIMA(1,1,3)을 최종 적정 모형으로 확정했다.

ARIMA(1,1,3) 모형은 1차 차분한 자료를 사용한 모형이며, AR(1)과 MA(3)를 기반으로 한다. AR(1) 모형은 1시점 전의 데이터가 현재 데이터에 영향을 준다고 가정한 모형으로, 연속된 1시점 정도의 데이터로 분석을 진행한다는 특징을 지닌다. 또한, MA(3) 모형은 직전 시점의 백색잡음, 시차가 2인 백색잡음, 그리고 시차가 3인 백색잡음의 결합으로 이뤄진 모형이다.

〈표 21〉 부동산 비기초가격의 ARIMA(p,d,q) 분석 결과

	LL	AIC
ARIMA(0,1,0)	234.93	-467.85
ARIMA(0,1,1)	234.97	-465.95
ARIMA(0,1,2)	243.97	-481.94
ARIMA(0,1,3)	246.41	-484.82
ARIMA(1,1,0)	234.94	-465.88
ARIMA(1,1,1)	241.32	-476.63
ARIMA(1,1,2)	244.60	-481.20
ARIMA(1,1,3)	250.97	-491.94
ARIMA(2,1,0)	241.36	-476.72
ARIMA(2,1,1)	247.76	-487.52
ARIMA(2,1,2)	249.72	-489.45
ARIMA(2,1,3)	251.43	-490.87
ARIMA(3,1,0)	247.33	-486.66
ARIMA(3,1,1)	248.70	-487.41
ARIMA(3,1,2)	250.70	-489.40
ARIMA(3,1,3)	251.52	-489.03

ARIMA(1,1,3) 모형의 예측오차는 〈표22〉와 같다. RMSE는 0.0161로, MAE는 0.0127로 나타났다.

〈표 22〉 ARIMA(1,1,3) 모형의 예측오차

	ME	RMSE	MAE	MPE	MAPE
ARIMA(1,1,3)	0.0005	0.0161	0.0127	8.6385	246.0255

2) ARIMAX 분석

앞선 ARIMA 모형 선정 과정과 동일한 절차를 거쳐, ARIMAX 적정 모형을 선별했다. 1단계로, 차분 변수에 대한 ACF와 PACF를 시각화해, ARIMAX의 AR 모형 p값과 MA 모형 q값을 추론했다. ACF 값은 세 개의 스파이크가 유의수준을 넘어섰고, PACF도 세 개까지 스파이크를 보였다. 해당 결과를 토대로 ARIMAX(p,d,q) 기준 모형을 ARIMAX(3,1,3)로 설정하고, ARIMAX 모형에 대한 차수를 최대 시차 3까지로 한정했다. Granger 인과관계 검정 결과를 참조해, 외생변수(X)인 부동산 신문기사 논조지수 자료의 시차는 2로 설정했다.

2단계로, ARIMAX 모형들의 LL 값과 AIC 값을 비교했다. AIC 값이 가장 작은 ARIMAX(2,1,3) 모형을 적정 모형으로 선정했다(〈표22〉).

마지막 단계로, ARIMAX(2,1,3) 모형의 잔차 독립성 검정을 시행했다. Box-Pierce, Ljung-Box 검정 결과, Q 통계량의 p-value가 유의수준을 넘어섰다(p>0.05). 이를 통해 ARIMAX(2,1,3) 모형의 잔차에 자기상관성이 존재하지 않음을 확인했다.

ARIMAX(2,1,3) 모형은 1차 차분한 자료를 사용한 모형이며, AR(2)과 MA(3)를 결합한 모형이다. AR(2) 모형은 2시점 전의 데이터와 1시점 전의 데이터가 현재 데이터에 영향을 준다고 가정한 모형으로, 연속된 2시점의 데이터를 분석 대상으로 한다. MA(3) 모형은 직전 시점의 백색잡음, 시차가 2인 백색잡음, 그리고 시차가 3인 백색잡음의 결합으로 이뤄진 모형이다.

〈표 23〉 부동산 비기초가격의 ARIMAX(p,d,q) 분석 결과

	LL	AIC
ARIMAX(0,1,0)	232.03	-460.06
ARIMAX(0,1,1)	232.08	-458.17
ARIMAX(0,1,2)	241.25	-474.49
ARIMAX(0,1,3)	243.39	-476.79
ARIMAX(1,1,0)	232.04	-458.08
ARIMAX(1,1,1)	238.60	-469.20
ARIMAX(1,1,2)	241.80	-473.61
ARIMAX(1,1,3)	248.01	-484.02
ARIMAX(2,1,0)	238.53	-469.06
ARIMAX(2,1,1)	244.87	-479.73
ARIMAX(2,1,2)	246.81	-481.62
ARIMAX(2,1,3)	249.18	-484.36
ARIMAX(3,1,0)	244.24	-478.47
ARIMAX(3,1,1)	245.74	-479.47
ARIMAX(3,1,2)	247.66	-481.31
ARIMAX(3,1,3)	249.23	-482.46

lag of the exogenous variable: 2

ARIMAX(2,1,3) 모형의 예측오차는 〈표24〉이며, RMSE는 0.0147, MAE는 0.0116이다.

〈표 24〉 ARIMAX(2,1,3) 모형의 예측오차

	ME	RMSE	MAE	MPE	MAPE
ARIMAX(2,1,3)	0.0004	0.0147	0.0116	64.4594	198.83261

이어 ARIMA 모형과 ARIMAX 모형의 RMSE, MAE의 값을 비교해 예측력의

정확도를 검증했다. 부동산 비기초가격 ARIMA 모형의 RMSE, MAE는 0.0161과 0.0127이다. 부동산 신문기사 논조지수가 추가된 부동산 비기초가격 ARIMAX 모형의 RMSE, MAE는 0.0147, 0.0116으로, 예측오차가 9.25%, 10.23% 향상되었다(〈표25〉). 이와 같이 제곱근 오차인 RMSE와 절댓값 평균인 MAE를 비교했을 때, 부동산 신문기사 논조지수가 부동산 비기초가격 변화 예측의 정확도를 향상시킴을 알 수 있다.

〈표 25〉 ARIMA(1,1,3) 모형과 ARIMAX(2,1,3) 모형의 예측오차 비교

	RMSE		MAE	
		change rate		change rate
ARIMA(1,1,3)	0.0161	-	0.0127	-
ARIMAX(2,1,3)	0.0147	9.25%	0.0116	10.23%

본문과 유사하게, 박재수·이재수(2021)는 신문기사 논조지수와 방송뉴스 논조지수가 아파트 매매가격지수 변화 예측의 정확도를 향상시키는지 분석했다. 신문기사 논조지수를 추가할 경우 RMSE와 MAE의 예측오차가 7.90%, 6.21% 향상했으며, 방송기사 논조지수를 추가할 경우 RMSE와 MAE의 예측오차가 각각 5.05%, 5.42% 향상했다. 본 연구에서도 부동산 비기초가격 기본 모형 대비 부동산 신문기사 논조지수 포함 모형의 예측력이 9.25%, 10.23% 향상해, 선행연구와 유사한 결과를 보였다. 나아가 부동산 비기초가격 예측의 본질적 어려움, 모형 예측오차의 감소율을 고려했을 때, 기존의 연구보다 일정 수준 발전된 결과를 도출했다.

제 4 절 소결

본 장에서는 부동산 비기초가격에 대한 부동산 신문기사 논조지수의 예측 유용성을 Granger 인과관계 분석, ARIMA 분석을 통해 검토했다.

분석 결과는 다음과 같다. 첫째, Granger 인과관계 분석 결과, 부동산 신문기사 논조지수는 시차 2에서 부동산 비기초가격에 대해 유의미한 인과관계를 가졌다. 반면, 부동산 비기초가격은 부동산 신문기사 논조지수에 별다른 관계를 보이지 않았다. Granger 인과관계 결과가 유의미하다는 것은 한 변수가 다른 변수의 예측에 유용하게 활용될 수 있음을 의미한다. 해당 결과는 부동산 신문기사 논조지수가 부동산 비기초가격 예측 모형의 외생변수를 고려될 수 있음을 보인다.

둘째, 부동산 비기초가격 기본 모형인 ARIMA 모형과 부동산 비기초가격에 외생변수로 부동산 신문기사 논조지수를 투입한 ARIMAX 모형을 비교했다. 분석 결과, ARIMA 모형에 비해 ARIMAX의 RMSE와 MAE가 각각 9.25%, 10.23% 감소하여 예측오차가 개선되었다.

종합하자면, 부동산 신문기사 논조지수가 부동산 비기초가격을 예측하는 데 유용하게 활용될 수 있다. 이는 부동산 시장에 내재된 심리적 요인을 측정하여 가격을 모니터링하고 예측하는 데 활용한다면, 기존의 임대가격 변수나 거시경제변수로 예측하기 어려웠던 변동적인 가격 흐름을 더욱 섬세하게 읽어낼 수 있다는 가능성을 시사한다.

제 7 장 결 론

부동산 가격에는 정량적으로 설명할 수 있는 영역과 설명할 수 없는 영역이 뒤엉켜 있다. 이론적으로는 부동산 가격이 임대 가치나 경기 동향으로 설명되어야 하나, 현실에서 그 외의 과열적 기대나 자기실현적 예측 등에 의한 가격의 증감이 존재한다. 그러나 관련한 학술적 논의가 파편화되어 있고 변수의 측정이 어려운 탓에, 부동산 비기초가격은 그 영향력과 중요성에 비해 연구가 더디게 진행되고 있다. 이에 본 연구는 부동산 신문기사 빅데이터를 활용하여 부동산 심리시장과 관련한 요인을 재구성했다.

본 연구는 서울시 아파트 매매가를 중심으로 부동산 비기초가격과 부동산 신문담론의 관계를 분석하는데 목적이 있다. 이를 위해, 먼저 부동산 비기초가격을 추출했다. 한국부동산원에서 제공하는 서울 아파트매매가격지수를 2014년 1월부터 2021년 12월까지 96개월간 월 단위로 수집했다. 서울 아파트매매가격지수를 내재가치로 설명되는 영역과 설명되지 않는 영역으로 분해하고, 전자를 부동산 기초가격으로 후자를 부동산 비기초가격으로 가정했다. 다음으로, 부동산 신문담론을 분석하기 위해 96개월간 4개 주요 일간지와 2개 주요 경제지를 웹 크롤링해 총 35,542개의 관련 기사를 수집했다. 토픽 분석을 활용해 부동산 신문기사를 정량화하고, 부동산 신문담론의 핵심 뼈대와 변화 추이를 파악했으며, 감성 분석을 통해 부동산 신문기사의 논조를 지수화했다.

부동산 비기초가격과 부동산 신문담론의 관계를 분석한 결과, 이들 사이에 유의미한 연관 관계가 존재했다. Granger 분석 결과, 부동산 비기초가격은 부동산 신문기사 중 부동산 정책 기사, 부동산 광고 기사, 생활환경 기사, 지역시장 기사 등과 유의미한 인과관계를 가졌다. 반면, 부동산 가격 기사, 부동산 공급 기사, 부동산 투자 기사 등 경제적 영역과 직접 연관된 기사와는 유의미한 인과관계를 보이지 않았다. 이는 부동산 가격을 설명하는 데에는 경제적 요인들뿐 아니라 정책, 심리, 정보 등과 같은 여타 요인들도 고려해야 한다는 이전 연구의 주장과도 상통

하는 결과이다.

더불어, VECM 모형 기반의 충격반응분석 및 분산분해분석을 통해, 부동산 신문담론이 내용의 구조적 변화가 부동산 비기초가격을 부풀리거나 누그러뜨리는 요인으로 작용할 수 있음을 확인했다. 충격반응분석 결과, 부동산 정책 기사의 충격은 부동산 비기초가격에 부정의 영향을 미쳤다. 이는 과세나 규제와 같은 부동산 정책이 부동산에 대한 기대를 어느 정도 진정시킬 수 있다는 것을 의미한다. 반면, 부동산 광고 기사와 생활환경 기사의 충격은 부동산 비기초가격에 긍정의 영향을 미쳤다. 이는 부동산에 대한 특정한 정보가 계속해서 반복되면 부동산에 대한 심리적 기대감을 부풀리고, 이러한 사회적 경향이 부동산 비기초가격에 반영될 수 있음을 내포한다. 한편, 지역시장 기사의 충격은 부동산 비기초가격에 부정의 영향을 미쳤다. 이는 해당 기사가 서울 인근의 수도권에 부동산 대장주 격의 지역을 대상으로 한다는 점을 고려해 해석할 필요가 있다. 주변 지역 부동산이 적극적으로 조명되면서 일시적으로 서울의 부동산이 지니는 상징성 내지 위치성이 약화될 수 있음을 의미한다.

나아가, 부동산 비기초가격에 대한 부동산 신문논조의 예측 유용성을 실증했다. 부동산 비기초가격의 ARIMA 예측 모형과, 비기초가격 모형에 부동산 신문기사 논조지수 외생변수로 투입한 모형의 ARIMAX 예측오차를 비교했다. 부동산 신문기사 논조지수가 부동산 비기초가격의 예측 정확도를 향상시키는지 분석한 결과, ARIMA 모형에 비해 ARIMAX 모형에서 RMSE, MAE 각각 9.25%, 10.23% 감소하여 예측오차가 개선되었다. 종합하면, 부동산 신문기사 논조지수가 부동산 비기초가격 예측 모형에 외생변수로 투입될 수 있으며, 통계적 유의성과 오차 측면에서 유의하게 활용될 수 있을 것이다.

이 연구는 부동산 가격에 대한 이해도를 높이고자 부동산 비기초가격과 부동산 신문담론의 관계를 분석했다. 부동산 가격은 부동산 기초가격과 부동산 비기초가격으로 이루어져 있으며 부동산 비기초가격은 심리적 요인으로 설명할 수 있다는 문제의식에 착안하여, 부동산 신문기사를 활용해 부동산 비기초가격을 설명했다. 본 연구는 신문기사 빅데이터를 활용해 정책이나 정보와 같이 정량화하기 어려운

요인을 변수로 구축할 방안을 제시했다. 또한, 기존에 선험적 혹은 이론적 판단에 의거해 부동산 비기초가격과 관련 있다고 가정되었던 요인들과의 관계를 실증적으로 파악했다. 본문의 결과는 기존 연구에서 부동산 비기초가격의 추정이 부동산 기초가격의 분석에 종속되었던 것에서 나아가 비기초가격에 대한 설명변수를 구축하여 추정방안을 고도화할 단초를 제공한다. 더불어, 부동산 정책이나 부동산 시장 심리를 더욱 적극적으로 지수화한다면 부동산 가격의 추세를 섬세하게 해석하고 예측할 수 있을 것이라는 가능성을 제시한다. 앞으로 부동산 비기초가격을 독해하는 연구가 계속해서 축적된다면 부동산 가격에서 일종의 미지 영역으로 간주하던 부분이 점차 축소될 수 있을 것이다.

다만 본문의 결과를 독해할 시, 이 연구의 한계에 대해 유념할 필요가 있다. 첫째, 부동산 비기초가격에 대한 추정이 기초가격에 대한 분석에 종속적이다. 부동산 비기초가격과 유의미한 관계를 맺는 요인에 대한 연구가 거의 부재한 만큼, 선행연구와 마찬가지로 본 연구도 부동산 시장가격에서 부동산 기초가격을 제외해 나머지 부분을 부동산 비기초가격이라 가정하였다. 그 때문에 연구의 문제의식과 연구의 한계가 맞닿아 있다는 역설적 한계가 있다. 둘째, 부동산에 관한 비정형 텍스트 데이터를 분석하기 위한 자료를 신문기사로 한정했다. 최근의 추세에 발맞춰 신문기사에서 방송기사, 유튜브 플랫폼, SNS, 프롭테크 등의 다양한 매체 자료를 폭넓게 활용하는 것도 추후 연구에서 고려해야 할 사항이다. 셋째, 수집한 부동산 신문기사로부터 부동산 감성사전을 더욱 정교하게 작성하기 위한 노력이 필요하다. 유관한 연구가 계속되고 있으나 각자의 발견과 성과가 하나로 누적되고 있지 않아, 부동산에 특화된 감성사전을 작성하는 방안에 대한 더욱 구체적이고 총체적인 연구가 요구된다. 넷째, 무엇보다도 부동산 비기초가격 변수와 부동산 신문기사 변수 모두 자료를 변수화하는 방안은 아직 학술적 합의가 종결되지 않은 영역이다. 따라서 불가피하게 본 연구에서 사용한 변수와 모형은 불완전하며, 시장 요소를 완전하게 포착하지 못하고 있음을 염두에 두어야 한다. 본 연구의 결과는 어떠한 절대적인 기준이라기보다 초기 연구의 탐색적 시도로서 이해할 필요가 있으며, 그럼에도 본문의 분석은 부동산 비기초가격 대한 후속 연구에 최소한의 방향성을 제시하는 자료로써 기능할 수 있으리라 생각한다.

참 고 문 헌

강준만, 2006,「강남, 낯선 대한민국의 자화상」, 서울: 인물과 사상사.

경정익·이국철, 2016. "Text Mining에 의한 부동산 빅데이터 감성분석 모형 개발",「주택연구」, 24(4): 115-136.

고현림·송선주·신종칠, 2018, “아파트 분양계약률에 영향을 미치는 특성요인들에 관한 분석”,「주거환경」, 16(4): 159-177.

구정우·조성권, 2017. "언론을 통해 본 한국의 ODA: 토픽 모델링을 활용한 신문기사 분석, 1993~2016,「국제・지역연구」, 26(3): 173-210.

국토교통부, 2017a, “주택시장의 안정적 관리를 위한 선별적 맞춤형 대응방안, 6월 10일, 보도자료.

국토교통부, 2017b, “실수요 보호와 투기수요 억제를 통한 주택시장 안정화 방안”, 8월 2일, 보도자료

국토교통부, 2018a, “수도권 주택공급 확대 추진 및 투기지역 지정 등을 통한 시장안정 기조 강화(주택정책과), 8월 17일, 보도자료.

국토교통부, 2018b, “주택시장 안정 방안 중 국토교통부 소관 정책 보도자료, 9월 13일, 보도자료.

국토교통부, 2019, “191216(3시 이후) 주택시장 안정화 방안(주택정책과), 12월 16일, 보도자료.

권순구·강명구, 2011. "부동산 수익에 대한 인식과 실제간 간극: 행동경제학적 해석",「서울도시연구」, 12(1): 127-138.

김경민, 2017. "통화정책 및 실물금융변수와 주택가격간 동학적 상관관계 분석",「부동산학보」, 70: 206-220.

김경민, 2018. "서울아파트시장과 거시경제변수 요인들간 동학적 상관관계 분석",「부동산학보」, 73(1): 115-129.

김경아·김선희, 2021. "지속가능발전목표 관점에서 토픽모델링을 활용한 지속가능한 도시 연구 동향 분석",「환경정책」, 29(1): 129-153.

김경환·이한식, 2000. "부동산 가격 거품과 가격 전망", 「대한부동산학회지」, 18: 59-81.

김대원·유정석, 2013. "주택가격에 대한 심리적 태도가 주택 매매 거래량에 미치는 영향 분석", 「주택연구」, 21(2): 73-92.

김동환, 2015. "VECM 모형을 이용한 주택시장과 거시경제변수 관계 분석", 「대한부동산학회지」, 41: 179-203.

김명수, 2021. "발전주의는 언제 일상이 되었는가?: 도시 중산층 가정성의 확립으로 본 현대 정착의 양상과 시간성".「경제와 사회」, 126-171.

김봉한, 2004. "부동산가격 버블의 존재 검정: 상태전환회귀식의 활용", 「주택연구」, 12(1): 71-96.

김상엽·엄수원·윤만식, 2020. "주택금융규제정책이 공동주택가격에 미치는 영향 연구", 「부동산학보」, 81(1): 87-102.

김선재·이영화, 2014. "벡터오차수정모형(VECM)을 이용한 인플레이션과 주요 경제변수와의 관계 분석", 「산업경제연구」, 27(3): 1139-1155.

김성용·오세준, 2016. "정부의 부동산 정책에 대한 언론의 보도 프레임 분석", 「한국콘텐츠학회 논문지」, 16(7): 221-230.

김원희·강원철, 2012. "부동산 버블 측정에 관한 연구 - 아파트를 중심으로", 「지역연구」, 28(2): 129-142.

김윤영, 2013. "한국 주택가격 변동은 펀더멘탈에 의해 주도되고 있는가?", 「경제학연구」, 61(4): 117-148.

김진유, 2006. "신문기사가 부동산가격변동에 미치는 영향 - '투기'가 포함된 신문기사와 주택가격간의 그랜저인과관계분석을 중심으로", 「주택연구」, 14(2): 39-63.

문태헌·김혜림, 2020. "신문기사 감성분석과 아파트매매실거래가의 관계 분석: 부산지역 사례", 「주거환경」, 18(3): 135-149.

문화체육관광부, 2022, 「문재인정부 국정백서(부동산_제13권 주택공급 확대와 집값 안정화를 위한 노력」, 서울: 문화체육관광부.

박승록, 2020.「(STATA를 이용한) 응용계량경제학」, 서울: 박영사.

박인권, 2022, “한국 주택정책 변천과정에 대한 비판적 고찰”, 「한국공간환경학회 학술대회」.

박재수·이재수, 2019. "벡터자기회귀모형을 이용한 온라인 뉴스기사와 아파트 매매가격의 동태적 관계 연구: 비정형 빅데이터를 활용한 감성분석 기법의 적용", 「감정평가학 논집」, 18(2): 83-113.

박재수·이재수, 2021. "부동산 감성지수의 주택가격 예측 유용성: 뉴스기사와 방송뉴스 빅데이터 활용 사례:, 「국토계획」, 56(4): 99-111.

박철수, 2008. "인쇄광고물 분석에 의한 서울 및 수도권의 1990년대 아파트 분양특성 연구", 「서울학연구」, 33, 75-105.

박헌수·이창원, 2012. "상태공간모형을 활용한 우리나라 지역별 아파트가격 버블 추정에 관한 연구", 「지역연구」, 28(4): 115-132.

방한솔·문호석, 2019, “텍스트마이닝을 이용하여 텍스트의 주요 토픽을 시계열적으로 표현하는 방법론 연구, 「한국데이터정보과학회지」, 30(6): 1259-1276.

방한솔·문호석, 2019. "텍스트마이닝을 이용하여 텍스트의 주요 토픽을 시계열적으로 표현하는 방법론 연구". 「한국데이터정보과학회지」, 30(6): 1259-1276.

백영민, 2020. 「R을 이용한 텍스트 마이닝」, 파주: 한울아카데미.

서정석·오지훈·김정섭, 2021. "부동산 뉴스와 주택경기의 동적 관계에 대한 고찰: 언론보도논조지수 개발을 중심으로". 「한국지역개발학회지」, 33(1): 89-112.

성채원, 2005. "한국의 지가변동 요인에 관한 분석", 연세대학교 대학원 석사학위논문.

신인섭·김병희, 2007, 「한국 근대 광고 걸작선 100: 1876~1945」, 서울: 커뮤니케이션북스.

신진호·황수성. 2015. "주택시장의 과신과 가격거품". 「부동산학연구」, 21(1): 5-29.

오근엽·김봉한·김흥기, 2005. "한국 주요 도시 아파트 가격의 버블 존재

검증 및 추정".「경제연구」, 23(3): 105-132.
우윤석·이은정, 2011. "언론보도와 시계열 주택가격 간의 관계에 관한 연구",「주택연구」, 19(4): 111-134.
윤성훈, 2020, “서울 지역별 아파트 거품 가능성 검토”, KIRI 리포트(포커스), 487, 1-31.
윤성훈·한성원. 2018. 서울 아파트 매매가격의 거품 가능성. KIRI 리포트(포커스), 440, 1-12.
이석원·정재호, 2017. "거시경제요인이 아파트가격 변동에 미치는 영향 연구 - 의사결정나무 방법론을 이용하여",「부동산학보」, 70: 28-41.
이영수, 2019. "통화정책과 주택시장: 부호제약을 이용한 베이지안 VAR 분석",「주택연구」, 27(1): 113-136.
이용만, 2000. "한국의 부동산시장은 비합리적인가: 주택시장을 중심으로 한 합리성 검정",「부동산연구」, 10: 49-64.
이우민, 2019. "주택정책에 대한 서울 자치구별 주택시장 반응의 공간적 차이 연구: VECM 모형을 이용하여", 서울대학교 대학원 석사학위논문.
이우민·김경민·김진석, 2019. "주택정책에 따른 서울 자치구별 주택시장 반응에 대한 연구",「한국경제지리학회지」, 22(4): 555-575.
이종민·이종아·정준호, 2017. "뉴스 빅데이터를 이용한 전세 가격 예측 - 토픽모형 분석을 중심으로 -",「부동산학보」, (69): 43-57.
이준구, 2017, “부동산 관련 정책에 관한 두 가지 단상”,「한국경제포험」, 9(4): 1-25.
이현재, 2001. "PHILLIPS - HANSEN 의 FM - OLS 공적분추정에 의한 실질자산효과 분석",「응용통계연구」, 14(2): 273-287.
이효섭·조대연, 2020. "토픽모델링과 키워드 네크워크분석을 활용한 국내 전환학습 연구 동향",「평생교육학연구」, 26(1): 1-24.
임대봉, 2015. "유동성이 주가 및 주택가격에 대한 파급효과 분석",「부동산학보」, 61: 80-93.
정민경·최규완, 2021. "텍스트 마이닝 기법을 활용한 COVID-19 발생

이전·이후의 배달주문플랫폼 서비스에 대한 주요 토픽 분석", 「지역산업연구」, 44(3): 283.
조경준, 2013. "주택 투자심리 변화가 주택시장 구조 변화에 미치는 영향", 「국토연구」, 76: 113-127.
조무상·남주하, 2019. "주택가격의 폭발적 거품의 추정", 「주택연구」, 27(1): 83-112.
조성배·신신애·강동석, 2018. "토픽 모델링을 이용한 개방형 혁신 연구동향 분석 및 정책 방향 모색", 「정보화 정책」, 25(3): 52-74.
조한국·김영곤, 2020. "아파트 가격 및 거래량과 거시경제변수 간의 동태적 관계연구 - 서울특별시를 중심으로 -", 「부동산학보」, 82: 45-58.
최차순, 2010. "주택가격 거품유무에 관한 연구", 「대한부동산학회지」, 30: 1-23.
홍정의·김형준·안세륭, 2022. "서울 아파트 가격은 거품인가?", 「부동산분석」, 8(1): 1-21.
홍지수, 2022. "강남 지역의 담론적 재현과 정체성 변환: 토픽 모델링을 활용한 신문 기사 분석을 중심으로, 1990~2021". 「한국도시지리학회지」, 25(2): 115-135.
Abraham, J. and Hendershott, P., 1996. "Bubbles in Metropolitan Housing Markets", Journal of Housing Research, 7(2): 191-207.
Akerlof, G. and Shiller, R., 2009. Animal Spirits: How Human Psychology Drives the Economy, and Why it Matters for Globay Capitalism, Princeton, NJ: Princeton University Press.
Bikhchandani, S., Hirshleifer, D., Welch, I., 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades". The Journal of Political Economy, 100(5): 992-1026.
Blanchard, O. and Watson, M., 1982. "Bubbles, Rational Expectations and Financial Markets". NBER Working Paper Series, 945.

Blei, D., 2012. "Probabilistic topic models", Communications of the ACM, 55(4): 77-84.

Bucchianeri. G., 2011. "The Anatomy of a Housing Bubble: Overconfidence, Media and Politics". SSRN Electronic Journal. doi:10.2139/ssrn.1877204.

Buenano-Fernandez, D., Gonzalez, M., Gil, D., and Lujan-Mora, S., 2020. "Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modeling Approach", IEEE Access, 8: 35318-35330.

Cagan, P., ed. 1956. "The Monetary Dynamics of Hyperinflation". in Studies in Quantity Theory of Money, edited by Friedman, M., 25-127. Chicage: University of Chicage Press.

Camerer, C., 1989. "Bubbles and Fads in Asset Prices", Journal of Economic Surveys, 3(1): 3-41.

Case, K. and Shiller, R., 2003. "Is There a Bubble in the Housing Market?", Brookings Papers on Economic Activity, 34(2): 299-362.

Engle, R. and Granger, C., 1987. "Co-Integration and Error Correction: Representation, Estimation, and Testing". Econometrica, 55(2): 251-276.

Griffiths, T. and Steyvers, M., 2004. "Finding scientific topics", Proceedings of the National Academy of Sciences, 5228-5235.

Gujarati, D.N. and Poter, D.C., 2009. 「(Gujarati의) 계량경제학」, 박완규・홍성표 역, 서울: 지필.

Johnson, M., 2010. "Studying How Changes in Consumer Sentiment Impact the Stock Markets and the Housing Markets", Ph. D. Dissertation, University of New Orleans.

Kahneman, D. and Tversky, A., 1979. "Prospect Theory: An

Analysis of Decision under Risk", Econometrica, 47(2): 263-292.
Kahneman, D. and Tversky, A., 1984. "Choice, Values and Frames", American Psychologist, 39(4), 341-350.
Levin, E. and Wright, R. 1997. "The impact of speculation on house prices in the United Kingdom". Economic Modelling, 14(4): 567-585.
Mercille, J. 2014. "The Role of the Media in Sustaining Ireland's Housing Bubble". New Political Economy, 19(2): 282-301.
Muellbauer, J. and Murphy, A. 1997. "Booms and Busts in the UK Housing Market". The Economic Journal, 107(445): 1701-1727.
Park, Min Jeong and Choi, Doo and Moon, Il Chul and Park, Sunyoung, Does News Affect the Housing Market? (March 24, 2016). Available at SSRN: https://ssrn.com/abstract=2754426
Phillips, P. and Hansen, B., 1990. "Statistical Inference in Instrumental Variables Regression with I(1) Processes". The Review of Economic Studies, 57(1): 99-125.
Roche, M., 2001. "The rise in house prices in Dublin: Bubble, fad or just fundamentals", Economic Modelling, 18(2): 281-295.
Schaller, H. and Van Norden, S., 2002. "Fads or bubbles?". Empirical Economics, 27(2): 335-362.
Shiller, R., 2000. Irrational Exuberance. Princeton, NJ: Princeton University Press.
Shiller, R., 2002. "Bubbles, Human Judgment, and Expert Opinion". Financial Analysts Journal, 58(3): 18-26.
Shiller, R., 2021. 「내러티브 경제학: 경제를 움직이는 입소문의 힘」, 박슬라 역, 서울: 알에이치코리아.
Simon, H., 1959. "Theories of Decision-Making in Economics and Behaviroal Science", The American Economic Review, 49(3):

253-283.

Stiglitz, J., 1990. "Symposium on Bubbles". The Journal of Economic Perspectives, 4(2): 13-18.

Summers, L., 1986. "Does the Stock Market Rationally Reflect Fundamental Values?". The Journal of Finance, 41(3): 591-601.

Thaler, R., 2009. Nudge: Improving Decisions about Health, Wealth, and Happiness, New York: Penguin Books.

Tversky, A. and Kahneman, D., 1991. "Loss Aversion in Riskless Choice: A Reference Dependent Model", The Quarterly Journal of Economics, 106(4): 1039-1061.

Tversky, A. and Kahneman, D., 1992. "Advanced in Prospect Theory: Cumulative Representation of Uncertainty", Journal of Risk and Uncertainty, 5:297-323.

Van Norden, S. and Vigfusson, R., 1998. "Avoiding the Pitfalls: Can Regime-Switching Tests Reliably Detect Bubbles?", Studies in Nonlinear Dynamics and Econometrics, 3(1): 1-22.

Van Norden, S., 1996. "Regime switching as a test for exchange rate bubbles", Journal of Applied Econometrics, 11(3): 219-251.

Weber, W., and Devaney, M., 1996. "Can Consumer Sentiment Surveys Forecast Housing Starts?", The Appraisal Journal, 64(4): 343-350.

Abstract

Bridging the Gap: Using News Articles on Real Estate to Explain and Forecast Non-Fundamental Real Estate Prices

Hong, Ji-Su

Department of Environmental Planning

Graduate School of Environmental Studies

Seoul National University

This study investigates the relationship between non-fundamental real estate prices and newspaper discourse on real estate. It is based on the theory that media coverage leads public sentiment in certain ways, and that public opinion has a substantial impact on market price. Non-fundamental real estate prices were estimated using fully modified OLS (FM-OLS), and newspaper discourse on real estate variables were constructed through topic modeling.

The time series relationship between non-fundamental prices and newspaper discourse was confirmed through Granger causal analysis, impulse response analysis, and variance decomposition

analysis. The results are as follows: First, non-fundamental real estate prices were negatively associated with newspaper discourse on real estate policy. Second, there was a positive relationship between non-fundamental real estate prices and information on real estate advertisements. Third, non-fundamental real estate prices had a negative relationship with information on the nearby metropolitan area. This result suggests a way to construct variables that are difficult to quantify, such as policies and information, using big data from newspaper articles.

The predictive usefulness of real estate newspaper articles for non-fundamental prices was conducted through ARIMA analysis. The sentiment index for real estate newspaper articles was constructed through topic analysis and sentiment analysis, and it was confirmed whether the sentiment index is useful for predicting non-fundamental real estate prices. The results are as follows: The analysis revealed that adding the sentiment index as an exogenous variable reduced the prediction errors of non-fundamental real estate prices by about 10%. This suggests that the sentiment index of real estate newspaper articles improves the accuracy of predicting changes in non-fundamental real estate prices.

Overall, this study analyzed the relationship between non-basic real estate prices and real estate newspaper discourse to enhance understanding of real estate prices. This study used real estate newspaper articles to explain non-fundamental real estate prices. The study proposed a way to construct variables for factors that are difficult to quantify, such as policies or information, using big data from real estate newspaper articles. Furthermore, the study suggested the possibility of using real estate newspaper articles to

analyze and predict real estate prices more delicately. If research continues to accumulate using new data to analyze the real estate market, the area that was once considered a kind of unknown territory in real estate prices could gradually shrink.

Keywords : Real Estate Price, Non-Fundamental Real Estate Price, Newspaper Articles, Newspaper Articles on Real Estate, Big Data

Student Number : 2021-29070