

# 포털 검색 지수를 활용한 전세 가격 예측

- 네이버·구글을 중심으로 -

The Jeonse Price Forecasting used by Portal Search Index

- Focusing on Naver · Google Trend -

이 종 민\* · 이 종 아\*\* · 정 준 호\*\*\*

Lee, Jong Min · Lee, Jong Ah · Jeong, Jun Ho

## 目次

- |            |             |
|------------|-------------|
| I. 서론      | IV. 분석결과    |
| II. 선행 연구  | 1. 기본적 검정   |
| III. 분석 설계 | 2. 기본모형 설정  |
| 1. 분석자료    | 3. 유용성 분석   |
| 2. 분석모형    | V. 결론 및 시사점 |
|            | <abstract>  |
|            | <참고문헌>      |

## ABSTRACT

### 1. CONTENTS

#### (1) RESEARCH OBJECTIVES

This study is an analysis of the jeonse price prediction using portal Naver and Google's search index. If the portal search index is effective to predict the jeonse price, its object to portal search index is statistically confirmed that the most effective for a period of time.

#### (2) RESEARCH METHOD

This research used the data of KB Kookmin Bank jeonse price index and Naver and Google portal search index from July 2010 to December 2015 applied the granger causality test and cross correlation analysis and compared the MAE of the ARIMA model, including jeonse portal search index and jeonse price change rate ARIMA model.

#### (3) RESEARCH FINDINGS

After analyzing cross correlation analysis portal search index go ahead of the jeonse price, and granger causality test found that the portal search index affects the jeonse price. The model of including

\* 주 저 자 : 강원대학교 부동산학과, 박사과정 수료, goodrealtor@korea.com

\*\* 공동저자 : KB금융지주, 선임연구위원, 부동산학 박사, jongah01@hanmail.net

\*\*\* 교신저자 : 강원대학교 부동산학과, 교수, D.Phil. jhj33@kangwon.ac.kr

▷ 접수일(2016년 10월 23일), 수정일(1차:2016년 11월 28일, 2차:2016년 12월 20일), 게재확정일(2017년 2월 10일)

the portal search index proved to be more improvement in predictive power than the jeonse price univariate time series models at 6~7 months time lags

## 2. RESULTS

This study showed that the participants in jeonse market search their jeonse stuff, jeonse market conditions, loan jeonse, via the portal 6 to 7 months ahead of their jeonse contract expiration date or the new jeonse contract. It is significant that the portal search index is based on providing more accurate result of price prediction jeonse to jeonse market participants with the usage of the portal search index.

## 3. KEY WORDS

- ARIMA, Cross correlation, Granger causality test, Jeonse, Search index

## 국문초록

본 연구는 Naver·Google포털의 검색 지수를 활용한 전세가격예측에 관한 실증분석이다. 이를 위하여 KB국민은행의 전세가격변동률 ARIMA 모형과 여기에 포털 검색 지수를 포함한 모형의 MAE를 비교분석하였다. 인과관계분석 결과 네이버 PC와 구글의 검색 지수가 일방향으로 전세가격변동률에 영향을 미쳤으며, 교차상관분석에서는 네이버 PC 검색 지수가 6~7개월의 시차를 두고 전세가격변동률과 동조하는 것으로 확인되었다. 전세가격변동률 단일변량 ARIMA모형과 포털 검색 지수를 포함한 모형의 예측력 비교 결과, 포털 검색 지수를 포함한 모형의 예측력이 6~7개월의 시차를 두고 향상되는 것으로 나타나, 앞으로 전세가격 예측시 포털 검색 지수를 활용한다면 보다 효과적이고 선제적인 정책입안과 시장대응이 가능할 수 있을 것으로 판단된다.

**핵심어 :** 교차상관분석, 예측력, 인과관계분석, 전세, 검색 지수, 아리마모형

## I. 서론

빅데이터 기술의 발전은 그동안 제한적 데이터로 시장을 예측해왔던 분석에 새로운 패러다임의 변화 가능성을 제시하고 있다. 이는 더 많은 데이터의 양과 빈도 측정이 가능하다는 것 이외에도 정형 데이터뿐만이 아니라 비정형 데이터까지도 일정한 패턴과 흐름을 분석하고 시장의 변화를 예측하는데 사용될 수 있기 때문이다. 실제 2008년 구글에서 독감 증상에 관한 키워드의 검색

빈도를 파악해 독감의 확산을 예측한 시도<sup>1)</sup>는 이러한 변화의 대표적인 사례라고 볼 수 있다. 최근의 경제 예측에도 이와 같이 비정형 빅데이터를 결합하여 분석하는 연구가 지속되고 있다. 경제활동의 예측은 소비자의 구매패턴<sup>2)</sup> 및 행태 분석을 통해 제품의 효율적 생산 및 서비스 방향 등을 결정할 수 있도록 지원하는 것은 물론 자산의 위험관리 차원에서도 유용한 데이터를 제시할 수 있기 때문이다. 한편, 예측 연구에 활용되는 비정형 빅데이터는 대체로 검색 포털사이트에서

1) 미국질병통제예방센터는 2008년 2월에 애틀란타 주의 독감 유행에 대한 발생을 공표하였으나, 이미 구글에서는 검색 트렌드 분석을 통해 2주전에 독감 유행 가능성으로 예측 보도함 (정지선, 새로운 미래를 여는 빅데이터 시대, 한국정보화진흥원, 2013, p.31)  
 2) 김현동·이상엽, “백화점형 할인점 개발을 위한 빅데이터 기반 소비자 행태 및 입지특성 분석에 관한 연구”, 부동산학회, 2015, 63집, p.174.

제공하는 검색 지수<sup>3)</sup>가 대부분인데, 이는 소비자가 직접 검색하는 키워드 중심으로 데이터가 만들어지므로 소셜네트워크나 블로그의 텍스트 분석보다 작성자의 의도성이나 오류가 적다<sup>4)</sup>.

검색 키워드 형태의 빅데이터를 이용한 경제활동 예측 연구로는 구글 검색 지수를 활용한 Choi and Varian(2009, 2011)<sup>5)</sup>과 야후 검색 빈도를 이용한 Goel et al. (2010)<sup>6)</sup>등이 대표적이는데, 해당 연구에서는 검색 키워드를 감안한 모형이 기존 모형보다 대체로 신규 주택판매나 비디오 게임 매출 등의 예측과 상관관계(correlation)향상에 더 유용하다는 것을 증명하고 있다. 김민희·김나경(2013)<sup>7)</sup>은 구글 트렌드를 통하여 브랜드순위와 검색량의 순위가 유사한 형태로 움직이고 있음을 실증하였다.

한편, 경제 지표와 자산에 대한 빅데이터 연계 분석은 대체로 실업률과 주식, 부동산 등으로

집중되고 있다. Ettedge et al. (2005)<sup>8)</sup>와 Tuhkuri(2015)<sup>9)</sup>는 미국의 실업률을, 나중화·김은섭(2013)<sup>10)</sup>과 권치명 외(2015)<sup>11)</sup>는 한국의 실업률에 포털 검색 지수가 유용한지에 대한 연구를 진행하였는데 이를 통해 대체로 실업률 연계 검색어의 검색 빈도 확대는 단기 예측에 유용성이 있다는 결과를 도출하고 있다. 주식의 경우 Preise et al.(2010)<sup>12)</sup>와 Bordino et al.(2012)<sup>13)</sup>등이 S&P 500, Nasdaq - 100 등의 주식데이터를 활용해 검색어의 연관성으로 분석하였으며, 검색 지수와 부동산에 대한 연구로는 Shichun et al.(2013)<sup>14)</sup>와 Wu and Brynjolfsson(2009)<sup>15)</sup>가 각각 중국과 미국의 주택가격예측에 구글의 검색어가 미치는 영향을 점검하였다. 우리나라에서도 김민희(2014)<sup>16)</sup>, 김대원·유정석(2014)<sup>17)</sup>등이 주택 매매, 전세 시장과 포털 검색 지수와의 관계에 대해 실증 분

- 3) 네이버와 구글은 포털을 통하여 검색된 검색어와 검색 횟수를 활용한 검색 지수를 0~100 숫자로 환산하여 상대적 지수로 표시하고 있다. 특정 검색어가 가장 많이 검색된 주간 최대값을 100으로 정하고 나머지 기간의 검색 횟수를 상대값(정수)으로 환산하여 지수로 제공한다.
- 4) 이궁희·황성필, "빅데이터를 이용한 경기판단지표 개발: 네이버 검색 경기지수 작성과 유용성 검토", 경제분석, 한국은행 경제연구원, 2014, 제20권 4호, p.5.
- 5) Hyunyoung Choi, Hal Varian, "Predicting the Present with Google Trends", ©Google Inc, 2009, pp.1-19.  
Hyunyoung Choi, Hal Varian, "Predicting the Present with Google Trends", ©Google Inc, 2011, pp.1-12.
- 6) Sharad Goel, Jake M. Hofman 외 3, "Predicting consumer behavior with Web search", PNAS, 2010, Vol.107, No.41, pp.17486~17490.
- 7) 김민희·김나경, "검색 데이터를 보면 소비트렌드가 보인다", LG Business Insight, LGERI, 2013.11.13, pp.34~36
- 8) Ettredge, M., Gerdes, J., Karuga, G, "Using web-based search data to predict macroeconomic statistics" *Communications of the ACM*, 2005, Vol.48, No.11, pp.87~92.
- 9) Joonas Tuhkuri, "Big Data : Do Google Searches Predict Unemployment?", *University of Helsinki Faculty of Social Sciences, Economics*, 2015, pp.1-83.
- 10) 나중화·김은섭, "소셜 미디어 정보를 이용한 실업률 예측", 한국산업정보학회지, 한국산업정보학회, 2013, 제18권 6호, pp.95~101.
- 11) 권치명·황성원·정재운, "실업률 예측을 위한 인터넷 검색 정보의 활용", 한국시물레이션학회 논문지, 한국시물레이션학회, 2015, 제24권 2호, pp.31~39.
- 12) Tobias Preis, Daniel Reith, H. Eugene Stanley, "Complex dynamics of our economic life on different scales : insights from search engine query data", *The Royal Society*, 2010, pp.5707~5719.
- 13) Ilaria Bordino, Stefano Battiston,외 3, "Web search queries can predict stock market volumes", *Plos one*, 2012, Vol.7, Issue7, e40014, pp.1~17.
- 14) Han Shichun, Wang Ping, Wang Bian, "A study on correlation between web search data and housing price evidence from Beijing and Xi'an in China", *ICSSR*, 2013, pp.603-608.
- 15) Lynn Wu, Erik Brynjolfsson, "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales", *Technical report, MIT*, 2009, pp.1~24.
- 16) 김민희, "검색데이터 주택시장의 단기예측에 유용하다" LG Business Insight, LGERI, 2014.7.2, pp.2~13.
- 17) 김대원·유정석, "인터넷 검색 활동과 주택 가격 및 거래량 간 동적 관계 분석", 부동산연구, 한국부동산연구원, 2014, 제24집 2호, pp.125~140.

석을 진행하였다.

이에 본 연구는 국내의 포털 검색에 가장 대표적인 네이버와 구글<sup>18)</sup>을 이용해 전세가격 예측<sup>19)</sup>에 포털 검색 지수의 유용성 여부를 분석하고, 유용성이 있다면 포털 검색의 시차(lag)가 어느 정도 일 때 예측력이 전세가격지수 단일 변수의 모형보다 통계적으로 유의성이 향상되는가를 밝히는 것을 목적으로 분석을 시도한다.

본 연구는 기존 연구가 한 개의 포털 검색을 활용한 것과 달리 네이버와 구글 모두를 분석하여 두 개의 검색포털의 차이를 점검하고자 하며, 주간 단위의 변수를 사용하여 보다 정밀한 예측력을 도출하고자 한다.

이를 위하여 해당 논문은 연구의 범위를 2010년 7월부터 2015년 12월까지의 기간으로 하고 KB전세가격지수와 네이버·구글의 '전세'와 '전세대출' 검색 지수를 연구대상으로 하여 시장 예측에 가장 보편적으로 활용되는 ARIMA모형을 통해 예측력에 미치는 효과를 측정하고자 한다.

이러한 연구는 최근 주거 불안에 따른 사회적 문제로 대두되고 있는 전세시장의 변화<sup>20)</sup>를 선제적으로 파악하고 이를 활용해 보다 빠르고 정교한 정책 수립은 물론 실수요자의 전세 거래에 도움이 될 수 있을 것으로 판단한다.

## II. 선행 연구

포털 검색활동을 경제 분야에 연관되어 연구한 최초의 논문은 Ettredge et al. (2005)<sup>21)</sup>의 연구로 미국의 실업율과 포털 검색량과의 관계를 언급한 것이라고 알려져 있다. 이후 포털 검색 지수 및 소셜 네트워크의 활동량을 이용하여 예측을 시도하는 연구는 크게 소비자의 경제 활동이 반영된 포털 검색이 제품의 판매등을 예측하는 것과, 주요 포털 검색어와 경제지표의 관계 및 예측력 개선 등을 분석하는 흐름으로 나뉜다.

Choi and Varian(2009)<sup>22)</sup>은 구글의 검색 지수를 포함한 AR모형을 사용하여 미국의 자동차 판매, 주택판매등에 대한 예측력을 비교하였다. 결과적으로 자동차판매 부문에서 MAE (mean absolute error)<sup>23)</sup>가 18.4%, 신규주택 부문에서는 12%의 예측력의 개선 효과가 있었음을 확인했다. Goel et al. (2010)<sup>24)</sup>은 Yahoo 검색량을 이용하여 box-office의 오픈 주말 수익과, 첫달 비디오게임의 매출, Billboard hot 100 chart의 노래순위를 AR모형을 통하여 상관관계를 비교하였는데, 기준모형 대비 검색량이 포함된 모형의 Billboard Rank는 24.3%의 상관관계(correlation) 향상의 결과를 보였다. 우리나라에서도 이와 유사한 연구가 진행되었는데 김민희·김나경(2013)<sup>25)</sup>은 네이버 검색 데이터 분석을 통하여 소비자들의 년도별 스마트폰 구매 결정에 미치는 요인에 차이가 있음을 증명하고, 구글 트렌드를 통하여 의

18) 국내 포털검색의 시장점유율은 2015년 12월 기준 네이버가 84.7%를 차지하는 등 압도적 비율을 보이고 있으며, 해외에 높은 점유율을 보이고 있는 구글은 상대적으로 1.2%수준에 국한되어 있음.

19) 이윤홍, "전국 아파트 전세변동률 기준 시장 유형화와 유형별 전세가격 변동률 영향요인 분석", 부동산학보, 한국부동산학회, 2016, 64집, pp.212~226.

20) 최근 국내 전세시장은 부족한 공급 물량과 급등하는 전세가격 등으로 시장의 불안이 커져가고 있다. 2015년 12월을 기준으로 최근 5년간 전국 매매가격은 17.6% 상승한 반면 전세가격은 43.0% 상승하며, 임차 거주자의 주거비용이 크게 증가하는 현상이 발생함.

21) Ettredge et al., *op.cit.*, pp.87~92.

22) Choi and Varian, *op.cit.*, pp.1~19.

23)  $MAE = \frac{1}{T} \sum_{t=1}^T |Y_t - \hat{Y}_t|$  여기서  $Y_t$ 는 실제값,  $\hat{Y}_t$ 는 예측값

24) Goel et al., *op.cit.*, pp.17486~17490.

25) 김민희·김나경, 전계논문, pp.32~34.

류 브랜드 갭(GAP), 자라(ZARA) 등의 브랜드 순위가 검색량의 순위와 유사하게 움직인다고 주장하였다. 또한, 권치명 외(2014)<sup>26)</sup>는 질병관련 네이버 포털 검색어를 조합한 회귀모형을 질병 감지에 활용할 경우 현재의 질병 감지 체계보다 1~2주 정도 선행하여 유행성 독감을 감지하는 것으로 분석했다.

경제지표의 관계 및 예측력에 대한 연구로 Preis et al.(2010)<sup>27)</sup>은 주간 구글 검색어 조사를 통하여 S&P 500 기업의 주식거래량과 해당 기업 이름 인터넷 검색량이 관계가 있으며, 현재의 주가 흐름은 차주의 기업 이름 검색량에도 영향을 미치고 있음을 발견하였다. Bordino et al.(2012)<sup>28)</sup>은 NASDAQ-100 주식 거래량과 해당주식의 Yahoo 검색량의 교차상관분석 및 그랜저 인과분석을 통하여 일일주식거래량과 검색량이 연관이 있으며, 인터넷 검색량 분석은 금융시스템위험을 조기에 발견하는데 기여할 수 있다고 주장하였다. Latoeiro et al.(2013)<sup>29)</sup>은 구글 검색 지수를 활용하여 구글의 'EURO STOXX' 검색활동이 EURO STOXX Index의 거래규모와 변동성에 선행하고, 1주 단위의 단기 예측에 유용하다는 것을 검증하였다. 이궁희·황상필(2014)<sup>30)</sup>은 네이버 경기관련 검색어를 이용하여 만든 경기지수와 교차상관분석 결과 네이버 경기지수<sup>31)</sup>가 경제심리지수, 경기동행지수에 2개월 선행하였고, 경제성장률과 민간소비증

감률의 예측력을 비교했을 때에도 네이버 검색 경기지수를 포함한 모형의 예측력이 특정기간<sup>32)</sup>에 단순 경제지수보다 예측력이 상대적으로 우수한 것으로 조사됐다.

한편 주식이외의 연구로는 Tuhkuri(2015)<sup>33)</sup>는 실업률지수의 AR기준모형과 구글 'Unemployment benefits' 검색 지수를 추가한 모형의 예측력을 비교한 결과, 포털 검색 지수를 포함한 모형이 단기예측력에 있어서 유용성이 있음을 증명한 연구가 있다. 국내에서도 실업률에 대한 분석을 시도한 연구로 나중화·김은섭(2013)<sup>34)</sup>과 권치명 외(2015) 논문이 있는데 나중화·김은섭(2013)은 구글 검색어 '채용'을 ARIMAX 모형에 포함하여 분석한 결과, ARIMA 모형보다 실업률 예측력(MAE)이 30.8% 개선이 있음을 확인하였다. 또한 권치명 외(2015)<sup>35)</sup>는 주간 네이버 검색 지수를 월간으로 변환하여 실업률 ARIMA 기준모형에 해당 검색 지수를 추가한 결과 MSE<sup>36)</sup>가 53.2% 우수한 것을 확인하였다.

본 연구와 유사한 부동산관련 학술 연구는 아직까지 주식이나 실업률에 비해 국내외적으로 활발하지 않다. 이는 부동산관련 데이터의 주기가 길어(월간 또는 분기) 주간으로 생성되는 포털 검색 자료와 비교가 용이하지 않기 때문일 것으로 추정된다. 그럼에도 불구하고 부동산 중주 택에 대한 다양한 분석 방법론으로 접근한 연구가 일부 존재한다. 우선 Wu and Brynjolfsson

26) 권치명·황성원·정재운, "인터넷 검색어를 활용한 유행성 독감발생 감지", 한국시물레이션학회 논문지, 한국시물레이션학회, 2014, 제 23권 4호, pp.31~39.

27) Preis et al., *op.cit.*, pp.5707~5719.

28) Bordino et al., *op.cit.*, pp.1~17.

29) Pedro Latoeiro, Sofia B. Ramos, Helena Veiga, "Predictability of stock market activity using Google search queries", *Universidad Carlos III de Madrid Working paper*, 2013, Vol.13, No.06, pp.1~33.

30) 이궁희·황상필, 전계논문, pp.1~37.

31) 호황검색통계의 평균에서 불황 검색통계의 평균을 차감한 후 100을 더한 지수(이궁희, 황상필 2014)

32) 금융위기 기간: 2008.4/4 ~ 2009.4/4

33) Tuhkuri, *op.cit.*, pp.1~83.

34) 나중화·김은섭, 전계논문, p.99.

35) 권치명·황성원·정재운, 전계논문, p.37.

36) MSE(mean square error) 예측치의 평균제곱오차 :  $\frac{1}{T} \sum_{t=1}^T |PE_t|^2$

(2009)<sup>37)</sup>는 미국의 주택판매량과 주택가격을 기본으로 하는 AR모형보다 구글의 'Real Estate Agencies'와 'Real Estate'의 검색어를 추가로 포함한 모형의 MAE가 76.9% 향상된다고 주장하였다. Shichun et al.(2013)<sup>38)</sup>은 구글 검색 지수와 주택가격과의 상관관계를 최소자승법을 통하여 비교한 후 중국 베이징과 시안을 비교분석하였다. 이 결과 베이징은 2개월, 시안은 1개월 주택가격지수보다 검색 지수가 선행하는 것을 확인하였다.

국내도 포털 검색 지수와 부동산 관련 연구가 초기 단계에 있다. 김민희(2014)<sup>39)</sup>는 교차상관계수를 통하여 구글 트렌드 전세검색량과 전세가격 상승률은 2~5개월의 시차를 두고 같은 방향으로 움직이는 것을 확인하였다. 김대원·유정석(2014)<sup>40)</sup>은 포털 검색활동과 주택가격 및 거래량간의 관계를 분석하기 위하여 서울시 23개 행정구를 패널로 하고 네이버 PC 트렌드를 검색 변수로 Arellano-Bond 동적패널 추정결과 구매자의 포털 검색활동은 1~4개월의 시차를 두고 영향을 미치는 것으로 확인하였다.

종합적으로 본 논문과 관련성 있는 선행연구에 대한 검토 결과, 기존 연구들과 본 연구의 차이점은 다음과 같다. 첫째, 본 연구는 포털 검색 지수를 이용한 전세 가격 예측력 향상 분석을 위해 ARIMA 모형으로 분석을 시도하는 첫 번째 연구로 보인다. 둘째, 본 논문은 구글이나 네이버를 선택해 예측력을 비교한 기존 선행연구와 달리, 구글과 네이버의 검색 지수를 모두 분석하고 그 결과를 비교하여, 각 포털별 차이를 점검하고자 하였다. 한편, 네이버 내에서 제공되는 검색 접근 경로별 검색어 지수를 활용하여 PC와 모바일 차

이 역시 비교할 수 있도록 하였다. 셋째, 대부분의 부동산 선행연구가 월간 검색 지수 자료를 이용하였으나, 본 연구는 2010년 7월부터 2015년 12월까지의 주간 자료(276주)<sup>41)</sup>를 활용하여 보다 정밀한 예측력 향상 정도의 검증을 시도하였다.

### Ⅲ. 분석설계

#### 1. 분석자료

전세가격지수는 KB국민은행에서 주간단위로 발표되는 아파트 지수<sup>42)</sup>를 활용하였으며, 네이버 모바일 검색 지수가 제공된 시점인 2010년 7월부터 2015년 12월까지를 사용하였다. 전세가격과 상관관계가 높을 것으로 추정되는 검색어의 선정은 의미가 명확하고 대표성을 가지는 '전세'와<sup>43)</sup> 검색 연관성이 높게 나타난 '전세대출'을 변수로 선정하였으며, 이에 대한 자세한 기준은 다음과 같다.

#### 1) 전세가격지수

전세가격지수의 선정에 있어서 KB국민은행은 전국 전세 가격 지수이외에도 각 지역별 기준으로 발표하고 있으나 각 포털에서 제공하는 검색 지수가 지역별로 제공되지 않아 전국 아파트 전세가격지수를 전세가격의 기준값으로 선정하였다. 한편, 본 연구에서 선정한 기간내 전세가격지수가 약 50.9% 상승한 것과 유사하게 서울아파트 역시 동일 기간 약 51.1% 상승하여 전세가격 상승세가 전국적인 흐름이었음을 알 수 있다.

37) Wu and Brynjolfsson, *op.cit.*, p.14.

38) Shichun et al., *op.cit.*, pp.603~608.

39) 김민희, 전계논문, pp.2~13.

40) 김대원·유정석, 전계논문, pp.125~140.

41) KB전세가격지수는 설날등 연휴가 포함된 날은 지수를 발표하지 아니하여, 1년 52주간 일치하지 않음

42) 통계청 발표 2015년 인구주택총조사 전수집계결과 전세주택에서 아파트가 차지하는 비중이 59.9%, 공동주택 비중은 74.5%으로 아파트가 대표적인 주택형태로 자리잡고 있음

43) 김민희, 전계논문, p.5.

2) 포털 검색 지수

네이버, 구글에서는 검색 통계를 Trend 라는 사이트를 통하여 제공하였다. 본 연구에서는 국내 포털 점유율이 가장 높은 네이버 모바일과 PC의 검색 지수와 경제대용변수로 활용가능 여부를 분석한 논문(나중화 외 2013, 김민희 2014)<sup>44)</sup>에서 변수로 인용된 구글의 검색 지수를 사용하였다.

〈표 1〉 국내 검색 포털 시장 점유율

검색포털	기간내 평균	기간 시작	기간 끝
Naver	75.62 %	55.60 %	84.66 %
Daum	17.32 %	32.39 %	11.83 %
Google	3.77 %	4.66 %	1.21 %

주 : 1) 기간: 2010. 7월~2015. 12월  
 2) 출처: [http://trend.logger.co.kr\(trendforward.tsp\)](http://trend.logger.co.kr(trendforward.tsp))

네이버는 포털 검색에서 발생하는 검색어 통계정보 service를 beta 버전으로 제공하고 있다 (<http://datalab.naver.com>). PC검색 통계는 2007년 1월부터, 모바일 검색 통계는 2010년 7월부터 주간단위로 제공한다.<sup>45)</sup> 네이버 검색 지수는 사람들이 검색한 검색량을 단순히 합하여 제공하지 않고 상대값으로 발표하고 있다. 검색지수는 검색횟수를 0~100 숫자로 환산하여 상대적 지표로 제공되고 있다. 특정한 검색어 검색이 가장 많이 되었던 주간 최대값을 100으로

정하고 나머지 기간의 검색횟수를 상대값으로 변환하여 정수인 지수로 표시하고 있다. 검색 지수는 최대값을 기준으로 작성되기 때문에 별도로 검색된 검색 지수간 값들을 서로 비교할 수 없다. 검색 지수 0은 검색이 없었던 것이 아니라 최대 검색횟수에 비해 검색횟수가 충분하지 않음을 의미한다.<sup>46)</sup> 검색어는 분석하고 싶은 '주제어'별로 최대 20개까지 입력할 수 있으며 띄어쓰기한 공백은 제거되고 검색어들 사이에 콤마 단위로 구분하여 입력된다. 그러나 검색어 맞춤법에 오류가 있거나 다르게 표현된 것은 다른 검색어로 처리된다. 구글 트렌드 검색통계도 역시 주 단위로 작성되는데 네이버와 유사하게 최대 검색량을 100으로 하여 상대적 값으로 지수형태로 표시된다.

3) 검색어 선정

검색어 중 전세가격지수 예측에 가장 많은 영향을 미치는 검색어로 해당 시장을 가장 대표적으로 표현하는 '전세'라는 검색어를 선정하였다. 통상 전세(Jeonse)는 우리나라 고유어로 검색어에는 거래 시장과 가격 시장을 통칭하여 사용되는 경향이 존재하며 기존 선행연구에서도 유의성이 검증된 바 있다.<sup>47)</sup> 그 외 검색어로는 구글 검색어 '전세' 검색 지수와 관련 검색어로 가장 높은 순위로 선정된 '전세대출'을 선택하였는데, 전세 관련

〈표 2〉 분석변수의 기초통계량

변수명	표기	평균	최대값	최소값	표준편차	관찰수
전세가격지수	JS_KB_I	100.087	117.347	77.844	10.747	276
Naver 모바일 '전세' 검색 지수	JS_NV_MB	38.261	100.000	1.000	20.552	276
Naver 모바일 '전세대출' 검색 지수	JSL_NV_MB	27.156	100.000	2.000	19.372	276
Naver PC '전세' 검색 지수	JS_NV_PC	28.848	100.000	14.000	14.477	276
Naver PC '전세대출' 검색 지수	JSL_NV_PC	33.388	100.000	20.000	12.843	276
Google '전세' 검색 지수	JS_GG	36.243	100.000	19.000	12.073	276
Google '전세대출' 검색 지수	JSL_GG	39.928	100.000	0.000	16.318	276

44) 나중화·김은섭, 전계논문, pp.95~101. : 김민희, 상계논문, pp.2~13.  
 45) 네이버는 기존에 네이버 트렌드에서 PC와 모바일을 구분하여 검색 지수를 제공하던 것을 2016년부터 네이버 데이터랩에서 PC 검색 통계와 모바일 검색 통계를 합하여 검색 통계를 제공하고 있다.  
 46) 이궁희·황상필, 전계논문, pp.6~7. : 네이버 트렌드  
 47) 김민희, 전계논문, pp.5~8.

검색어 상위 순위에 대출관련 검색어가 위치하고 있는 것은 전세가 월세와 달리 목돈이 들어가므로 전세를 구하는 임차인의 관심사는 자금 마련일 것이다.<sup>48)</sup> 네이버 트렌드에서도 '전세'와 '전세대출'이 타 검색어(임대, 월세, 반전세등) 대비 높은 검색량 추이를 보여, 본 연구에서는 전세가격지수와 상관관계를 실증분석 할 검색어로 '전세'와 '전세대출' 2개의 검색어를 최종 선정하였다.

## 2. 분석모형

본 논문에서 사용되는 모형은 전세가격 ARIMA 모형과 여기에 포털 검색 지수를 포함한 ARIMA 모형이다. 이는 어떤 특정치를 분석할 때 그 특정치와 인과관계에 있는 다른 변수들을 고려하지 않고 그 자신의 현재 및 과거의 값들만을 가지고 그 특성을 파악하고자 하는 방법이다.<sup>49)</sup>

시계열 분석에서는 기본적으로 시계열자료, 즉 연속적인 시간에 대한 한 변수의 관측치에 대한 미래예측을 수행하는데 대표적인 확률적 시계열모형에는 AR(autoregressive model, 자기회귀 확률 과정 모형)과 MA (moving average mode, 자기회귀 이동 평균 모형)이 있으며, 확률과정이 AR과 MA과정을 동시에 지니고 있는 경우, 효과적으로 분석하는 방법이 Box and Jenkins이 고안해낸 ARIMA(autoregressive integrated moving average model, 자기회귀 결합이동평균모형)를 주로 사용한다. Box-Jenkins 방법은 4단계로 이루어져 있다. 1단계 모형의 식별은 산점도, 상관도표와 편상관도표를 사용하여 적절한 p, d, q를 찾는다. 계열이 추세형인지 또는 확률적 도형인지를 구분하고 시계열의 안정성등을 검정한 후 불안정시계열이면 차분하여 안정화시킨다. 2단계 모형의 추정 1단계에서 추정된 모수가 통계적으로 유의한지 여부를 가리는 단계이다. 모수 선택기준(criteria)은 모수절약(parsimony), 안정성(stationarity) 및 AIC와 SBC의 적합도, 가역성(invertibility)

등이며 추정된 모수가 통계적으로 유의하지 않으면 다시 1단계로 회귀한다. 3단계 모형의 진단은 최종적으로 모형을 선정하는 단계이다. 잔차가 k차 후행까지 자기상관이 존재하지 않는다는 귀무가설을 검증하기 위해 Ljung-Box Q-통계량과 p값 검정을 실시한다. 마지막 4단계 예측(forecast)은 최종 선정된 모형을 이용하여 미래를 예측한다.

포털 검색 지수가 전세가격 예측에 의미가 있는지 검증 위해 KB국민은행에서 제공하는 전세가격지수를 활용한 ARIMA 기준모형과 기준모형에 포털 검색 지수 정보를 추가한 ARIMA모형을 활용하여 포털 검색어 검색량의 변화가 전세가격 예측에 미치는 영향과 유의미한 영향을 미치는 시차를 실증적으로 분석한다.

## IV. 분석결과

### 1. 기본적 검정

#### 1) 단위근 검정

시계열 변수가 불안정적일 경우 변수 내에 단위근이 존재하게 되며, 이를 검정하는 방법이 단위근 검정이다. 시계열 자료인 KB국민은행 2010년7월~2015년12월까지 주간 자료인 전세가격지수와 포털 검색 지수의 단위근을 검정하기 위하여 시간추세와 상수가 포함된 ADF검정법을 이용한 단위근 검정을 실시하였다. 자료가 주간단위이므로 계절변동에 대하여는 검정하지 않았다. 기준모형으로 사용할 전세가격지수에 대한 단위근 검정결과 5% 유의수준에서 단위근을 갖는다는 귀무가설을 기각하지 못하므로 전세가격지수 수준변수가 단위근을 가지고 있는 것으로 판단된다. 시계열이 단위근을 갖는다는 것은 시계열이 추세를 내포하고 있어 차분등에 의하여 시계열을 안정성을 확보하여야 한다는 것을 의미

48) 김민희, 상계논문, p.11.

49) 우경·이성석, "지가변동률 예측을 위한 시계열 모형 분석", 부동산학보, 한국부동산학회, 2015, 60집, p.144.



한다. 그래서 전세가격지수를 로그 차분하여 변화율로 전환하고 단위근 없는 시계열로 변경하였다. 50) 네이버와 구글 '전세' '전세대출' 검색 지수는 수준변수에서 모두 1% 유의수준에서 귀무가설을 기각하여 안정성을 갖는다고 판단할 수 있다. 최종적으로 전세가격지수는 로그차분 변수, 포털 검색 지수는 수준변수를 모형의 변수로 선정하였다.

〈표 3〉 단위근 검정결과

변수명	수준변수		로그 1차 차분변수	
	t 값	p-value	t 값	p-value
JS_KB_I	-3.069	0.116	-5.302	0.000
JS_NV_MB	-5.596	0.000	-19.769	0.000
JSL_NV_MB	-6.310	0.000	-20.729	0.000
JS_NV_PC	-4.607	0.001	-18.035	0.000
JSL_NV_PC	-5.413	0.000	-15.037	0.000
JS_GG	-6.668	0.000	-20.357	0.000
JSL_GG	-5.539	0.000	-16.301	0.000

주: 1) 시차결정은 SIC를 기준으로 결정함  
 2) 유의수준 임계치 1%(-3.992), 5%(-3.426), 10%(-3.136)

## 2) 인과관계 검정

전세가격변동률과 포털 검색 지수간의 상호 연관성을 검증하기 위하여 그랜저 인과관계 검정을 실시하였다. 그랜저 인과관계 검정법은 시차분 포모형을 이용하여 원인변수와 결과변수의 구분을 명확하게 위한 방법이다. 전세가격변동률과 포털 검색 지수간의 그랜저 인과관계 검정분석결과 네이버 모바일 '전세' 검색 지수는 전세가격변동률에 시차 5 에서만 1% 유의수준으로 인과관계가 존재하였으나, 다른 시차에서는 인과관계가 존재하지 않았다. 네이버 모바일 '전세대출' 검색 지수는 전 시차에서 전세가격변동률과 상호 인과관계가 존재 하지 않아, 네이버 모바일 검색어는 전세가격변동률과 상호 영향을 미치지 않는 것으로 확인되었다. 네이버 PC, 구글의 '전세' '전세대출' 검색 지수는 1% 유의수준에서 전세가격변동률에 영향을 미쳤 으나, 반대로 전세가격변동률은 네이버 PC, 구글 '전세' '전세대출' 검색 지수에 영향을 미치지 않았다. 이것은 전세시장 참여자들이 전세 계약 전에 PC를 통하여 부동산시장의 정보를 획득하고 일정 기간 후에 실행에 옮기는 것으로 볼 수 있다.

〈표 4〉 그랜저 인과관계 검정결과

Null Hypothesis:	Lags: 2	Lags: 3	Lags: 4	Lags: 5
	p-value	p-value	p-value	p-value
JS_NV_MB → DLOGJS_KB_I	0.063*	0.112	0.025**	0.001***
DLOGJS_KB_I → JS_NV_MB	0.038**	0.060*	0.103	0.132
JSL_NV_MB → DLOGJS_KB_I	0.194	0.261	0.097	0.036**
DLOGJS_KB_I → JSL_NV_MB	0.242	0.234	0.207	0.235
JS_NV_PC → DLOGJS_KB_I	0.000***	0.000***	0.000***	0.000***
DLOGJS_KB_I → JS_NV_PC	0.419	0.216	0.454	0.537
JSL_NV_PC → DLOGJS_KB_I	0.000***	0.000***	0.000***	0.000***
DLOGJS_KB_I → JSL_NV_PC	0.968	0.501	0.156	0.063*
JS_GG → DLOGJS_KB_I	0.000***	0.000***	0.000***	0.000***
DLOGJS_KB_I → JS_GG	0.254	0.391	0.398	0.463
JSL_GG → DLOGJS_KB_I	0.001***	0.000***	0.000***	0.001***
DLOGJS_KB_I → JSL_GG	0.142	0.260	0.086*	0.202

주: \*, \*\*, \*\*\*은 10%, 5%, 1%의 유의수준에서 귀무가설을 기각하므로 변수간 인과관계가 있음을 의미함

50) 금기조·김병량, "KOSPI지수와 금융변수가 주택매매가격과 전세가격에 미치는 영향 분석", 부동산학보, 한국부동산학회, 2015, 60집, pp.188~189.

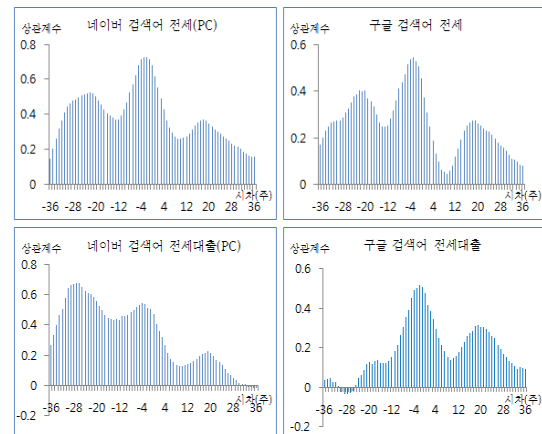
### 3) 교차상관 분석

전세가격변동률과 포털 검색 지수의 상관관계를 분석하기 위하여 ±36주 시차를 가지고 교차상관분석을 실시하였다.<sup>51)</sup> 네이버 PC '전세' 검색 지수는 분석기간 모두 양의 상관관계를 보였으며, -22주(5~6개월)에서 상관계수가 1차 피크를 기록한 이후 다시 -12주부터 상관계수가 상승하여 시차 -3주에 상관관계가 0.726으로 최고점을 기록한 후 급격하게 하락하는 모습을 보였다. 구글 '전세' 검색 지수 또한 네이버 PC '전세' 검색 지수와 상관계수만 다를 뿐 시기별로 유사한 흐름 패턴을 보였으며, 김민희(2014)<sup>52)</sup>도 구글 '전세' 검색량이 변화하면 2~5개월 후에는 전세가격 상승률도 유사한 흐름을 보인다고 하였다. 이것은 주택임대차보호법<sup>53)</sup> 임대인은 임대차계약의 존속여부를 임차인에게 계약종료 6개월전부터 1개월 전 사이에 통보하여야 한다는 법규정에 따라 임대인은 6개월 전부터 인근 주택의 전세 시세, 임대차 시장 동향등을 파악하기 위하여 포털 검색을 시작하고 임차인 역시 임대차기간이 종료되기 전에 전세 시세 파악과 새로운 전세 주택을 검색하므로 검색의 빈도가 높아지는 결과로 해석된다. 특이한 것은 네이버 모바일 '전세' 검색 지수는 분석 전 기간 음의 상관관계를 보이고 있다는 것이다. 해당 분석에 대해 시차를 늘려 확인한 결과 -54주 이전 시차에서는 양의 상관관계를 나타내기는 하였으나, 모바일 검색결과와 PC의 결과가 일치하지 않는다는 것은 검색 접근 경로에 따라 검색하는 성향의 차이와 용도가 다르기 때문일 것으로 추정할 수 있다. 다만,

모바일의 경우 본 분석기간 내 사용자가 급증하여<sup>54)</sup> 검색의 빈도의 증감<sup>55)</sup>이 단순 전세 시장의 흐름 변화와 연관성이 떨어질 수 있다는 점 또한 본 결과를 해석하는 하나의 관점이 될 수 있을 것으로 판단된다.

한편, '전세대출' 검색 지수의 교차상관분석 결과는 '전세'와 유사하였다. 네이버 PC '전세대출' 검색 지수는 -36주 전부터 상관계수가 상승하여 -26주(6~7개월)시점에 상관계수(0.680)가 가장 큰 것으로 확인됐다. 이것은 전세대출이 필요한 임대인 임차인이 계약만료시점 6~7개월 전부터 대출은행, 금리등 대출조건을 사전에 검색하는 결과로 해석된다. 구글 검색어 전세대출도 6~7개월 전 부터 상관관계가 꾸준히 증가하였다.

〈그림 1〉 교차상관 분석 결과



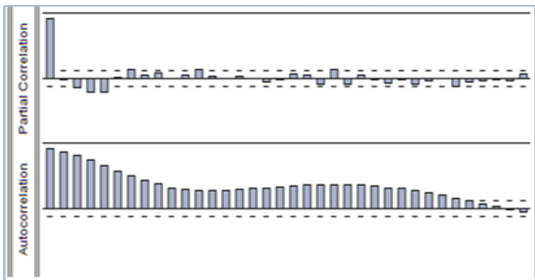
주: 시차가 마이너스(-)인 것은 포털 검색 지수가 전세가격변동률에 선행하는 것임.

51) 이강용·이종아·정준호, "주택시장과 주식시장의 동적 네트워크 구조 비교", 부동산학보, 한국부동산학회, 2015, 61집, p.198.  
 52) 김민희, 전계논문, pp.6~7.  
 53) 주택임대차 보호법 제6조(계약의 갱신) ① 임대인이 임대차기간이 끝나기 6개월 전부터 1개월 전까지의 기간에 임차인에게 갱신거절(更新拒絕)의 통지를 하지 아니하거나 계약조건을 변경하지 아니하면 갱신하지 아니한다는 뜻의 통지를 하지 아니한 경우에는 그 기간이 끝난 때에 전 임대차와 동일한 조건으로 다시 임대차한 것으로 본다. 임차인이 임대차기간이 끝나기 1개월 전까지 통지하지 아니한 경우에도 또한 같다.  
 54) 스마트폰 보유율은 2011(24.2%), 2012(54.0%), 2013(71.9%), 2014(79.5%), 2015(83.2%)으로 급속도로 상승하였으나, 데스크탑 컴퓨터는 2012(66.7%) ⇒ 2015(63.0%) 소폭 하락하고 노트북은 2012(24.4%) ⇒ 2015(26.5%), 테블릿PC는 2012(3.5%) ⇒ 2015(6.0%)로 상승하여 PC의 보유를 변화는 적었음. (2015년 한국미디어패널조사 (정보통신정책연구원) <표 5-1> 연도별 미디어 기기 보유현황)  
 55) 이금희·황상필, 전계논문, p.9.

## 2. 기준모형의 설정

전세가격지수 수준변수의 단위근 검정결과 단위근이 발견되어 전세가격지수를 로그 1차 차분하여 단위근 검정결과 1% 유의수준에서 시계열이 안정화된 것으로 판명되었다.

〈그림 2〉 전세가격변동률 계열상관도



기준모형의 AR, MA차수를 결정하기 위하여 전세가격변동률의 계열상관도를 진단한 결과 자기상관함수(ACF: autocorrelation function)는 지수적으로 감소하고, 편자기상관함수(PACF: partial autocorrelation function)는 시차1에서 스파이크가 신뢰 한계선 밖으로 벗어나, 가능성 있는 모형<sup>56)</sup>을 ARIMA(1, 1, 2), ARIMA(1, 1, 3), ARIMA(1, 1, 4), ARIMA(1, 1, 5) 4가지를 선정하였다.

〈표 5〉 ARIMA 모형 식별통계량

구 분	ARIMA (1, 1, 2)	ARIMA (1, 1, 3)	ARIMA (1, 1, 4)	ARIMA (1, 1, 5)
adj R <sup>2</sup>	0.9097	0.9156	0.9191	0.9188
AIC	-13.0576	-13.1202	-13.1578	-13.1507
SC	-12.9918	-13.0413	-13.0658	-13.0454
HQC	-13.0312	-13.0885	-13.1209	-13.1084

ARIMA 모형의 선정은 adj R<sup>2</sup>인 가장 높으며, AIC(Akaike information criterion), SC(Schwarz criterion), HQC(Hannan-Quinn criterion)값이 작을수록 모형의 적합도가 높은 것으로 추정되므로 ARIMA(1, 1, 4)를

기준모형으로 선정하였다.

〈표 6〉 ARIMA(1, 1, 4) 모형 추정치

구 분	계수	t 값	adj R <sup>2</sup>	F 값	DW
C	0.0014*	3.1121	0.9191	519.6883*	1.9991
AR(1)	0.8998*	41.6710			
MA(1)	-0.0148	-0.4217			
MA(2)	0.1996*	3.3350			
MA(3)	0.2425*	4.4899			
MA(4)	0.2302*	5.2803			

주: \* 유의수준 1%

기준모형의 적합성을 진단하기 위하여 Ljung-Box Q-통계량값을 검정하였다. 검정 결과 Q통계량이 chi-square( $\chi^2$ )보다 작아 귀무가설인 ACF는 백색잡음 ACF와 유의적인 차이가 없다(ACF=0)를 기각하지 않았다. 그리고 자기상관함수와 편자기상관함수가 모든 시차에서 신뢰한계선 내에 존재하는 형태를 보였다. 따라서 선정된 모형의 잔차가 백색잡음으로 모형의 타당성이 검정되어 최종 모형으로 선정하였다.

〈그림 3〉 ARIMA(1, 1, 4) 잔차 독립성 검정

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	-0.000	-0.000	7 E-05		
2	0.000	0.000	7 E-05		
3	0.027	0.027	0.2097		
4	0.035	0.035	0.5555		
5	0.021	0.021	0.6745		
6	-0.022	-0.023	0.8103	0.368	
7	0.077	0.075	2.4742	0.290	
8	-0.035	-0.038	2.8308	0.418	
9	-0.023	-0.024	2.9875	0.560	
10	-0.090	-0.094	5.3357	0.376	
11	-0.112	-0.116	8.9264	0.178	
12	-0.040	-0.043	9.3845	0.226	
13	-0.012	-0.002	9.4233	0.308	
14	-0.026	-0.020	9.6133	0.383	
15	-0.008	0.011	9.6323	0.473	
16	0.064	0.073	10.821	0.458	
17	0.030	0.047	11.091	0.521	
18	-0.046	-0.032	11.706	0.552	
19	0.058	0.052	12.710	0.549	
20	0.164	0.148	20.726	0.146	
21	-0.096	-0.121	23.483	0.101	
22	0.122	0.106	27.934	0.046	
23	-0.041	-0.079	28.447	0.056	
24	0.053	0.037	29.285	0.062	
25	0.068	0.074	30.714	0.059	
26	0.047	0.056	31.388	0.067	
27	0.075	0.074	33.096	0.061	
28	-0.020	0.018	33.221	0.077	
29	0.040	0.021	33.714	0.090	
30	0.074	0.131	35.406	0.081	
31	0.029	0.040	35.670	0.098	
32	0.009	0.008	35.697	0.122	
33	0.041	0.051	36.230	0.137	
34	0.042	0.042	36.795	0.152	
35	-0.109	-0.104	40.559	0.094	
36	-0.080	-0.091	42.599	0.080	

56) ARIMA적정모형을 결정하기 위한 차수는 최대 시차5로 제한함.

### 3. 유용성 분석

포털 검색 지수가 인과관계 분석을 통하여 전세가격변동률에 영향을 미치는 것으로 분석되었고, 교차상관분석을 통하여 일정한 시차를 두고 검색 지수 상관계수가 높아지는 것을 알 수 있었다. 따라서 포털 검색 지수가 전세가격변동률의 예측에 유용하게 이용될 수 있다고 판단된다. 포털 검색 지수가 포함된 모형과 전세가격변동률 기준모형과의 예측력 비교를 위하여 실제값과 예측치 차이 절대값의 평균인 MAE를 사용하여 모형간의 예측력의 정확도를 검증하였다. MAE의 값이 작을수록 예측에 대한 정확도가 높은 것을 의미한다.

예측력 향상 정도를 분석하기 위하여 기준모형(Model(0))에 인과관계와 교차상관 분석에서 양의 상관관계를 보인 4개의 검색 지수를 포함한 모형을 추가하여 비교분석 하고자 한다.

〈표 7〉 기준모형과 검색 지수 포함모형

Model명	검색 지수	모형
Model(0)		$\Delta \log(y_t) = \beta_0 + \beta_1 \Delta \log(y_{t-1}) + e_t - \theta_0 - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3 e_{t-3} - \theta_4 e_{t-4}$
Model(1)	$w_t^1$ (네이버PC '전세')	Model(0) + $\beta_{23}(w_t^1)$
Model(2)	$w_t^2$ (네이버PC '전세대출')	Model(0) + $\beta_{24}(w_t^2)$
Model(3)	$w_t^3$ (구글 '전세')	Model(0) + $\beta_{25}(w_t^3)$
Model(4)	$w_t^4$ (구글 '전세대출')	Model(0) + $\beta_{26}(w_t^4)$

주 : 1) 시차(lag)는 인터넷 검색어 지수의 선행 최대 시차 32주 중 MAE가 최소가 되는 시차임.  
 2) 향상률은 기준모형(Model(0))대비 검색 지수 포함모형(Model1~4)의 adj R<sup>2</sup>과 MAE의 향상 정도

〈표 8〉 기준모형 대비 예측력 비교

Model	시차(lag)	AIC	adj R <sup>2</sup>	adj R <sup>2</sup> 향상률	MAE	MAE 향상률
Model(0)	0	-13.15782	0.919082	-	0.000831	-
Model(1)	0	-13.22617	0.924602	0.601%	0.000707	14.922%
	29	-13.24899	0.925315	0.678%	0.000684	17.690%
Model(2)	0	-13.23559	0.925376	0.685%	0.000763	8.183%
	30	-13.41597	0.935513	1.788%	0.000639	23.105%
Model(3)	0	-13.16186	0.919665	0.063%	0.000805	3.129%
	27	-13.25663	0.928684	1.045%	0.000701	15.644%
Model(4)	0	-13.15420	0.919058	-0.003%	0.000819	1.444%
	29	-13.24682	0.925160	0.661%	0.000697	16.125%

기준모형에 현시점(t=0)의 검색 지수를 포함한 모형의 예측력을 비교한 결과 구글 '전세' 검색 지수를 포함한 Model(3)은 3.129%, '전세대출' 검색 지수를 포함한 Model(4)는 1.444%의 예측력 향상을 보였으나, 네이버 PC '전세' 검색 지수를 포함한 Model(1)은 14.922%, '전세대출' 검색 지수를 포함한 Model(2)가 8.183%의 예측력 향상을 보여, 1개월 이내의 단기 예측에는 네이버 PC 검색어 '전세'와 '전세대출'이 유용하다는 것이 확인되었다. 최대 검색 시차를 32주로 하여 비교 결과, 구글 '전세' 검색 지수가 포함된 Model(3)은 시차 27주에서 15.644%, '전세대출' 검색 지수가 포함된 Model(4)는 시차 29주에서 16.125%의 예측력 향상을 보였다. 네이버 PC '전세' 검색 지수를 포함한 Model(1)은 시차 29주에서 17.690%, '전세대출' 검색 지수를 포함한 Model(2)가 시차 30주에서 4개의 모형중 최대 예측력 향상률인 23.105%를 보였다. 이것은 전세 시장참여자들이 전세 계약 실행 6~7개월 전 부터 사전에 전세 관련 정보를 검색하는 결과로 추정되어, 6개월 이상의 전세시장 동향 예측 시 전세 관련 포털 검색어들이 효과가 있음이 실증분석 되었다.

Choi and Varian(2009, 2011)<sup>57)</sup>은 기준모형 대비 구글 검색 지수가 포함된 모형이 실업률은 6.0%, 주택판매는 12.0%, 자동차판매는 18.4% MAE가 향상되는 것을 확인했으며,

57) Choi and Varian, 2009, *op.cit.*, pp.6~14. ; Choi and Varian, 2011, *op.cit.*, p.4.

이금희·황상필(2014)<sup>58)</sup>은 GDP성장률과 네이버 검색 지수를 포함한 모형의 MAE 비교 결과 17.263%의 예측력 향상을 보였다. Tuhkuri(2015)<sup>59)</sup>는 실업률 모형과 구글 검색 지수를 포함한 모형의 MAPE(mean absolute percentage error)<sup>60)</sup>의 비교 결과 전체기간은 1개월전의 구글 검색 지수를 사용할 때 7.48%, 경제위기 기간은 현시차의 구글 검색 지수는 17.95%, 1개월 전의 트렌드 지수를 사용할 때는 34.50%의 예측력이 향상됨을 확인하였다.

본 연구도 6~7개월의 시차에서 기준모형대비 예측력이 15.6~23.1% 향상됨을 보여 선행연구와 유사한 결과 보였다고 할 수 있다.

## V. 결론 및 시사점

본 연구는 ARIMA모형을 이용하여, 2010년 7월부터 2015년 12월까지의 네이버와 구글 '전세' '전세대출' 검색 지수가 전세가격 예측에 효과가 있는지 여부와 효과가 있다면 어느 시차에서 가장 유용성이 높은지 실증분석 하였다. 결과적으로 포털 검색 지수가 경제지표로서의 유용성이 있는지의 여부를 검증한 선행연구와 유사하게 포털 검색 지수를 활용할 경우 전세시장 예측 정확도 향상에 유의미한 결과를 나타내고 있음이 검증되었다. 그 외에도 검색어 검색 매체와 선택 검색어에 따라 예측 정확도가 상이하다는 것을 확인하였다.

실증분석결과는 다음과 같다. 첫째, 전세가격변동률과 포털 검색 지수의 교차상관분석결과 네이버 PC '전세' 검색 지수가 시차 3주시 최대의 상관계수 0.7259를 나타냈으며, 네이버 PC '전세대출' 검색 지수는 시차 26주시 상관계수

0.6798으로 분석되어 단기 전세가격변동률 예측에는 네이버 PC 검색어 '전세'가, 6개월 이상의 전세가격변동률 예측에는 네이버 PC 검색어 '전세대출'이 도움이 될 것으로 분석되었다. 둘째, 인과관계분석에서는 네이버 PC와 구글의 '전세' '전세대출' 검색 지수가 전세가격변동률에 시차2~5 전체에서 영향을 미치는 것으로 나타나, 교차상관분석시 포털 검색 지수가 전세가격변동률에 선행한 것이 실증적으로 증명 되었다. 이는 김민희(2014)<sup>61)</sup>의 전세 검색량이 전세가격 상승률에 대해 선행성을 가진다는 분석과 같은 결과이다. 셋째, ARIMA 기준모형과 포털 검색 지수를 추가한 모형의 예측 정확도를 비교하기 위하여 MAE를 분석한 결과, 현시점(t=0)에서는 대체로 예측력이 향상된 결과를 보였다. 그중에서 네이버 PC '전세' 검색 지수가 가장 높은 예측력 향상을 가져왔다. 시점의 차이를 두었을 때는 모든 검색어가 대체로 6~7개월의 시차에서 MAE가 최소값을 보였다. 이는 전세시장 참여자들이 계약만료 또는 신규 전세계약을 실행에 옮기기 6~7개월전부터 전세물건, 전세 시장 현황, 전세대출등의 정보를 인터넷을 통하여 사전에 검색하는 것으로 해석할 수 있다.

본 연구는 포털 검색 지수를 활용한다면 전세시장 정책입안과 향후 전세가격의 변동에 관심이 많은 전세시장 참여자들에게 보다 정확한 전세가격 예측 결과를 제공할 수 있는 근거를 마련하였다는데 의의가 있다. 국내 포털 시장 점유율 70%대 이상을 차지하는 네이버와 4%대인 구글의 예측력 향상 검증에서 점유율에서는 큰 차이가 있었지만 유사한 예측력 향상을 보였다. 이것은 활용 매체와 상관없이 전세시장 참여자들의 정보검색 시기와 실제 실행에 옮기는 행동양식에 유사성이 있음이 확인되었다.

끝으로 본 연구에 사용된 포털 검색 지수는

58) 이금희·황상필, 전계논문, pp.30~33.

59) Tuhkuri, *op.cit.*, pp.48~54.

60)  $MAPE = \frac{100}{T} \sum_{t=1}^T |Y_t - \hat{Y}_t|$ ,  $Y_t$ 는 실제값,  $\hat{Y}_t$ 는 예측값이며 %가 낮을수록 예측력이 좋다

61) 김민희, 전계논문, pp.6~7.

지역에 따른 검색량의 차이를 구분하여 제공하고 있지 않다. 그러나 실제 전세시장은 지역에 따라 전세의 가격 변동이 다소 상이할 때가 존재한다. 이에 따라 보다 더 세밀한 예측을 위해서는 지역

에 따른 포털 검색 지수와 지역별 전세가격 지수의 비교가 필요할 것으로 판단된다. 이에 대한 연구는 향후 과제로 남긴다.

---

 參考文獻

- 권치명 · 황성원 · 정재운, "인터넷 검색어를 활용한 유행성 독감발생 감지", 한국시물레이션학회 논문지, 한국시물레이션학회, 2014, 제23권 4호
- 권치명 · 황성원 · 정재운, "실업률 예측을 위한 인터넷 검색 정보의 활용", 한국시물레이션학회 논문지, 한국시물레이션학회, 2015, 제24권 2호
- 금기조 · 김병량, "KOSPI지수와 금융변수가 주택매매가격과 전세가격에 미치는 영향 분석", 부동산학보, 한국부동산학회, 2015, 60집
- 김대원 · 유정석, "인터넷 검색 활동과 주택 가격 및 거래량 간 동적 관계 분석", 부동산연구, 한국부동산연구원, 2014, 제24집 2호
- 김민희, "검색데이터 주택시장의 단기예측에 유용하다" *LG Business Insight*, LGERI, 2014
- 김민희 · 김나경, "검색데이터를 보면 소비트렌드가 보인다" *LG Business Insight*, LGERI, 2013
- 김현동 · 이상엽, "백화점형 할인점 개발을 위한 빅데이터 기반 소비자 행태 및 입지특성 분석에 관한 연구", 부동산학보, 한국부동산학회, 2015, 63집
- 나중화 · 김은섭, "소셜미디어 정보를 이용한 실업률 예측", 한국산업정보학회지, 한국산업정보학회, 2013, 제18권 6호
- 이강용 · 이종아 · 정준호, "주택시장과 주식시장의 동적 네트워크 구조 비교", 부동산학보, 한국부동산학회, 2015, 61집
- 우경 · 이성석, "지가변동률 예측을 위한 시계열 모형 분석", 부동산학보, 한국부동산학회, 2015, 60집
- 이금희 · 황상필, "빅데이터를 이용한 경기판단지표 개발: 네이버 검색 경기지수 작성과 유용성 검토", 경제분석, 한국은행 경제연구원, 2014, 제20권 4호
- 이운홍, "전국 아파트 전세변동률 기준 시장 유형화와 유형별 전세가격 변동률 영향요인 분석", 부동산학보, 한국부동산학회, 2016, 64집
- 정지선, "新 가치창출엔진 빅데이터의 새로운 가능성과 대응전략", 새로운 미래를 여는 빅데이터 시대, 한국정보화진흥원, 2013
- Ettredge, M., Gerdes, J., Karuga, G., "Using web-based search data to predict macroeconomic statistics" *Communications of the ACM*, 2005, Vol.48, No.11
- Han Shichun, Wang Ping, Wang Bian, "A study on correlation between web search data and housing price evidence from Beijing and Xi'an in China", *ICSSR*, 2013
- Hyunyoung Choi, Hal Varian, "Predicting the Present with Google Trends", ©Google Inc, 2009
- Hyunyoung Choi, Hal Varian, "Predicting the Present with Google Trends", ©Google Inc, 2011
- Ilaria Bordino, Stefano Battiston, 외 3, "Web search queries can predict stock market volumes", *Plos one*, 2012, Vol.7, Issue7, e40014
- Joonas Tuhkuri, "Big Data : Do Google Searches Predict Unemployment?", *University of Helsinki Faculty of Social Sciences, Economics*, 2015
- Lynn Wu, Erik Brynjolfsson, "The Future of Prediction : How Google Searches Foreshadow Housing Prices and Sales", *Technical report, MIT*, 2009
- Pedro Latoeiro, Sofia B. Ramos, Helena Veiga, "Predictability of stock market activity using Google search queries", *Universidad Carlos III de Madrid Working paper*, 2013, Vol.13, No.06
- Sharad Goel, Jake M. Hofman 외 3, "Predicting consumer behavior with Web search", *PNAS*, 2010, Vol.107, No.41
- Tobias Preis, Daniel Reith, H. Eugene Stanley, "Complex dynamics of our economic life on different scales : insights from search engine query data", *The Royal Society*, 2010
- 네이버 트렌드 <http://trend.naver.com>.
- 구글 트렌드 <http://www.google.co.kr/trends>.
- KB국민은행 홈페이지 <http://www.kbstar.com>.