

뉴스 기사가 아파트 매매·전세 가격 국면 지속기간에 미치는 영향 분석*

김대원** · 유정석***

How does News Article Affect on the Apartment Sale and Chonsej Price Regime Duration*

Dai-Won Kim** · Jung-Suk Yu***

요약 : 이 연구는 전국과 서울시를 대상으로 뉴스 기사 속 어조와 아파트 매매·전세 가격 국면 지속기간 간 관계를 확인하기 위한 실증 분석을 수행하였다. 분석을 위해 ‘주택 가격’ 또는 ‘전세 가격’으로 검색된 뉴스 기사 제목에 나타난 ‘상승’ 또는 ‘하락’ 단어의 빈도수와 아파트 매매가와 전세가 지수를 분석 자료로 구성하고, Cox 비례위험 모형으로 양자 간의 지속 의존성 관계를 분석하였다. 분석 결과, 주택 가격과 관련된 뉴스 제목 속 어조와 아파트 매매 가격 국면 지속기간 간에는 통계적으로 유의미한 관계를 포착할 수 없었으나, 전세 가격과 관련된 뉴스 기사 제목 속 어조는 전국과 서울시 모두에서 전세 가격 상승 또는 하락 국면의 위험률을 낮추는 역할을 하는 것으로 확인되었다. 이 같은 결과는 뉴스 기사 속 어조에 따라 주택 전세 가격 상승 또는 하락 국면의 지속기간이 영향을 받고 있음을 의미하는 결과로, 향후 전세 시장 가격과 동향 예측에 있어 뉴스 기사가 활용될 수 있음을 시사한다.

주제어 : 뉴스 기사, 아파트 매매가 국면, 아파트 전세가 국면, 지속 의존성 분석

ABSTRACT : In order to identify the relationship between the tone of news article and apartment sale price and Chonsej price regime duration, an empirical analysis was implemented on all the country of Korea and Seoul. We set up the analysis data with the frequency of the word ‘upturn’ or ‘downturn’ in news article headlines and the apartment sale and Chonsej price indices, then we estimated duration dependence analysis on both variables using Cox proportional hazard model. The analysis result showed that the tone of news article headlines decreased hazard ratio of Chonsej price regime duration in both all the country of Korea and Seoul, although there’s no significant relationship between the tone of news articles and housing price regime duration. The results of this study provide the empirical evidence that Chonsej price can be influenced by the tone of news articles. Hence, it could be possible to use news articles to predict the Chonsej price and its trend in the near future.

Key Words : news article, apartment sale regime cycle, apartment Chonsej regime cycle, duration dependence analysis

* 이 논문은 2015년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2015S1A3A2046715). 그리고 제1저자의 박사학위논문 내용 중 일부를 수정·보완하여 작성되었습니다.

** 단국대학교 부동산학 박사(Ph.D. in Urban Planning and Real Estate, Dankook University)

*** 단국대학교 도시계획·부동산학부 부교수(Associate Professor, School of Urban Planning and Real Estate, Dankook University),
교신저자(E-mail: jsyu@dankook.ac.kr, Tel: 031-8005-3338)

I. 서론

주택 가격은 다른 자산과 마찬가지로 상승과 하락의 국면을 번갈아 나타내며 일정 기간 동안 지속되는 양상을 나타낸다(Bracke, 2013). 이러한 주택 가격 국면은 때로 오랫동안 지속되기도 하지만, 어떤 때에는 짧게 유지되다가 다른 국면으로 전환되어 다양한 사회 현상을 수반하기도 한다(김대원·유정석, 2013).

특히, 주택 시장의 가격 국면 변화는 다른 시장보다 사회·경제적으로 더욱 높은 여파를 미치게 되는데, 이는 주택 가격이 다른 재화에 비해 월등히 높은 가격에 거래되기 때문이다. 실제로, 1970년대 이후 산업화를 거치는 동안 국내 주택 가격은 지속적인 상승세를 기록하면서, 하락보다는 상승 국면이 주도적인 양상을 나타냈다. 주택 가격의 상승은 기존 주택 보유자들에게 경제적 부를 안겨 주면서 국내 경제 성장의 상징적인 지표로 받아들여지기도 하였으나, 한편으로는 높은 주택 가격으로 인한 서민 주거 안정의 저해, 물가 상승, 불로소득으로 인한 빈부 격차의 증가 등 다양한 사회 문제의 원인으로 인식되기도 하였다.

이와는 반대로, 2010년까지 지속적인 상승 국면을 기록하던 국내 주택 가격은 2011년 무렵을 기점으로 상승세를 멈추고 보합세 또는 하락세를 유지하면서 또 다른 사회·경제적 문제를 불러오기도 하였다. 주택 가격 하락 국면의 지속으로 인한 주택 담보 가치의 하락, 이로 인한 가계 대출의 불안정, 미분양 주택의 증가, 주택시장 참여자들의 투자 심리 위축, 전세가 폭등 등 주택 가격 상승세에 나타났던 문제와는 또 다른 문제가 나타나기도 하였으나, 이는 결국 서민 주거 불안정이라는 현상으로 귀결된다는 점에서 상승 시기와

공통적인 특징을 보이기도 하였다.

어떠한 경우든 주택 가격 국면의 지속은 주택 가격 자체뿐만 아니라 금융 시장, 건설 시장, 가계 및 주거 안정 등 다양한 사회·경제 현상에 중대한 영향을 미치는 요소 중 하나인 것만은 분명한 것으로 보인다.

주택 가격 국면이 갖는 중요성은 주택 시장과 관련 산업의 시장 흐름을 좌우한다는 데 있다. 김대원·유정석(2013)의 연구에서 분석된 바와 같이, 일단 주택 가격 국면이 탄력을 받아 일정기간 유지되면 그 국면을 지속시키려는 경향을 강하게 나타내기 때문이다. 이에 대하여 김대원·유정석(2013)은 주택 가격 형성 초기의 가격 변화량이 시장 참여자들의 심리적 태도에 영향을 미치기 때문으로 해석하여, 국내 주택 시장에서 시장 참여자들의 심리적 결정 요인으로 초기 가격 변화량의 중요성을 강조한 것이다.

이 때, 주택 시장 참여자들의 심리적 태도는 초기 가격의 변동량뿐만 아니라 다양한 요소로부터 영향을 받게 된다. 일례로, 주변 사람들의 의견이나 경험담, 전문가 그룹의 컨설팅, 온라인 공간상의 정보, 사업시행자가 제시하는 홍보물 등 수많은 정보가 시장 참여자들의 심리적 태도에 영향을 미치게 된다. 하지만, 현대사회에서 일반 대중에게 가장 커다란 영향력을 미치는 매체로는 언론, 특히 뉴스 기사를 들 수 있다. 이는 뉴스 기사는 누구나 쉽게 접근할 수 있다는 점과 규칙적·지속적으로 공급되는 정보라는 특징으로 일반 대중에게 많은 영향을 미치는 매체이기 때문이다. 게다가 최근 모바일 인터넷이 급속하게 발전·보급되고 있는 국내 상황에서 뉴스 기사는 더욱 빠르고 널리 유통되고 있다는 점도 주목해야 할 점이다.

이러한 맥락에서 시장 참여자들의 심리적 태도가 시장 분위기 형성과 주택 가격 국면의 지속기간을 결정짓는다고 할 때, 보편적 대중 매체인 뉴스 기사도 시장 참여자들의 심리적 태도에 영향을 미치게 되고, 이로 인해 결국 주택 가격 국면 지속기간에도 영향을 미치게 될 것이라고 가정할 수 있다. 이 연구는 주택 가격과 관련된 뉴스 기사 속 어조가 아파트 매매·전세 가격 국면 지속기간에 미치는 영향에 대하여 지속 의존성 분석을 실시하여, 뉴스 기사와 주택 가격 국면의 지속기간 간 관계에 대한 실증적 분석을 시도하였다. 만일 지속 의존성 분석 결과가 뉴스 기사 속 어조에 의해 주택 가격 국면 지속 의존성이 민감도를 갖는 것으로 확인된다면, 이는 뉴스 기사의 내용에 따라 주택 가격 지속기간이 변화할 수 있음을 의미하며 뉴스 기사에 내포된 어조에 의해 주택 가격의 변화가 유도됨을 의미하게 될 것이다. 이와 더불어 분석 결과는 뉴스 기사에 따른 주택 가격 변화의 방향성도 제시할 수 있을 것이다.

논문은 다음과 같이 구성된다. II 장에서는 뉴스 기사와 주택 시장 간의 관계에 대한 선행연구를 검토하고, 이를 통해 이 연구의 차별성에 대하여 제시하였다. 이후 III 장에서는 이 연구에서의 실증 분석 방법을 소개하고, 분석 자료 및 이에 대한 기초 통계를 제시하였다. IV 장에서는 실증 분석 결과와 이에 대한 해석을 제시하여 뉴스 기사와 아파트 매매·전세 가격 국면 지속기간 간의 관계에 대하여 고찰하였다. 마지막 V 장에서는 실증 분석 결과를 요약하고, 이 연구 결과가 갖는 의미와 활용 방향에 대한 기대를 제시하였다.

II. 선행연구 검토 및 이 연구의 차별성

언론 또는 뉴스 기사와 부동산 시장 간 관계에 대한 실증 연구는 양자 간의 높은 관련성이 쉽게 예측됨에도 불구하고, 국내의를 막론하여 그다지 많이 다루어지지 않은 편이다.

국내에서 발표된 논문 중 언론과 부동산 시장 간 관계에 대한 연구는 3건으로 확인되며, 이 중 국내 출간 논문임에도 불구하고 미국 부동산 시장에 대해 분석한 진창하·Gallimore(2012)의 연구를 제외하면, 2건의 연구가 언론과 부동산 시장 간의 관계를 분석한 순수 국내 연구로 확인된다.

먼저, 김진유(2006)의 연구는 국내 연구 중 최초로 언론과 부동산 시장 간 관계에 대하여 실증적으로 분석한 연구로 확인된다. 이 연구는 조선, 동아, 매일, 한경 등 주류 언론사를 대상으로 2001년부터 2005년까지 총 75,620건의 뉴스 기사를 수집, 이 중 '투기'라는 단어가 포함된 기사들을 분류하여 기사 내용 중 상승 또는 하락이라는 어조가 포함된 기사를 구분하고 그 수를 산정하여 주택 가격과의 인과적 방향성을 확인하였다. 분석 결과, 주택 가격의 변화는 상승 기사 수에 인과적인 영향력을 미치는 것으로 나타났으나, 하락 기사 수에는 인과적 영향력이 미치지 않는 것으로 확인되었다. 한편, 뉴스 기사 수도 주택 가격 변화에 별다른 인과성을 나타내지는 않는 것으로 확인되었다. 이에, 주택가격의 변동이 주택 가격 상승과 관련된 뉴스 기사 수에 영향을 미친다는 점 이외의 주택 가격과 뉴스 기사와의 인과적 영향은 뚜렷하지 않음을 이 연구 결과는 제시하고 있다.

우윤석·이은정(2011)은 김진유(2006)의 연구와 다소 다른 관점을 통해 뉴스 기사와 주택 가격 간의 관계에 대한 실증 분석을 실시하였는데, 이

들은 언론 보도의 빈도수가 주택 가격 변화에 영향을 미칠 것이라는 관점에서 실증 분석을 시도하였다. 이들은 한국언론재단의 KINDS 서비스를 활용해 2003년 2월부터 2008년 7월까지 서울 강남 지역 아파트와 관련된 2,757건의 뉴스 기사를 수집, 강남, 강북, 수도권, 광역시 아파트 가격에 미치는 인과적 영향력 및 방향성에 대하여 실증 분석하였다. 분석 결과, 강남 지역의 아파트 가격 상승에 대한 기사 빈도수의 증가가 타지역 주택 가격 상승에 영향을 미치고 있음을 확인할 수 있었다.

해외 연구에서도 뉴스 기사와 부동산 시장 간의 관계를 다룬 연구는 제한적이다. 이 중 Gayer and Viscusi(2002)는 미국 미시건주 Grand Rapids 지역을 대상으로 1988년 1월부터 1993년 12월까지 72개월간 3,702건의 주택 매매거래를 대상으로 뉴스 기사와의 관련성을 연구하였다. 이들의 연구는 위험 폐기물 지역에 대한 뉴스 기사가 주택 가격에 미치는 영향을 실증 분석하였는데, 뉴스 기사가 1만كم 증가할 때마다 약 100~200달러 정도 주택 가격이 상승하는 현상을 발견하였다. 이에 대하여 Gayer and Viscusi(2002)는 매스컴의 관심이 주민들의 위험 인지 정도를 낮추거나 위험 요소 제거에 대한 기대를 상승시키기 때문으로 해석하였다.

상기와 다소 다른 관점에서 Tsftati and Cohen(2003)은 제3자 효과(Third-Person Effect)¹⁾를 이론적 근거로 주변 지역에 대한 미디어 보도가 주민들의 이사 욕구(moving desire)에 미치는 영향을 검정하였다. 이들은 이스라엘 지역 472개 표본

자료를 사용하여 구조방정식 모형을 설정, 제3자 효과가 이사에 대한 인식과 행동에 영향을 미치는지에 대하여 실증 분석을 실시하였다. 분석 결과, 사람들은 뉴스 기사가 다른 사람들에게 영향을 미치는 것으로 인지하면 이사 여부를 고려한다는 점을 밝혔다. 이러한 현상에 대하여 Tsftati and Cohen(2003)은 주변 개발 지역에 대한 낙인화(stigmatizaion)가 거주민들의 거주 만족도에 영향을 미치게 되어 이사 여부를 결정하는 것으로 해석하였다.

또한 진창하·Gallimore(2012)는 미국 애틀랜타주 CMSA(Consolidated Metropolitan Statistical Areas)를 대상으로 뉴스 기사와 주택 가격 간 인과성을 실증 분석하였다. 이들은 2002년 12월부터 2008년 10월까지 Atlanta Journal-Constitution 자료를 대상으로 real estate, real estate price, home price를 키워드로 605건의 뉴스 기사 자료를 수집, 주택 가격 간의 관계에 대한 Granger 인과성을 검정하였다. 분석 결과, 주택 가격의 변화는 뉴스 기사 속 긍정적인 단어의 사용보다는 부정적인 단어의 사용이 영향을 미치는 것으로 확인되었다.

한편, Mercille(2014)은 뉴스 미디어의 역할로 인하여 아일랜드 주택 가격 버블이 지속되었음을 정치경제학적 관점에서 분석하였다. 이 연구 결과, 2007년 아일랜드의 주택 가격 버블이 붕괴 직전까지 지속되었던 이유를 언론사들이 정부와 기업의 입장을 반영하여 뉴스 기사를 작성하였기 때문으로 지적하였다. 이는 뉴스 조직이 엘리트 집단으로 신자유주의적 사고를 갖고 있다는 점, 정치 집단과 기업 집단과 다중의 연결 고리

1) 제3자 효과(Third-Person Effect)는 W. Phillip Davison에 의해 제시된 인간의 편향된 인지적 특성 효과를 의미한다. Davison(1983)에 따르면, 매스 미디어에 대하여 사람들은 자신보다는 타인들이 더 영향을 많이 받는다고 생각하는 인지적 편향성을 갖는데, 이를 제3자 효과라 지칭하였다. 즉 본인 자신은 매스 미디어에 의해 영향을 적게 받을 것이라고 과신하지만, 제3자인 누군가는 매스 미디어에 의해 영향을 많이 받게 된다고 믿는 인지적 오류를 의미한다.

를 유지하였다는 점, 부동산 개발 회사 등 기업들의 광고 수주에 대한 부담감을 갖고 있다는 점, 사건의 정리에 있어 엘리트 집단 출신의 소위 전문가에 전적으로 의존하고 있다는 점 등에서 관련성이 유추된다. 즉, 주택 가격 버블이 붕괴되기 직전까지 버블 붕괴에 대한 조짐이 실제로 나타났더라도, 뉴스 미디어들은 정치·경제계와의 관계로 인하여 이에 대한 언급을 하지 않았기에 버블 상태가 끝까지 지속된 것으로 동 연구는 밝혔다.

상기에서 살펴본 바와 같이, 비록 관련 연구의 수가 많은 편은 아니지만 뉴스 기사와 부동산 시장이 직·간접적으로 크고 작은 관련성을 맺고 있다는 사실은 자명해 보인다. 이는 일반 대중에게 새로운 소식을 전하여 여론 형성에 기여하는 뉴스 기사의 특성이 부동산 시장에서도 동일한 원리로 작동, 부동산 시장 동향 변화에 크고 작은 영향을 미치기 때문일 것이다.

이 연구는 뉴스 기사 속 어조가 주택 시장 분위기 형성에 중요한 요인으로 작용할 것으로 상정, 기존 선행 연구의 범위를 보다 확장하여 뉴스 기사 속에 내포된 주택 시장에 대한 어조와 주택 가격 국면 간의 관계를 파악하고자 하였다. 이를 위해 뉴스 기사 속 어조가 아파트 매매 및 전세 가격 국면 지속 기간에 미치는 영향을 실증 분석하여, 기존 연구와의 차별성을 도모하였다. 이와 더불어 이 연구는 인터넷 텍스트 마이닝 기법²⁾을 활용하여, 기존 선행 연구에 비해 장기간의 뉴스 기사를 수집·분석할 수 있었고 기존 연구 대비 긴 시계열 자료를 사용하였다는 점에서도 차별성이 있다.

Ⅲ. 연구 방법과 분석 자료

1. 지속의존성 분석

지속 의존성 분석(duration dependence analysis)은 특정 이벤트가 발생할 때까지의 시간에 대한 분석 기법이다. 지속 의존성 분석은 다양한 분야에서 활용되면서 명칭도 다양하게 불리고 있는데, 의학 분야에서는 생존 분석(survival analysis)으로 불리고 있으며, 공학 분야에서는 신뢰도 분석(reliability analysis), 사회학에서는 사건사 분석(event history analysis), 경제학 분야에서는 지속 의존성 분석으로 불리고 있다. 다양한 명칭에도 불구하고 이 분석 기법은 특정 이벤트의 발생 시점까지의 기간을 측정하는 통계적 방법의 하나로, 중도절단 자료(censored data)와 절단 자료(truncated data)를 포함하여 분석할 수 있다는 공통점을 갖고 있다(송경일·최종수, 2011; 김대원·유정석, 2013).

지속 의존성 분석의 유형은 크게 모수적(parametric) 모형, 비모수적(non-parametric) 모형, 반모수적(semi-parametric) 모형의 3가지로 분류된다. 모수적 모형은 지속 의존성에서의 생존 함수(survival function)가 특정 분포 형태를 갖는다고 가정하는 모형으로, 지수(exponential) 모형, Weibull 모형, Gompertz 모형, Lognormal 모형, 그리고 Loglogistic 모형 등이 여기에 속한다. 모수적 모형은 분석 대상 자료의 생존 함수가 상기 나열된 분포 형태 중 어떠한 분포를 만족시키기에 대한 가정이 우선적으로 충족되어야 하므로, 분석 자료의 분포 형태에 대한 검정이 먼저 실시되어야 한다.

한편 비모수 모형은 분석 자료에 대한 생존 함수

2) 이 연구에서의 뉴스 기사 자료 수집에 사용된 텍스트 마이닝 기법에 대하여는 이 논문 Ⅲ장 2절과 각주 5)의 내용을 참조하기 바란다.

수의 형태를 가정하지 않는 모형으로, Kaplan and Meier(1958) 모형, Nelson(1972)과 Aalen (1978) 모형이 이에 속한다. 비모수 모형은 생존 함수의 형태에 대한 어떠한 전제도 가정하지 않고, 자료 자체를 통해 특정 이벤트가 발생할 때까지의 지속 기간이 유지될 확률을 구하는 방식을 사용한다. 비모수 모형은 분석 자료 자체가 특정 이벤트가 발생할 때까지 지속되는 기간 또는 생존 기간에 대한 확률을 구하는 경우에는 유용한 정보를 제공할 수 있으나, 단변량 분석 모형이라는 점과 연속형 변수를 모형 내에 투입할 수 없다는 한계를 갖는다(송경일·최중수, 2011).

모수적 모형과 비모수적 모형의 장점과 한계점을 절충한 모형으로, 반모수적 모형은 모수적 특성의 독립변수와 비모수적 특성의 독립변수를 동시에 모형 내에 포함할 수 있는 다변량(multivariate) 모형이라는 점과 연속형 자료를 모형 내 포함할 수 있다는 특징을 갖는다. 반모수적 모형은 Cox 비례위험 모형(Cox proportional hazard model)으로도 불리는데, 이는 영국의 통계학자 Cox(1972)가 제시한 비례위험 모형이 곧 반모수적 모형을 의미하기 때문이다. Cox 비례위험 모형은 생존함수를 특정 모수적 분포로 제한하지 않고 공변량의 효과를 모수화하여 기저(baseline) 생존 함수를 변화시키는 방식을 사용하여, 생존함수에 미치는 독립변수의 효과를 계량적으로 측정할 수 있는 모형의 우수성과 계산의 용이성으로 인하여 현재까지 널리 이용되고 있는 분석방법이다(Cleves et al., 2010).

지속 의존성 분석에서 이벤트 발생 기간 t 에 영향을 미치게 되는 독립변수에 대한 다변량 모형은 <식 1>과 같이 일반적인 다중회귀 모형으로 가정할 수 있다.

$$t = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon \quad \langle \text{식 1} \rangle$$

하지만, <식 1>의 다중회귀모형은 종속변수가 정규 분포를 따를 것을 가정하고 있으므로, 시간 t 를 종속변수로 설정하는 경우에는 이러한 가정을 충족시키지 못하게 된다. 또한, 시간 t 는 음수(-)를 갖지 못한다는 점과 일반적인 회귀모형에서는 중도절단 자료를 처리할 수 없다는 한계를 갖기 때문에 지속기간을 추정하기 위한 모형으로서 <식 1>은 적합한 모형이 될 수 없다.

Cox(1972)는 이벤트 발생 시점까지의 지속기간과 공변량들의 관계를 지수 함수(exponential function) 관계로 가정하여 <식 2>와 같은 모형을 제시하였다.

$$h(t|x_j) = h_0(t) \exp(x_j \beta_x) \quad \langle \text{식 2} \rangle$$

- $h(t|x_j)$: 개체 j 의 t 시점에서의 위험률
- $h_0(t)$: 기저 위험함수(baseline hazard)
- x_j : 공변량(covariate)
- β_x : 공변량 x 에 대한 회귀계수

<식 2>에서 $h_0(t)$ 는 모든 x 가 0일 때의 위험률인 기저선(baseline)을 의미하며, β_x 는 지속기간에 대한 회귀 계수를 의미한다. 이 때, <식 2>는 <식 3>과 같이 변형될 수 있다.

$$\ln \left(\frac{h(t|x_j)}{h_0(t)} \right) = x_j \beta_x \quad \langle \text{식 3} \rangle$$

변형된 <식 3>은 기저 위험률 대비 x_j 에 대한 위험률의 변화율을 종속변수로, x_j 를 독립변수로 설정한 모형으로, 이는 독립변수의 1단위 변화에 따른 기저 위험률 대비 위험률 변화를 측정할 수 있는 모형이 된다. 즉, Cox 비례위험 모형이라는 명칭에서 비례 위험(proportional hazard)의 의미는, 독립변수에 대한 비례 위험률의 변화를 측정하는 모형임을 뜻한다.

〈식 2〉에서 β_x 는 Cox 비례위험 모형에서 추정되어야 하는 추정 계수를 의미하는데, Cox(1972)는 이를 산정함에 있어 부분 우도(partial likelihood) 함수를 통해 추정하는 방법을 제시하였다³⁾. 이 때, 〈식 3〉의 부분 우도 함수를 사용한 β_x 의 추정은 이벤트의 발생 순서가 명확하면서 동시에 발생하지 않은 경우를 가정하고 있을 때 사용하는 추정 방법인데, 동일 시점에서 복수의 실패 이벤트가 발생하거나 이벤트의 발생 순서가 명확하지 않을 경우에는 변형된 형태의 우도 함수를 사용하게 된다. 실패 이벤트의 발생 순서가 명확하지 않은 상황에서, 이벤트가 동일 시점에 복수로 발생하지 않은 것이 확실한 경우에는 exact-marginal 계산법(Kalbfleish and Prentice, 2002)을 사용하게 되며, 동일 시점에 복수의 이벤트가 발생하는 경우에는 exact-partial 계산법(Cox, 1972; Kalbfleish and Prentice, 2002)을 사용하게 된다.

이 연구에서는 아파트 매매 또는 전세 가격 국면 지속 기간 자료를 구성함에 있어, 가격 상승 추세와 하락 추세가 종료되어 국면이 전환되는 시점별로 자료를 구성하였기 때문에, 이벤트(국면의 종료) 발생 시점이 연속적이지 않은 것은 명확하다. 하지만 가격 국면 지속기간 산정에 있어서 국면의 종료 시점에서 국면 시작 시점의 차이를 구하는 방식을 사용하였기 때문에, 자료만으로는 지속 기간의 순서를 판단하기 어려워 이벤트 발생 순서에 대한 명확성은 떨어지게 된다. 그

래서 이 연구에서는 분석 자료의 이러한 특성을 감안하여 exact-marginal 계산법을 사용하여 $\hat{\beta}_x$ 를 추정하였다⁴⁾.

한편, Cox 비례위험 모형 추정의 적합성 여부를 검정하기 위한 방법은 일반적인 회귀 모형의 적합성 검정 방식과 크게 다를 바가 없다(Cleves et al., 2010). Cox 비례위험 모형 추정에 대한 적합성 검정 방법에는 모형 오류(mis-specification) 검정법, 이상치(outliers) 검정법, 영향점(influential points) 검정법 등 다양한 방식이 존재하는데, 이 절에서는 추정된 $\hat{\beta}_x$ 가 올바르게 모수화 되었는지 여부를 검증하는 방식 중 하나인 Link Test를 사용하여 독립 변수 추정치의 적정성 여부를 검정하고자 한다.

Link Test는 추정된 계수의 제공이 통계적으로 유의한지 여부를 검정하는 방법으로서, Cox 비례위험 모형 추정을 통해 일차적으로 $\hat{\beta}_x$ 를 추정한 후, 〈식 4〉에 제시된 LRH(Log Relative-Hazard) 모형을 통해 Cox 비례위험 모형 추정치의 적합성 여부를 검정한다.

$$LRH = \beta_1(x\hat{\beta}_x) + \beta_2(x\hat{\beta}_x)^2 \quad \langle \text{식 4} \rangle$$

상기 〈식 4〉에서 $\hat{\beta}_x$ 의 추정이 적합하다면, $\beta_1 = 1$ 이면서 $\beta_2 = 0$ 이라는 조건을 만족할 것이라는 것이 본 검정 방법의 원리이다. 따라서 Link Test는 $\beta_1 = 1$ 및 $\beta_2 = 0$ 의 통계적 유의성을 검정하게 된다(Cleves et al., 2010).

3) Cox 비례위험 모형의 기저 위험함수 $h_0(t)$ 와 부분 우도함수의 추정에 대한 구체적 내용은 Klein and Moeschberger(2003), Nelson(1972), Aalen(1978), Cox(1972)를 참고하기 바란다.

4) 이 연구에서는 범용 통계 패키지인 Stata를 사용하여 Cox 비례위험 모형을 추정하였는데, Stata는 exact-marginal 로그 우도비의 계산에 있어 15-Point Gauss-Laguerre 구적법을 사용한다. exact-marginal 계산법의 원리에 대한 보다 자세한 사항은 Kalbfleish and Prentice(2002)에서 확인할 수 있다.

〈표 1〉 분석 자료(원자료) 기초 통계

구 분	AP(전국)	CP(전국)	AP(서울)	CP(서울)	K_HP_UP	K_HP_DOWN	K_CP_UP	K_CP_DOWN
관측수	294	294	294	294	294	294	294	294
평 균	65,101	57,240	67,113	59,271	5,588	4,245	5,752	3,276
표준편차	21,449	22,642	28,074	22,839	9,397	8,236	7,392	4,435
최소값	37,539	25,207	33,106	26,344	0,000	0,000	0,000	0,000
최대값	101,458	108,723	108,972	111,305	59,000	77,000	38,000	35,000

주 1) AP, CP 단위: 지수

2) K_HP_UP, K_HP_DOWN, K_CP_UP, K_CP_DOWN 단위: 빈도수

3) 기간: 1990년 1월 ~ 2014년 6월

2. 분석 자료와 기초 통계

이 연구에서는 뉴스 기사와 아파트 매매·전세 가격 국면 지속기간 간의 관계를 확인하기 위하여 1990년 1월부터 2014년 6월까지 294개월 동안의 KINDS⁵⁾ 뉴스 기사 자료와 KB부동산에서 수집한 AP(아파트 매매가 지수)와 CP(아파트 전세가 지수)를 사용하였다.

먼저, KINDS에서의 뉴스 기사 검색은 특정 기간별, 장르별, 면종별로 가능하며, 특정 검색어를 입력하여 기사 제목이나 본문에 검색어가 포함된 기사만을 검색할 수 있다. 이 연구에서는 ‘주택 가격’과 ‘전세 가격’이라는 검색어를 입력하여 뉴스 기사를 수집하였다. 뉴스 기사의 수집 기간은 KINDS에서 ‘주택 가격’과 ‘전세 가격’ 검색어로 최대한 검색이 가능한 시점인 1990년 1월 1일부터

2014년 6월 30일까지로 설정하여 일간 뉴스 기사 제목을 수집하였다⁶⁾. 수집된 뉴스 기사 건수는 ‘주택 가격’으로 검색할 경우 310,774건, ‘전세 가격’으로 검색할 경우 105,025건으로 나타났다. 이렇게 수집된 뉴스 기사 제목 중 ‘상승’이나 ‘하락’ 단어의 빈도수를 월 단위로 집계하여 K_HP_UP, K_HP_DOWN, K_CP_UP, K_CP_DOWN으로 명시하였다.

한편, 지속의존성 분석의 종속 변수에 해당되는 주택 매매가와 전세가의 대응 변수인 AP나 CP 자료는 KB국민은행에서 제공하는 전국 및 서울시 아파트 매매가·전세가 지수 중 계절 조정된 지수⁷⁾를 사용하였다. 분석을 위하여 수집된 원자료에 대한 기초 통계는 〈표 1〉에 제시된 바와 같다.

지속 의존성 분석을 위하여 AP와 CP에 대해

5) KINDS(Korean Intergrated News Database System)는 전국 종합일간지, 경제지, 영자신문, 지역 신문을 모두 통합하여 검색할 수 있는 기사 원문 검색 서비스로서 한국언론진흥재단에서 정보를 제공하고 있다. KINDS는 1990년부터 현재까지 최신 업데이트 된 뉴스 기사를 검색할 수 있는 서비스를 제공하고 있어, 일반인들의 기사 검색 목적은 물론이고 학문적 연구 목적을 충족시키기 위해 충분한 데이터 베이스를 구축하고 있다.

6) 1990년 1월 1일부터 2014년 6월 30일까지의 KINDS 뉴스 기사를 일간으로 사람이 직접 검색하여 자료를 수집할 경우 상당한 시간과 노력이 수반된다. 이에, 이 연구에서는 스크립트 언어인 파이썬(Python) 플러그인 중 인터넷 웹페이지에 게시된 내용을 수집하는 웹 크롤링(web crawling) 목적의 BeautifulSoup 패키지를 사용하여 KINDS 뉴스 기사를 검색어와 설정 기간에 맞게 수집하였다. 이 연구에서 사용된 파이썬 스크립트는 〈부록〉에 제시된 바와 같다.

7) 국민은행 부동산에서 제공하는 계절조정지수는 매매가격지수 및 전세가격지수에 X12-ARIMA 방법을 적용하여 산출된다(KB부동산, <http://nland.kbstar.com/quics?page=B025969>).

여 이 연구는 먼저 <식 5>와 같이 시계열 자료인 AP 및 CP 를 차분하여 ΔAP 및 ΔCP 를 산정하였다.

$$\begin{aligned} \text{만약, } \Delta AP \text{ 또는 } \Delta CP > 0 &\Rightarrow \text{상승 국면} \\ \Delta AP \text{ 또는 } \Delta CP < 0 &\Rightarrow \text{하락 국면} \\ \Delta AP \text{ 또는 } \Delta CP = 0 &\Rightarrow \text{보합} \end{aligned}$$

<식 5>

이 때, <식 5>와 같이 ΔAP 또는 ΔCP 가 0보다 크게 되면 이는 상승 국면을 의미하며, 반대로 0보다 작게 되면 하락 국면을 의미하고, 0일 경우에는 보합 국면을 의미한다. 따라서 이 연구에서는 ΔAP 또는 ΔCP 를 기준으로 상승 국면과 하락 국면을 구분하여 각각 국면별 지속 기간을 집계하여, AP 및 CP 의 가격 국면 지속기간을 산정하였다.

또한, 독립변수로 사용될 뉴스 기사 중 ‘상승’ 또는 ‘하락’ 단어수는 각 국면별 기간 동안의 평균 빈도수를 산정하여 가격 국면 지속기간에 대응시켰다.

상기 과정을 통해 작성된 지속 의존성 분석 자료 중 전국 아파트 매매·전세 가격 국면에 대한 기초 통계는 <표 2>에 제시된 바와 같다.

한편, <표 2>에 제시된 지속 의존성 분석 자료를 분석 대상별로 시간의 흐름에 따라 시각화하면 <그림 1>과 같다. <그림 1>에서 가로 막대는 AP , CP 의 상승 또는 하락 가격 국면 지속기간을 의미하며, 세로 막대는 ‘주택 가격’ 또는 ‘전세 가격’으로 검색한 KINDS 뉴스 기사 제목 중 가격 국면 지속기간 동안 ‘상승’ 또는 ‘하락’ 단어의 평균 빈도수를 의미한다.

<표 2>와 <그림 1>을 살펴보면, AP 의 상승 국면은 21번, 하락 국면은 20번이 반복되었음을 알 수 있으며, CP 또한 상승 국면은 21번, 하락 국면은 20번 반복되었음을 확인할 수 있다. AP 상승

국면이 지속기간 중 가장 짧았던 기간은 1개월로 확인되며, 가장 길게 지속된 기간은 35개월로 확인된다. 한편, AP 하락 국면 또한 가장 짧게 국면이 지속된 기간은 1개월이었으며, 가장 오래 지속된 기간은 12개월임을 알 수 있다. 전세가를 의미하는 CP 의 경우, 가장 오랫동안 상승 국면이 지속된 기간은 63개월로 나타났으며, 반대로 가장 오랫동안 하락 국면이 지속된 기간은 12개월임을 알 수 있다.

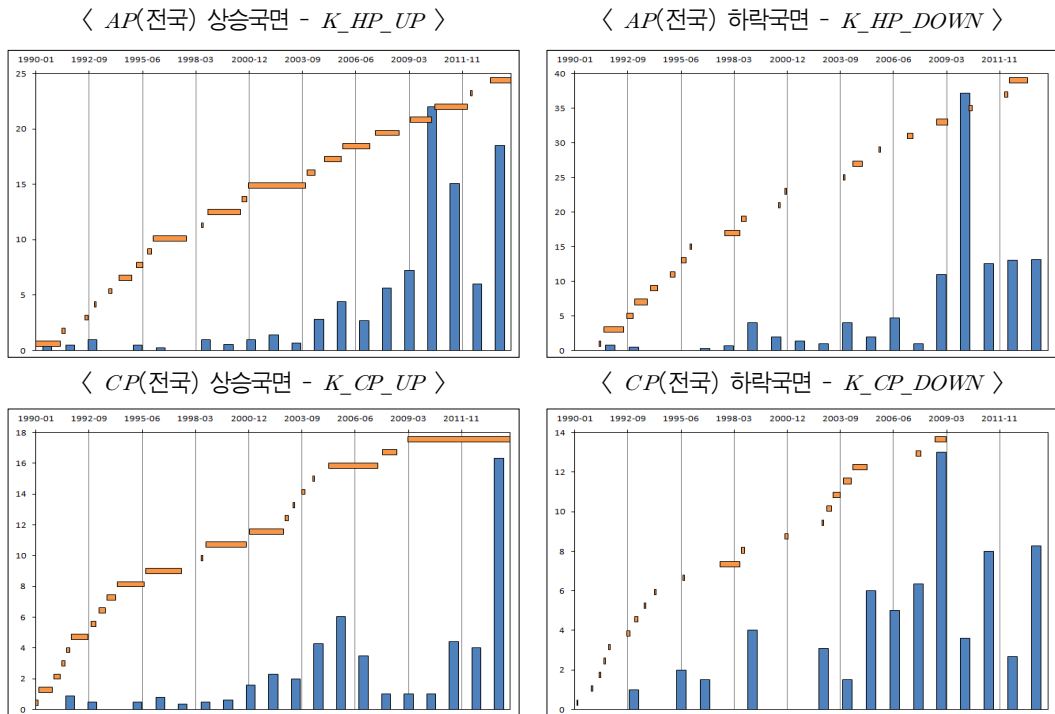
한편, 서울시 아파트 매매가와 전세가를 대상으로 작성한 지속 의존성 분석 자료의 기초 통계는 <표 3>에 제시된 바와 같으며, 이를 시간의 흐름에 따라 시각화한 그래프는 <그림 2>에 제시된 바와 같다. <표 3>과 <그림 2>를 살펴보면, 서울시의 경우 해당 기간 동안 AP 상승 및 하락 국면은 각각 28번씩 나타났으며, CP 의 상승 및 하락 국면은 각각 24번, 23번 나타난 것을 확인할 수 있다. 또한, AP 상승 국면 지속기간이 가장 길었던 기간은 23개월, 하락 국면 지속기간이 가장 길었던 기간은 29개월로 확인되며, CP 상승 국면이 가장 오래 지속되었던 기간은 36개월, 하락 국면이 가장 오래 지속된 기간은 13개월로 확인된다.

아파트 매매가 또는 전세가 국면 지속기간과 뉴스 기사 속 ‘상승’ 또는 ‘하락’이란 단어의 빈도수 간의 관계를 시각적 정보로 제시하고 있는 <그림 1>과 <그림 2> 그래프를 통해 확인할 때, 일부 구간에서 가로 막대의 길이와 세로 막대의 길이가 상관성을 갖는 것처럼 보이는 경우가 존재하는데, 이는 주택 가격 지속기간과 뉴스 기사 속 키워드 빈도수 간의 상관성이 존재할 수 있음을 의미하게 된다. 이에 대한 수치 정보 확인을 위해 다음 절에서는 지속 의존성 분석 결과를 제시하여 양자 간의 지속의존성 관계를 확인하였다.

〈표 2〉 지속 의존성 분석 자료 기초 통계: 전국

구 분		AP(전국) 상승국면	AP(전국) 하락국면	CP(전국) 상승국면	CP(전국) 하락국면
주 기 수		21	20	21	20
지속 기간	평 균	10.00	4.15	11.10	3.00
	min	1.00	1.00	1.00	1.00
	max	35.00	12.00	63.00	12.00
독립 변수	구 분	평균 K_{HP_UP}	평균 K_{HP_DOWN}	평균 K_{CP_UP}	평균 K_{CP_DOWN}
	평 균	4.38	5.45	2.45	3.30
	min	0.00	0.00	0.00	0.00
	max	22.00	37.14	16.35	13.00

주) 기간: 1990년 1월 ~ 2014년 6월



주 1) 가로 막대: 가격 국면(상승/하락) 지속기간

2) 세로 막대: 뉴스 기사 중 '상승' 또는 '하락' 단어 빈도수

〈그림 1〉 전국 아파트 매매·전세 가격 국면 - 뉴스 기사 단어 빈도수

IV. 실증분석

뉴스 기사에 노출된 '상승' 또는 '하락' 단어 빈도수에 대한 아파트 매매 또는 전세 가격 국면 지

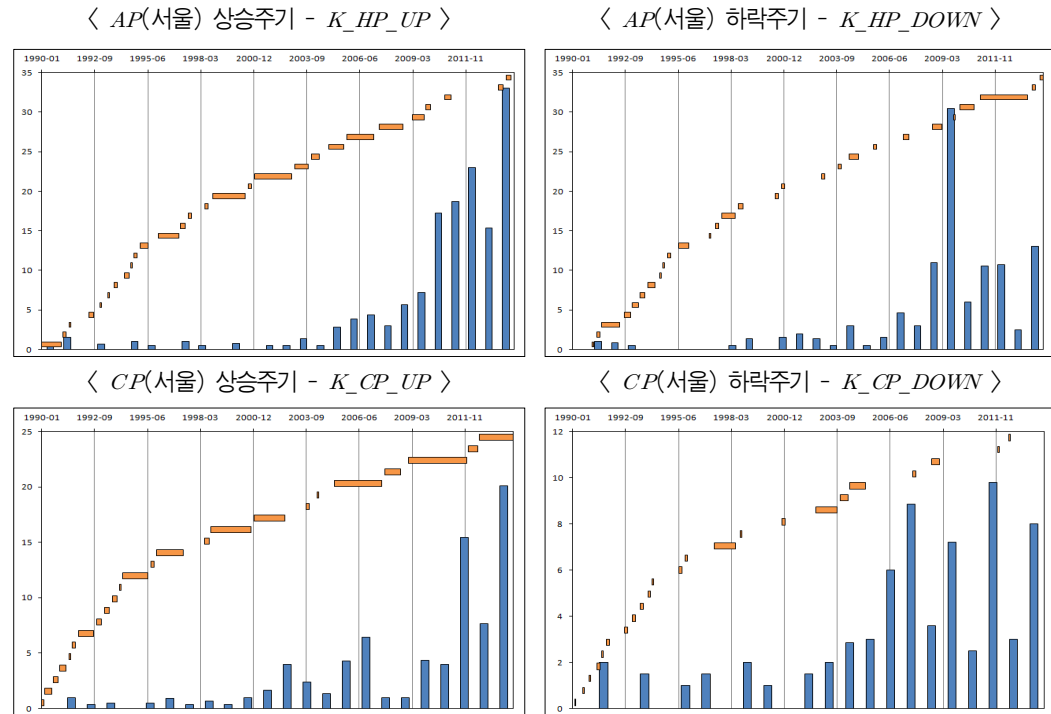
속기간 의존성 측정을 위한 Cox 비례위험 모형 추정 결과는 〈표 4〉에 제시된 바와 같다.

실증 분석에 있어, 지속기간 의존성 분석은 AP 또는 CP의 대상 지역에 따라 전국과 서울시로 구

〈표 3〉 지속 의존성 분석 자료 기초 통계: 서울

구 분		AP(서울) 상승주기	AP(서울) 하락주기	CP(서울) 상승주기	CP(서울) 하락주기
주 기 수		28	28	24	23
지속 기간	평 균	6.14	4.32	9.21	3.13
	min	1.00	1.00	1.00	1.00
	max	23.00	29.00	36.00	13.00
독립 변수	구 분	평균 K_{HP_UP}	평균 K_{HP_DOWN}	평균 K_{CP_UP}	평균 K_{CP_DOWN}
	평 균	5.13	3.80	3.30	2.93
	min	0.00	0.00	0.00	0.00
	max	33.00	30.50	20.10	9.80

주) 기간: 1990년 1월 ~ 2014년 6월



- 주 1) 가로 막대: 가격주기(상승/하락) 지속기간
2) 세로 막대: 뉴스 기사 중 '상승' 또는 '하락' 단어 빈도수

〈그림 2〉 서울 아파트 매매·전세 가격 주기 - 뉴스 기사 단어 빈도수

분하여 실시하였으며, Cox 비례위험 모형은 아파트 매매가 상승 및 하락 국면, 전세가 상승 및 하락 국면에 따라 4개의 모형으로 구성되었다. 이때, 독립변수는 각 가격 유형별 국면에 맞춰 국면

지속기간 동안의 평균 뉴스 기사 단어 빈도수로 설정하여 분석을 실시하였다.

먼저 전국에 대한 실증 분석 결과를 살펴보면, 4개의 Cox 비례위험 모형 중 추정 계수가 통계적

〈표 4〉 가격 국면 지속기간에 대한 Cox 비례위험 모형 추정 결과

종속변수 (가격국면 지속기간)	독립변수	전 국			서 울		
		위험률	계수(β_x)	LR $\chi^2(1)$	위험률	계수(β_x)	LR $\chi^2(1)$
AP 상승국면	K_HP_UP	0.9477	-0.0537 (-1.16)	1.60	0.9929	-0.0071 (-0.25)	0.06
AP 하락국면	K_HP_DOWN	0.9894	-0.0107 (-0.42)	0.19	0.9327	-0.0697 (-1.63)	3.71*
CP 상승국면	K_CP_UP	0.6039	-0.5043 (-2.43)**	13.92***	0.6451	-0.4383 (-2.58)**	18.14***
CP 하락국면	K_CP_DOWN	0.8298	-0.1866 (-2.05)**	5.27**	0.7966	-0.2274 (-2.22)**	6.39**

주 1) 분석 모형: $h(t|x_j) = h_0(t) \exp(x_j\beta_x)$

2) 계수 아래 괄호안 수치는 z-value를 의미

3) LR $\chi^2(1)$ 는 추정 모수 적합성에 대한 우도비(Likelihood Ratio) 검정 통계량을 의미

4) p-value: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

으로 유의성을 갖는 것은 CP(전국) 상승국면 - K_CP_UP 모형과 CP(전국) 하락국면 - K_CP_DOWN 모형 2개로 확인된다. 이 때, CP(전국) 상승국면 - K_CP_UP 모형에서 K_CP_UP 의 계수값은 -0.5043으로, 이에 대한 위험률(hazard ratio)은 0.6039로 확인되며,⁸⁾ CP(전국) 하락국면 - K_CP_DOWN 에서 K_CP_DOWN 의 계수값은 -0.1866이고, 이에 대한 위험률은 0.8298로 확인된다. 이 연구에서 사용한 Cox 비례위험 모형은 앞서 〈식 2〉에서 제시한 바와 같이 $h(t|x_j) = h_0(t) \exp(x_j\beta_x)$ 의 형태를 갖게 된다. 이에, 추정 계수 $\hat{\beta}_x$ 가 양(+)수라면 독립변수 증가에 따른 위험률은 점점 증가하게 되는 반면, $\hat{\beta}_x$ 가 음(-)수라면 독립변수 증가에 따른 위험률은 점점 감소하게 된다. 따라서 상기 통계적으로 유의한 두 개의 모형 모두에서 추정 계수값이 음(-)의 값을 갖는다는 것은 독립변수의 증가에 따라 가격 국면 지속기간이 종료될 위험률이 낮아짐을 의미한다. 즉, CP(전국) 상승국면 - K_CP_UP

모형의 추정 결과는 ‘전세 가격’과 관련된 뉴스 기사에 ‘상승’이란 단어가 많아질수록 전세가 상승국면 지속기간이 더욱 길어질 수 있음을 의미하는 것이며, CP(전국) 하락국면 - K_CP_DOWN 모형의 추정 결과는 ‘전세 가격’ 관련 뉴스 기사에 ‘하락’이란 단어의 빈도수가 증가할수록 전세가 하락 국면 지속기간이 유지될 확률이 높아짐을 의미하는 것이다.

또한, 서울시에 대한 실증 분석 결과도 전국에 대한 실증 분석 결과와 유사하게 4개의 Cox 비례위험 모형 중 CP(서울) 상승국면 - K_CP_UP 모형과 CP(서울) 하락국면 - K_CP_DOWN 모형에서 추정 계수 및 LR(likelihood ratio) 검정 결과가 통계적 유의성을 갖는 것으로 나타났다. 이 때, K_CP_UP 및 K_CP_DOWN 추정 계수는 각각 -0.4383과 -0.2274로 나타났으며, 위험률은 0.6451과 0.7966으로 산출되었다. 즉, 전국에 대한 분석 결과와 동일하게, 서울시에서도 ‘전세 가격’과 관련된 뉴스 기사에 ‘상승’ 또는 ‘하락’ 단어의

8) Cox 비례위험 모형 $h(t|x_j) = h_0(t) \exp(x_j\beta_x)$ 에 따라 위험률은 $\exp(-0.5043) = 0.6039$ 로 산정된다.

빈도수가 증가하게 되면 전세가 상승 또는 하락 국면 지속기간도 영향을 받게 됨을 의미하는 결과로 해석된다.

상기 추정 결과에서 주목할 만한 또 다른 점은 전국 및 서울시 모두에서 뉴스 기사에서의 ‘상승’ 또는 ‘하락’이란 단어의 빈도수가 아파트 매매가 국면 지속기간에 영향을 미치고 있지 못한 것으로 보이나, 전세가 국면 지속기간에는 통계적 유의수준 내에서 영향을 미친다는 것이다. 이 같은 결과는 트위터와 뉴스 기사에서의 주택 시장에 대한 어조가 주택 매매 시장보다는 전세 시장에 영향력을 발휘하고 있음을 의미하는 것으로 해석된다.

한편, 상기에서 추정된 Cox 비례위험 모형에 대한 모형 적합성 검정을 위하여 실시한 Link Test 결과는 <표 5>에 제시된 바와 같다.

검정 결과, 8개의 Cox 비례위험 모형 추정 결과에 대한 검정 모두에서 $(x\hat{\beta}_x)^2$ 가 통계적으로 유의성을 갖지 않는 것으로 나타났다. 또한, $x\hat{\beta}_x$ 에 대한 검정 추정 계수는 Cox 비례위험 모형의 통계적 유의성 결과와 유사한 결과를 나타냈는데, 전국의

경우 CP(전국) 상승국면 - K_CP_UP 모형 및 CP(전국) 하락국면 - K_CP_DOWN 에서의 $x\hat{\beta}_x$ 추정 계수가 10% 수준에서 통계적 유의성을 갖는 것으로 나타났으며, 서울시의 경우에는 CP(서울) 상승국면 - K_CP_UP 모형에서만 $x\hat{\beta}_x$ 추정 계수가 5% 수준의 통계적 유의성을 갖는 것으로 나타났다. 다만, 서울시의 CP(서울) 하락국면 - K_CP_DOWN 모형에서는 $x\hat{\beta}_x$ 추정 계수에 통계적 유의성이 존재하지 않는 것으로 나타났으나, Cox 비례위험 모형 추정 결과에서 추정 계수가 5% 이내의 통계적 유의성을 갖는다는 점과 Link Test에서의 $x\hat{\beta}_x$ 추정 계수가 통계적 유의성을 갖지 않는다는 점을 고려할 때, 추정된 모형은 적절한 것으로 판단된다.

<표 4>와 <표 5>에서 제시된 Cox 비례위험 모형 추정과 Link Test 결과 중 모형 및 독립변수가 통계적으로 유의하게 나타난 모형에 대한 위험함수 곡선은 <그림 3>에 제시된 바와 같다. 이 때, 위험함수 제시에 있어 독립변수의 변화에 따른 위험함수 곡선의 민감도 변화를 살펴보기 위하여 독립

<표 5> Cox 비례위험 모형 추정에 대한 적합성 검정: Link Test

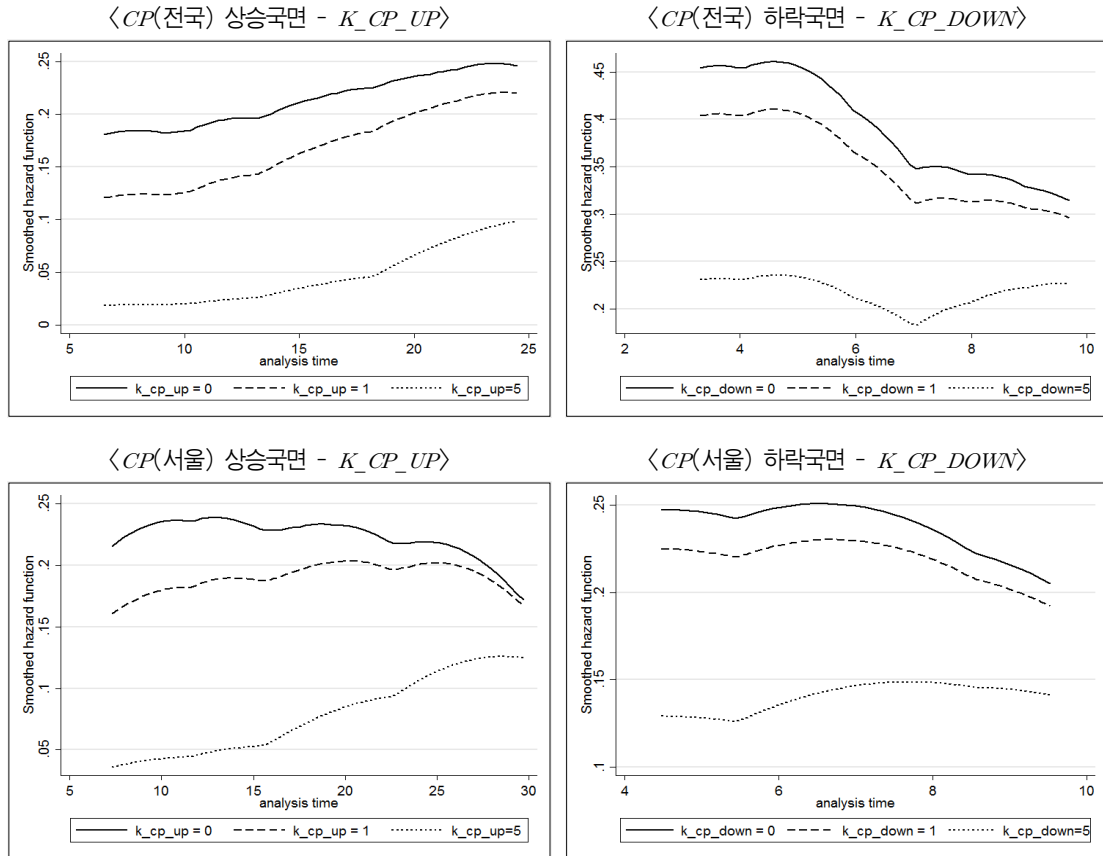
종속변수 (가격국면 지속기간)	독립변수	전 국			서 울		
		$x\hat{\beta}_x$	$(x\hat{\beta}_x)^2$	LR $\chi^2(1)$	$x\hat{\beta}_x$	$(x\hat{\beta}_x)^2$	LR $\chi^2(1)$
AP 상승국면	K_HP_UP	3.1566 (1.05)	2.1320 (0.78)	1.60	15.9090 (1.32)	76.1610 (1.39)	0.06
AP 하락국면	K_HP_DOWN	-1.0896 (-0.16)	-5.2847 (-0.28)	0.19	2.5920 (1.81)*	1.0175 (1.42)	3.71*
CP 상승국면	K_CP_UP	1.177 (1.75)*	0.0405 (0.33)	13.92***	1.2094 (2.00)**	0.0655 (0.71)	18.14***
CP 하락국면	K_CP_DOWN	2.1274 (1.80)*	0.662 (1.34)	5.27**	1.3963 (1.13)	0.3760 (0.69)	6.39**

주 1) Link Test 모형: $LRH = \beta_1(x\hat{\beta}_x) + \beta_2(x\hat{\beta}_x)^2$

2) 계수 아래 괄호안 수치는 z-value를 의미

3) LR $\chi^2(1)$ 는 추정 모수에 대한 우도비(Likelihood Ratio) 검정 통계량을 의미

4) p-value: * p < 0.1, ** p < 0.05, *** p < 0.01



〈그림 3〉 Cox 비례위험 모형에 대한 위험함수 곡선

변수를 0, 1, 그리고 5인 경우에 대하여 위험함수 곡선을 살펴보았다.

먼저, CP(전국) 상승국면 - K_{CP_UP} 모형에 대한 위험함수 곡선을 살펴보면, 시간의 흐름에 따라 곡선의 형태가 우상향하는 양상을 보이고 있으며, 이러한 우상향 추세는 약 24개월 정도에서 국면이 전환되는 것을 알 수 있다. 이는 전국 전세가 상승 국면이 종료될 위험률이 24개월까지 지속적으로 상승하게 됨을 의미하는 것으로, 만일 전세 가격 상승 국면이 24개월 정도까지 지속된다면 이후 가격 상승 국면이 종료될 위험률은 일정해 질 것임을 의미하게 된다. 이 때 독립변수인 K_{CP_UP} 의 값이 높아질수록 위험률이 점점 낮

아지는 것을 알 수 있는데, 이는 '전세 가격' 관련 뉴스 기사 속에 '상승'이란 단어가 많이 나타날수록 전세 가격 상승 국면이 오랫동안 지속됨을 의미하는 것으로, <표 2>의 추정 결과와 동일한 결과를 보여주고 있다.

CP(전국) 하락국면 - K_{CP_DOWN} 모형의 위험함수 곡선은 다소 다른 모양을 보여주고 있는데, 5개월까지 위험함수 곡선은 완만한 상승세를 보이다가 5개월 이후로는 급격하게 위험률이 떨어지는 모습을 보이고 있다. 이후 7개월 시점에서 위험률의 기울기가 다시 한 번 변화하는 것을 알 수 있다. 이는 5개월 시점까지 전세가 하락 국면이 종료될 위험률이 점진적으로 증가하지만, 하락 국

면이 5개월 이상 지속되게 되면 가격 하락 국면이 종료될 확률보다는 계속 지속될 확률이 더욱 높아지게 됨을 의미하는 것이다. 이 때, 독립변수인 K_CP_DOWN 의 민감도에 따른 위험률의 변화를 살펴보면, K_CP_DOWN 의 값이 높을수록 위험률이 낮아지는 양상을 나타내고 있음을 알 수 있다. 즉, ‘전세 가격’ 관련 뉴스 기사 속 ‘하락’이란 단어의 증가는 전세가 하락 국면이 지속될 생존률을 높이게 되어, 이로 인해 전세가 하락 국면의 지속기간이 더욱 길어지게 됨을 의미한다. 다만, K_CP_DOWN 의 민감도가 5정도로 높은 경우, 7개월 시점 이후부터 위험함수 곡선이 우상향하는 모습을 보이고 있다는 점에 차이가 있다.

CP (서울) 상승국면 - K_CP_UP 모형에서는 K_CP_UP 의 민감도에 따라 다른 위험함수 곡선의 양상을 나타내고 있다. 우선, K_CP_UP 가 0일 때의 기저 위험함수는 약 10개월까지 상승하다가 이후 완만하게 기울기가 우하향하는 모습을 나타내는 반면, K_CP_UP 가 1이거나 5일 때에는 지속적으로 우상향하는 모습을 보이고 있다. 하지만 위험함수의 상승 또는 하락 양상에도 불구하고 K_CP_UP 의 값이 높을수록 절대적인 위험률이 낮아지는 모습을 보이고 있어, 뉴스 기사 속 ‘상승’이란 단어의 증가가 서울시 전세가 상승 국면의 위험률을 낮추는 역할을 하고 있음을 알 수 있다. 또한 K_CP_UP 의 민감도에도 불구하고 상당한 기간이 지난 후 전세가 상승 국면의 위험함수가 비슷해지는 것을 알 수 있다.

CP (서울) 하락국면 - K_CP_DOWN 모형도 시간 경과에 따라 위험함수 곡선이 완만하게 우하향하는 모습을 보이고 있는데, 이러한 추세가 7개월이 경과되면서부터 더욱 급격해짐을 알 수 있다. 하지만, K_CP_DOWN 의 값이 5일 때에는 약

6개월 이후 시점부터 위험함수가 다소 증가하는 모습을 나타내고 있다. 한편, K_CP_DOWN 값의 증가에 따라 서울시 전세가 하락 국면의 위험률은 낮아지고 있음을 알 수 있어, 뉴스 기사에 ‘하락’이란 단어가 증가할수록 전세가 하락 국면이 더욱 오래 지속됨을 확인할 수 있다.

V. 결론

이 연구는 뉴스 기사 속 어조와 주택 시장 분위기 형성의 중요 요소인 가격 국면 지속기간 간 관계 확인을 위하여, ‘주택 가격’ 또는 ‘전세 가격’으로 검색된 뉴스 기사 제목에 나타난 ‘상승’ 또는 ‘하락’ 단어의 빈도수와 아파트 매매 또는 전세 가격 국면 지속기간 간의 관계를 실증 분석하였다.

분석을 위해 이 연구는 전국 및 서울로 구분하여 지속 의존성 분석 모형을 설정하였으며, 분석 모형으로는 모수적 및 비모수적 지속 의존성 분석 모형의 장점이 절충된 Cox 비례위험 모형을 사용하여 뉴스 기사 속 ‘상승’ 또는 ‘하락’이란 단어의 빈도수에 따른 가격 국면 지속기간 위험률의 변화를 살펴보았다.

Cox 비례위험 모형 추정 결과, 뉴스 기사 속 가격 관련 정보는 아파트 매매 가격 국면 지속기간에는 별다른 영향을 미치지 못하는 것으로 확인되었으나, 전세 가격 국면 지속기간에는 영향을 미치고 있었으며, 이러한 현상은 전국과 서울 모두에서 동일한 것으로 나타났다. 이러한 분석 결과를 통해, 뉴스 기사에 나타난 주택 시장 관련 어조가 전국 및 서울 전세 시장에서 일관된 방향으로 영향력을 미치고 있음을 확인할 수 있었다.

또한, 분석 결과는 뉴스 기사 제목에 나타난 ‘상승’ 또는 ‘하락’ 단어의 빈도수가 증가할수록 전세

가격 상승 또는 하락 국면 종료의 위험률이 낮아지는 것으로 나타났다. 이는 전세 가격과 관련된 뉴스 기사에 '상승' 또는 '하락'이라는 단어가 증가할수록 전세가의 상승 또는 하락 추세가 더욱 오랫동안 지속됨을 의미하는 것으로 뉴스 기사 속 어조에 따라 주택 가격 상승 또는 하락 국면 지속기간이 영향을 받고 있음을 의미하는 결과로 해석된다. 이에 이 연구의 실증 분석 결과는 시장 분위기 형성에 중요한 역할을 하는 가격 국면 지속기간 결정에 있어 뉴스 기사 속 어조가 중요한 변수로 작용하고 있음을 보여주고 있으며, 특히 매매 가격보다는 전세 가격 지속기간 결정에 영향을 미치고 있음을 제시하고 있다는 점에서 의의가 있을 것이다.

다만, 상기 제시된 연구의 의의에도 불구하고 이 연구는 다음과 같은 한계를 갖는다⁹⁾.

먼저, 이 연구 수행을 위한 뉴스 기사 자료 수집에 있어 대상 기간은 1990년부터 2014년까지 폭넓은 구간으로 설정되어 있다. 이에 1990년대와 2000년대에 수집된 뉴스 기사의 자료는 뉴스 미디어의 증가와 뉴스 기사 유통 채널의 다양화 등으로 인하여 양적인 차이가 존재하게 되는데, 이 연구에서는 이러한 차이를 통제하지 않고 있는 그대로의 수준 변수를 사용하였다는 한계를 갖는다. 이를 극복하기 위한 방안으로는 시기별 언론사 수에 대한 가중치 부여 또는 표준화 작업 등을 제시할 수 있을 것이다.

또한, 이 연구에서는 뉴스 기사 수집에 있어서 '주택 가격'이라는 키워드를 사용한 반면, 주택 가격에 대한 대응변수로서 아파트매매가지수를 사용하였다는 점에서 논란의 여지가 있을 수 있

다. 이러한 변수 선정은 국내 주택 시장의 매매 및 전세 가격 결정에 있어 아파트 시장이 주도적인 역할을 한다는 가정에 따른 것이나, 이로 인해 단독 및 다세대 주택 등이 배제된다는 한계를 갖는다.

아울러, 이 연구에서는 뉴스 기사 수집에 있어 '주택가격'과 '전세가격'을 키워드로 사용하여 검색된 자료를 수집하였는데, 이 때 기획 광고성 기사와 해외 시장에 대한 뉴스 등 다양한 내용과 형태의 뉴스 기사도 같이 검색될 수 있다. 이에 대하여 이 연구에서는 기사의 목적과 형태를 구분하지 않고 있는 그대로 수집된 자료를 사용하였는데, 이는 시장에서 유통되는 정보에 존재하는 노이즈(noise)도 시장 정보의 일부로 인정하고자 하였기 때문이다. 이에 이 연구에서의 분석 결과 해석과 적용에 있어 이를 감안하여야 할 것이다.

이러한 연구의 한계는 향후 후속 연구 수행에 있어 기술적 보완 및 자료 수집 기법의 개선을 통해 극복될 수 있기를 기대하며, 모쪼록 이 연구에서 도출된 결과를 토대로 향후 전세 시장 가격 및 동향 예측에 있어 뉴스 기사가 유용한 정보로 활용될 수 있기를 기대해 본다.

참고문헌

- 김대원·유정석, 2013, "초기 주택가격 형성이 가격 주기 지속기간에 미치는 영향 분석", 『부동산학연구』, 19(4): 87~108.
- 김진유, 2006, "신문기사가 부동산가격변동에 미치는 영

9) 이 연구에 대한 한계점과 보완점은 초고에는 간과되었던 부분이나, 심사과정 중 익명의 심사위원들의 의견을 바탕으로 제시되었다. 이 논문에 대하여 아낌없이 좋은 의견을 제시해 주신 익명의 심사위원들께 감사드린다.

- 향: '투기'가 포함된 신문기사와 주택가격간의 그래프
저인과관계분석을 중심으로, 『주택연구』, 14(2):
39~63.
- 송경일·최종수, 2011, 『(SPSS 15를 이용한) 생존자료의
분석』, 서울: 한나래.
- 우윤석·이은정, 2011, “언론보도와 시계열 주택가격 간의
관계에 관한 연구”, 『주택연구』, 19(4): 111~134.
- 진창하·Gallimore, P., 2012, “신문기사 내용과 주택가격:
인식, 사유, 그리고 투자심리”, 『부동산학연구』,
18(2): 125~142.
- Aalen, O., 1978, “Nonparametric Inference for a Family
of Counting Processes”, *The Annals of Statistics*,
6(4): 701~726.
- Bracke, P., 2013, “How Long Do Housing Cycles Last? A
Duration Analysis for 19 OECD Countries”,
Journal of Housing Economics, 22(3): 213~230.
- Cleves, M., et al., 2010, *An Introduction to Survival
Analysis Using Stata*, 3rd ed., Stata Press.
- Cox, D. R., 1972, “Regression Models and Life-Tables”,
*Journal of the Royal Statistical Society, Series
B(Methodology)*, 34(2): 187~220.
- Davison, W. P., 1983, “The Third-person Effect in
Communication”, *Public Opinion Quarterly*, 47(1):
1~15.
- Gayer, T. and Viscusi, W. K., 2002, “Housing Price
Responses to Newspaper Publicity of Hazardous
Waste Sites”, *Resource and Energy Economics*,
24(1): 33~51.
- Kalbfleisch, J. D. and Prentice, R. L., 2002, *The
Statistical Analysis of Failure Time Data*, 2nd ed.,
Wiley-Interscience.
- Kaplan, E. L. and Meier, P., 1958, “Nonparametric
Estimation from Incomplete Observations”,
Journal of the American Statistical Association,
53(282): 457~481.
- Klein, J. P. and Moeschberger, M. L., 2003, *Survival
Analysis: Techniques for Censored and Truncated
Data*, Springer Science & Business Media.
- Mercille, J., 2014, “The Role of the Media in Sustaining
Ireland’s Housing Bubble”, *New Political
Economy*, 19(2): 282~301.
- Nelson, W., 1972, “Theory and Applications of Hazard
Plotting for Censored Failure Data”, *Technometrics*,
14(4): 945~966.
- Tsfati, Y. and Cohen, J., 2003, “On the Effect of the
‘Third Person Effect’: Perceived Influence of
Media Coverage and Residential Mobility
Intentions”, *Journal of Communication*, 53(4):
711~727.
- KB부동산 <http://nland.kbstar.com/quics?page=B025969>

원 고 접 수 일 : 2015년 10월 29일
1 차 심 사 완 료 일 : 2016년 1월 29일
2 차 심 사 완 료 일 : 2016년 3월 31일
최 종 원 고 채 택 일 : 2016년 5월 17일

<부록> KINDS 뉴스 기사 제목 크롤링(Crawling) Python Scripts

```

# Python Ver.3.XX 전용 Scripts
# 주택가격 검색어에 대한 Crawling Scripts
# 검색어 및 조건 변경시 "url = ..." 부분 변경

import urllib.request
from bs4 import BeautifulSoup

for irt in range(62, 63):          # 수집할 페이지 범위 설정
    i = ((irt - 1) * 500) + 1
    #i = 2001                    #1이 1페이지 (11페이지는 11, 21페이지는 21)
    k = i + 499                  #10이 10페이지 (20페이지는 20, 30페이지는 30)

    dw = open('kinds__article(' + str(i) + '-' + str(k) + ').csv', 'w')

    for m in range(i - 1, k):
        s = m * 10
        url =
        'http://www.mediagaon.or.kr/jsp/sch/mnews/search.jsp?collection=mkind&query=%C1%D6%C5%C3+%
        B0%A1%B0%DD&startDate=1990.01.01&endDate=2014.06.30&pageStartNumber=' + str(s)
        html = urllib.request.urlopen(url)
        soup = BeautifulSoup(html)
        ns = soup.find_all('a',{'class':'mnews_kinds_not_ori_link'})
        dt = soup.find_all('td',{'class':'sort'})

        z = 0

        for a in ns:
            b = dt[z].text
            print(a.text, b)
            dw.write(a.text + b + '\n') #검색 결과를 txt로 출력
            z = z + 1

dw.close()

# Python Scripts for web crawling citation information
# Dai Won, Kim(2015). https://sites.google.com/site/phdkimdaiwon/

```