

부동산 과소거래시장에 대한 안정적인 실거래가 지수 산정에 관한 연구*

A Stable Transaction Based Price Index for Real Estate Thin Market

송 영 선 (Song, Young-Sun)**
신 혜 영 (Shin, Hye-Young)***
이 창 무 (Lee, Chang-Moo)****

< Abstract >

Recently, the scale of investment in the domestic office market has grown rapidly, and the demand for market information transparency is increasing as movements to revitalize indirect real estate investment are active. But information disclosure is quite limited and indicators for detailed diagnosis of market conditions are not enough. In particular, the price index, which can be used as a basic index for time point correction in evaluating asset value, an important factor as a denominator of the capitalization rate, has been released in various private companies in the past, but only some are maintained.

Compared to the housing market, the office market has significantly less stock of the transaction unit and thus extremely low frequency of transactions, making it difficult to estimate a stable index using actual transaction cases. Therefore, this study intends to present a more advanced methodology for estimating the Seoul Office Price Index. To this end, the Bayesian approach was introduced instead of the OLS Repeat Sales model, which has been widely used, and the index was estimated using three Bayesian trend models.

After estimating the index, as a result of comparing it with the index using the existing methodology, significantly improved index estimation performance was confirmed in various aspects such as the index stability, statistical reliability, and index change(Revision). However, additional considerations are still needed regarding the Revision, which is important in terms of the utility of the index. Nevertheless, this study suggested a way to solve problems that have been difficult to estimate a stable index due to low transaction frequency, and in the future, it is expected that more detailed market information can be provided by estimating sub-indices such as district or size category.

Keyword : Thin Market, Office Price Index, Repeat Sales Index, Bayesian Trend Model, Markov Chain Monte Carlo

* 이 논문은 2021년 한국부동산분석학회 상반기 학술대회 발표내용의 일부를 수정·보완하였음.
** 한양대학교 일반대학원 도시공학과 석박사통합과정, youngsun_song@naver.com, 주저자
*** 한양대학교 일반대학원 도시공학과 박사과정, shy@lh.or.kr, 공동저자
**** 본 학회 정회원, 한양대학교 도시공학과 교수, changmoo@hanyang.ac.kr, 교신저자

I. 서론

부동산 가격지수는 부동산 시장의 상황을 파악하기 위한 대표적인 지표 중 하나로 꾸준히 관련 연구가 진행되어져 왔다.

과거에는 조사가격 혹은 평가가격을 기초로 한 부동산 가격지수가 국내에서 대표적으로 활용되었으나 실거래가격을 완전히 대변하지는 못하였으며(이상경, 2007), 2006년부터 부동산 거래신고제도가 시행되고 실거래 자료가 누적되어 확보되면서 실거래가격에 기초한 부동산 가격지수의 산정에 관한 연구가 활발히 진행되어져 왔다. 특히 실거래가격 지수 중에서도 다양한 특성 변수를 필요로 하는 헤도닉 모형 대신에 거래가격과 거래시점 정보만으로 간편하게 산출이 가능한 반복매매모형이 국내외에서 보편적으로 실거래가격지수 산정에 이용되고 있다.

그러나 실거래가격 지수는 부동산거래가 관측된 정보만을 활용하기 때문에 거래 빈도가 적은 특정 부동산 유형 혹은 세분화된 하부시장에 대해서는 안정적인 지수 산정에 어려움이 있다. 특히, 시간과 비용 등의 경제성 측면에서 헤도닉모형에 비해 선호되는 반복매매모형은 2번 이상 거래가 관측된 자료를 이용함으로써 근본적으로 비효율성의 문제를 갖고 있는데 거래 빈도가 적은 시장에서는 그 한계가 더욱 증폭되어 나타날 수밖에 없다. 국내에서는 이러한 한계 개선을 위하여 MIT 부동산 연구센터에서 개발된 2단계 추정법(Geltner and Bokhari, 2008)을 활용하거나(황규완·손재영, 2017; 류강민·송기욱, 2020) S&P/Case-Shiller Home Price Index에 이용되는 자료 중첩 방식을 활용하여(송영선 외, 2020) 거래빈도가 적은 시장의 실거래가격 지수를 산정하고자 하였으나 여전히 만족스럽지 못한 측면이 있다.

거래 빈도가 적은 부동산 유형으로는 오피스 빌딩이 대표적인 한 예이다. 국내 서울 오피스 시장의 투자 규모는 지속적으로 상승세를 보여왔으며, 부동산 간접투자가 점차 활성화되면서 오피스 투자규모 증대와 투자자 다변화가 이루어졌다(류강민·송기욱, 2020). 이에 따라 시장 투명성에 대한 요구가 증가하고 있으나 여전히 시장 상황을 판단하고 투자 여부를 결정하는데 활용할 수 있는 시장 정보의 제공이 부족한 것이 현실이다. 시장 상황을 보여줄 수 있는 다양한 정보들 중에서도 오피스 매매가격 지수는 오피스 빌딩에 투자했을

때 얼마의 수익을 얻을 수 있는가를 판단하기 위해 기본적으로 요구되는 정보이다(최성호 외, 2010). 국내에서 오피스 매매가격지수는 이지스자산운용·대신증권에서 생산·발표되었으나 2020년 3분기를 마지막으로 중단된 상태이며 그나마 에비슨영코리아에서 마켓리포트를 통해 제공하고 있으나 미국의 RCA(Real Capital Analytics)에서 전세계 15개국 72개 지역을 대상으로 작성하고 있는 CPPI(Commercial Property Price Index)와 비교하면 안정적인 지수 산정에 큰 어려움을 겪고 있는 상황이다.

관측되는 거래 빈도가 적은 시장에 대한 안정적인 지수의 작성은 보다 세분화된 정보 제공을 통해 시장 투명성 제고에 기여할 수 있으며, 급격한 성장에도 불구하고 폐쇄적인 정보 형태를 갖고 있는 국내 오피스 시장의 적용은 정부의 「공모형 부동산간접투자 활성화 방안」(2019)과도 그 목적과 방향성을 공유하는 측면이 있다는 데에 의의가 있다.

따라서 본 연구에서는 서울시 오피스 시장을 대상으로 하여 기존에 국내에서 이용되어져 온 반복매매지수 산정 방식에서 탈피하여 인접 시점을 서로 연결하는 형태의 추세모형과 베이지안 방법론을 활용하고자 한다. 이러한 방법론을 적용하면 시점별 거래 자료 부족으로 이상치에 크게 영향을 받아 과적합된 지수 추정을 방지하고 시점 간 연결고리로 보다 안정적인 실거래가격 지수의 추정이 가능할 것으로 기대한다. 연구모형의 지수 추정 성능을 평가하기 위해서 가장 기본적인 형태의 OLS 일반 반복매매모형, MIT/CRE 2단계 추정법과 비교하며 통계적 신뢰도, 지수 안정성, 지수변화(Revision), 모형의 예측력의 측면에서 평가하고자 한다.

II. 선행연구 고찰

1. 부동산 가격지수 산정 방식에 관한 연구

부동산 가격지수는 이용 자료에 따라 시세조사에 기반한 지수와 실거래가격에 기반한 지수로 나눌 수 있다. 과거에 국내에서는 실거래 자료의 수집이 어려워 시세조사에 기반한 지수 작성이 주가 되었다. 그러나 실거래 자료의 수집이 가능해지면서부터 실거래가격

에 기반한 지수의 산정 방식에 관한 연구가 꾸준히 진행되었는데 시세 지수와 실거래가 지수가 각각 가지는 장단점이 있겠으나 시세 지수가 실거래가 지수에 비해 변동성이 낮고, 시장가격의 변동에 비해 그 움직임이 늦게 나타나며(이용만·이상한, 2008), 특히 가격이 하락하는 시기에 그 차이가 확대된다(이창무 외, 2005; 박연우·방두완, 2011)는 지적이 꾸준히 제기되었다. 따라서 2006년 부동산거래 신고제의 시행으로 실거래 자료 수집이 가능해진 것과 더불어 시세 지수가 가진 문제점을 개선하여 시장상황을 보다 잘 반영할 수 있는 대안으로 실거래가격 지수에 관한 연구가 활발히 이루어졌다.

실거래가격 지수의 한 가지로 널리 활용된 헤도닉 모형(Hedonic Model)은 Lancaster(1966), Rosen(1974)를 통해 그 이론적인 토대가 마련되었으며 국내에서도 이상경(2005, 2007), 이용만(2007), 이영유·이상경(2013) 등을 통해 헤도닉가격지수에 관한 연구가 진행되었다. 헤도닉 모형은 부동산의 특성가격을 추정한 후에 이를 이용하여 매 시점 추정 가격을 지수화 하는 방식으로(이용만, 2007) 부동산의 이질성을 극복할 수 있다는 장점을 가지고 있으나 방대한 부동산의 각종 특성에 관한 자료가 요구되어 많은 시간과 비용이 요구되며 변수 누락으로 인한 편의가 발생할 수 있다는 단점이 있다(이창무 외, 2002; 최성호 외, 2010).

실거래가격 지수 산정 방식의 다른 한 가지로 현재 가장 보편적으로 활용되고 있는 반복매매모형(Repeat Sales Model)은 Bailey et al.(1963)을 통해 제안되었고 Case and Shiller(1989), Shiller(1991)을 통해 수정된 형태로 널리 적용되면서 현재에도 S&P/Case - Shiller Home Price Index, Costar CCRSI(Costar Commercial Repeat-Sale Indices), 한국부동산원 공동주택 실거래가격지수 등 많은 기관에서 부동산 가격지수 작성에 이용하고 있다. 헤도닉 모형과 달리 지수작성 기간 내에 두 번 이상 거래된 부동산의 거래가격과 거래시점 자료만을 필요로 하며 거래가 발생한 두 시점 사이에 부동산 가격에 영향을 미치는 특성이 변하지 않는다는 가정을 기초로 모형이 구성되기 때문에 헤도닉 모형의 단점에 대한 대안으로 활용될 수 있다. 그러나 이로 인하여 자료 이용의 효율성이 떨어지고 표본선택의 편의에 관한 문제를 갖고 있고(Clapp and Giaccotto, 1992), 관측되는 반복 거래시점 간

차이가 길어지면서 이분산(heteroscedasticity)의 문제가 발생할 수 있다(이창무 외, 2002).

이와 같이 반복매매모형을 적용하는 경우 나타나게 되는 문제들에 대한 개선점을 찾고자 하는 연구도 다양하게 이루어져왔다. 이창무·백익민(2008)은 SPAR 모형의 공시가격과 실거래가격 비율의 개념에서 착안하여 시세가격도 함께 활용한 반복매매지수를 산정하였다. 이를 통해 자료 이용의 효율성과 표본추출 오류의 영향, 통계적 신뢰도를 크게 개선하고 지수 안정성을 확보하는 결과를 얻을 수 있었다. 류강민·이상영(2010), 최성호 외(2010)은 Case and Shiller(1987)가 제안한 WRS(Weighted Repeat Sale) 모형을 활용하여 이분산의 문제를 해결하고자 하였다. 표본의 편의 문제를 보완할 수 있는 방안으로는 Coulson and McMillen(2007), McMillen and Thorsnes(2006)가 분위회귀(Quantile Regression) 모형을 제안하였는데 이창무 외(2013)에서는 분위회귀 반복매매모형을 이용하여 서울시 아파트 실거래가 지수를 산정하였고 지수의 불안정성을 개선할 수 있는 가능성을 확인하였으나 표본 수가 극히 적은 경우에는 한계가 나타났다.

2. 거래빈도가 적은 시장에 대한 지수 산정 방식

단독, 연립/다세대주택과 같은 비아파트, 동별로 세분화 된 아파트 혹은 상업용부동산 시장과 같이 기본적으로 거래 발생 빈도가 적은 시장에 대해서는 반복매매지수를 이용하여 지수를 작성할 경우 반복매매모형의 단점이 보다 극명하게 두드러져 상당히 불안정한 지수가 산정된다. 이러한 문제를 해결하기 위한 다양한 방식의 연구가 국내외에서 진행되어져 왔는데, 해외에서는 이러한 연구를 통해 실제 지수 공표까지 이어져 기본적인 틀이 정립이 되고 있는 반면에 국내에서는 기존의 방법론의 한계 범위 내에서 가능한 수준까지만 지수의 작성 및 공표가 이루어지고 있고 더 발전된 연구 성과를 통해 보다 세분화되고 다양한 시장 지표 개발에 대한 노력이 부족한 실정이다.

특히 상업용부동산은 주택에 비하면 재고량이 크게 적고 물건단위 규모가 큰 관계로 거래가격이 높아서 거래 빈도가 적은 시장이다. 그 중에서도 오피스는 국내 상업용부동산 시장에서 가장 큰 거래 비중을 차지하는 대표적인 부동산 유형으로 국내의 관련 선행연구는 대부분 오피스를 대상으로 하고 있다.

국내에서 오피스 가격지수에 관한 연구는 1990년대 말부터 이루어져 왔으나 초기에는 임대료를 대상으로 한 연구가 주를 이루었고(손재영·김경환, 2000; 손진수·김병욱, 2002 등), 충분한 시계열이 확보되지 못한 관계로 헤도닉 모형을 이용한 연구가 진행되었다(이상경, 2005; 이상경, 2007). 특히 박헌수(2007)에서 변동성이 큰 헤도닉 가격지수의 대안으로 시계열 분석 방법을 적용하여 지수를 추정하였다. 과거 거래 자료와 현재의 거래 자료를 연결시킴으로서 비교거래 자료의 집합을 늘리는 형태로 지수 작성이 이루어졌는데, 이는 본 연구에서 활용한 추세 모형과 개념적으로 유사한 측면이 있다. 기본적으로 헤도닉 모형의 형태를 활용하였고 시계열에 대한 확률과정을 간단하게 가정했다는 점에서 한계를 지적하고 있으나 시계열 분석을 통한 가격지수 산정 결과 월 거래 자료가 2개인 경우에도 비교적 안정적인 지수 산정이 가능한 것을 확인하였다는 점에서는 의미가 있다.

이상경(2009)에서는 반복매매지수를 이용하여 오피스 매매가격지수를 산정하였으나 연 단위 지수를 산정함으로써 시점별로 비교적 적지 않은 거래 쌍을 확보하였고, 도심권, 강남권, 여의도권 등 투자 선호도가 높은 주요 지역의 거래 사례들로 자료가 구성되어 한계가 있었다. 반면에 최성호 외(2010)에서는 비교적 충실한 오피스 거래자료의 구축을 통해 오피스 반복매매지수를 작성하였다. 특히 기본적인 반복매매모형(Bailey et al., 1963)의 형태를 활용한 이상경(2009)와는 달리 GLS(Generalization Least Squares)를 이용한 WRS(Weighted Repeat Sales)를 통해 이분산을 고려하고(Case and Shiller, 1987; Shiller, 1991; 류강민·이상영, 2010) 연면적을 가중한 반복매매지수 산정을 시도하였다. Yeon(2016)은 벌점화 함수를 적용한 분위회귀 반복매매모형을 광주시 연립주택 실거래가 지수 산정에 활용하여 지수 안정성이 개선되는 결과를 얻었다.

과거 오피스 실거래 자료 구축이 어려움이 있던 당시에는 평가가치 기반 가격지수에 대한 연구가 진행되었고, 그 이후에는 실거래 자료를 활용하고자 하는 움직임이 커지면서 헤도닉 모형을 이용한 가격지수 연구가 주를 이루었다. 2006년 부동산 실거래 신고제가 시행된 이후로 실거래 자료의 수집이 가능해졌는데, 일정 수준 시계열이 쌓이면서부터는 반복매매지수의 작성이 시도되었다. 여기까지 보면 오피스 정보 및 투자

지수의 필요성에 대한 인식이 오래전부터 있어왔던 것과는 달리 자료 수집의 한계와 기본적으로 거래 빈도가 낮은 시장 특성으로 인해 실거래자료를 이용한 오피스 매매가격지수를 작성하는 것 자체에 집중된 연구들이 진행되어져 왔던 것으로 판단된다. 이후 국내에서는 해외만큼 통계적 기법 등을 통하여 보다 안정적이고 세분화된 지수 작성에 대한 연구가 활발히 이루어지지 않았다. 몇 가지 관련 연구들을 살펴보면 먼저 박헌수·유은영(2014)에서는 거래 빈도가 낮은 경우에 대한 실거래 가격지수를 작성하는 방법을 개발하고 더 나아가 미래의 주택가격 수준을 예측할 수 있는 모형에 대한 연구를 진행하였다. 오차항이 1차 자기회귀 확률과정을 따르는 상태공간모형을 활용하여 칼만 필터와 EM 알고리즘을 통해 서울시 종로구와 강남구 아파트에 대하여 규모별로 세분화한 지수와 강남구 삼성동의 지수를 추정하였다. 비교분석 결과 상태공간모형을 도입하는 경우 일반반복매매모형으로 추정하는 경우에 비해 지수 안정성이 개선되는 것을 확인하였으나 일부 규모에 대해서는 큰 효과가 없는 것으로 나타났다.

황규완·손재영(2017)과 류강민·송기욱(2020)은 오피스 가격지수 산정을 위해 Bokhari and Geltner(2012)의 연구에서 제안한 2단계 추정법(Two-stage Estimation)을 활용하여 오피스 매매가격지수를 작성하였다. 2단계 추정법은 과거 RCA(Real Capital Analytics)의 Moody's/Real CPPI 작성에 활용되었던 지수산정 방법론으로 검증된 방식으로 볼 수 있다. 2단계 추정법으로 지수를 추정한 결과 기존의 반복매매지수와 비교하여 지수 안정성이 크게 개선된 형태의 지수 작성이 가능하였으나 추정 방법의 특성상 시차가 발생하는 한계가 있다.

송영선 외(2020)은 안정적인 지수 산정을 위한 자료 처리 방식에 대한 연구를 진행하였다. 연이은 시점의 자료를 중첩하여 새로운 자료를 만들어 지수 작성에 이용하는 것으로 S&P / Case-Shiller Home Price Index의 작성에 이용되고 있는 방식(S&P Dow Jones Indices, 2020)이다. 서울시 강남구, 중구, 종로구의 아파트 실거래가 지수를 산정한 결과 기존의 일반반복매매지수 뿐만 아니라 이동평균지수, 2단계 추정법과 비교하여도 지수 안정성이 크게 개선된 지수 작성이 가능한 것을 확인하였다. 거래자료가 없는 특정 시점에 대해서도 자료를 새로이 만들어낸다는 점, 그리고

사전적인 자료 처리 방식에 관한 것으로 다양한 모형들에 적용하여 활용할 수 있다는 점에서 장점이 있다. 기본적으로 자료 중첩 방식으로 자료 수가 많은 시점의 가중치가 크게, 상대적으로 자료 수가 적은 월의 영향력이 작게 반영되어 안정적인 지수의 작성이 가능해지지만, 부동산 가격과 거래량 사이에 양의 상관관계가 존재하여 가격이 급격하게 하락하는 시장에서 관측치 감소의 우려가 있다는 점(Fisher et al., 2003; de Wit et al., 2013)을 함께 고려하면 시장 특성에 따라 가격 하락 폭에 대하여 과소평가할 가능성에 대한 고민이 필요할 것으로 보인다.

Goetzmann(1992)의 연구에서는 OLS(Ordinary Least Squares), GLS(Generalized Least Squares), WRS(Weighted Repeat-Sales) 모형, Stein-like estimator, Bayes estimator 등 다양한 방식을 통하여 지수를 산정하고 비교하였으며, 베이저안 접근법이 지수 추정 정확도를 높인다는 것을 확인하였다.

Francke(2010)은 Goetzmann(1992)에서 더미변수 대신에 도입한 이동량을 고려한 랜덤워크(Random Walk with drift) 확률과정을 개선한 보다 일반화된 형태의 국지선형추세(Local Linear Trend, LLT) 모형을 이용하여 지수를 추정하였다. 이용자 수가 많은 경우에는 기존의 방법론과 차이가 적고 사전확률의 영향력이 적어 데이터에 의존하는 결과가 나타났으나, 자료 수가 적은 과소거래시장에 대해서는 거래가격의 잡음(noise)에 민감한 기존 방법론과 달리 LLT 모형을 활용하는 경우에 안정적인 지수 추정이 가능했다.

van de Minne et al.(2020)는 안정적인 지수 추정과 더불어 반복매매모형의 한계로 인해 필연적으로 발생할 수밖에 없는 revision을 최소화 할 수 있는 방안으로 1차 자기회귀 과정을 도입한 AR 모형과 지수산정 대상 시장의 지역과 부동산 유형을 포함하는 교차하는 종합지수를 설명변수로 도입한 AG 모형을 결합한 형태의 방법론을 제안하였다.

이상에서 살펴본 바와 같이 과거 실거래가 지수가 국내에 자리 잡기 시작하던 당시에는 헤도닉모형과의 비교와 함께 반복매매모형의 구조적 특성이 가지는 표본선택의 편익, 이분산성 등의 한계 개선에 관한 연구가 적지 않게 이루어졌으나 이후에 거래 빈도가 적은 시장의 안정적인 지수 산정에 관한 연구는 주류가 되지 못하였다. 특히, 황규완·손재영(2017), 류강민·송기욱(2020)에서 제안한 방법론 정도가 실제 지수 생

산까지 이어졌을 뿐 그 외의 관련 연구들은 단순히 지수 안정성 개선을 시도해보는 정도에서 그쳤다. 그러나 시장 규모가 점차 확대되면서 이제는 과소거래시장에 대해서도 보다 세분화된 지수에 대한 요구가 증가하는 상황이며, 따라서 해외에서 연구가 선행되어 실제 지수 생산까지 이어진 모형의 기본적인 구조를 도입하여 시도하고자 한다. 본 연구에서 활용하는 베이저안 방법론은 최근 들어 컴퓨터 기술의 발전으로 보다 다양한 형태의 모형을 유연하게 시도해 볼 수 있는 장점을 바탕으로 기존 연구들이 가지는 한계를 보완하는데 유용하다. 또한 해외에서 검증된 모형의 형태를 차용하지만 특성이 다른 국내 시장 상황을 고려한 연구로서 한국형 지수 개발을 위한 기틀을 마련한다는 점에서 의미가 있다.

III. 이론적 논의

1. 반복매매모형

실거래가격을 이용한 부동산 가격지수 산정에 있어 반복매매모형은 현재 가장 보편화된 방법 중 하나이다. 충분한 반복거래쌍 구성이 가능하지만 하면 부동산의 가격에 영향을 미칠 수 있는 다양한 헤도닉 변수의 구축 없이 거래시점과 거래가격 정보만을 이용하여 시장에서의 가격변화를 잘 반영하는 지수의 산정이 가능하다.

반복매매모형은 다음의 식 (1)에서 부터 출발한다.

$$\ln P_{it} = \mu_t + \beta_i + \epsilon_{it}, \quad \epsilon_{it} \sim N(0, \sigma_\epsilon^2) \quad (1)$$

P_{it} 는 부동산 i 의 t 시점 거래가격을 나타내며 ϵ_{it} 는 오차항으로 평균이 0이고 분산이 σ_ϵ^2 인 정규분포를 따른다. μ_t 는 t 시점의 로그가격지수를 나타내고, β_i 는 자산고정효과를 나타낸다. 여기서 식 (1)은 시장의 종단적 가치 요소와 횡단적인 자산 특성의 가치에 대한 요소로 구성된다. 반복매매모형은 식 (2)와 같이 '동일 부동산' 가정을 따르는 반복 거래쌍의 두 번째 거래시점의 거래가격 모형에서 첫 번째 거래시점의 거래가격 모형을 차분한 형태로 구성된다.

$$\begin{aligned} \ln P_{is} - \ln P_{if} &= \mu_s - \mu_f + \epsilon_{is} - \epsilon_{if} \\ &= \delta_{f+1} + \dots + \delta_s + \epsilon_{ist}, \quad \epsilon_{ist} \sim N(0, 2\sigma_\epsilon^2) \end{aligned} \quad (2)$$

식 (2)의 좌변은 부동산 i 의 두 번째 거래 시점(s) 로그거래가격에서 첫 번째 거래 시점(f)의 로그거래가격을 뺀 것으로 거래쌍의 두 매매 시점 사이의 누적 로그가격변동률을 의미한다. 이는 두 번째 거래시점의 로그가격지수 μ_s 에서 첫 번째 로그가격지수 μ_f 를 뺀 형태로 구성이 되며, 직전 시점 대비 변동률을 나타내는 δ_t 의 $f+1$ 시점부터 s 시점까지의 합으로 달리 표현할 수 있다. $\epsilon_{is} - \epsilon_{if}$ 으로 나타내어지는 오차항은 평균이 0이고 분산이 $2\sigma_\epsilon^2$ 인 정규분포를 따르는 ϵ_{ist} 로 표현 가능하다.

식 (2)는 Bailey et al.(1963)의 가장 기본적인 반복매매모형의 형태로 이후에 Case and Shiller(1987), Goetzmann(1992) 등의 연구들을 통해 다양한 형태로 발전해왔다. 그러나 본 연구에서는 식 (2)를 가장 기본적인 반복매매모형의 형태로 활용하고 ‘일반 반복매매모형’으로 기술한다.

2. 확률적 추세 반복매매모형

Goetzmann(1992)는 과소거래시장을 대상으로 하는 가격지수 산정을 위해 시간고정효과로 포함되었던 로그가격지수 향을 랜덤워크 과정을 따르는 형태로 구성한 형태의 모형을 제안하였다.

$$\mu_{t+1} = \mu_t + \kappa_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2) \quad (3)$$

$$\kappa_{t+1} = \kappa_t + \zeta_t, \quad \zeta_t \sim N(0, \sigma_\zeta^2) \quad (4)$$

여기서 $\sigma_\zeta^2 = 0$ 이고 $\kappa_1 = 0$ 이면, $\mu_{t+1} = \mu_t + \eta_t$ 의 형태로 표현되며 Random Walk(RW) 모형으로 표현한다. $t+1$ 시점의 가격수준 μ_{t+1} 이 직전 시점의 가격수준 μ_t 로부터 확률적인 보행의 결과로 결정된다는 가정을 통해 인접 시점 간의 로그가격지수가 연결된 모형이 구성됨으로써 보다 안정적인 지수의 산정이 가능해진다. Goetzmann(1992)는 여기에 drift라고 하는 모든 시점에서 동일한 값을 가지는 이동량을 추가적으로 포함한 Random Walk with Drift(RWD) 모형을 제안하였다($\sigma_\zeta^2 = 0$). 여기서 이동량(drift)을 나타내는 κ 는

경사(slope)의 의미를 가지는 것으로 해석되며 장기적인 가격변동 추세를 반영하는 것으로 이해할 수 있다 (Francke, 2010).

그러나 가격의 상승과 하락이 반복되는 부동산 시장에서 RWD 모형과 같이 일정량의 가격변동 추세가 모든 시점에서 동일한 상수로 보는 것이 적절한 가정이 아닐 수 있다. 따라서 Francke(2010)에서는 κ 가 식 (4)와 같이 시간에 따라 변화하는 구조를 추가적으로 포함하였다. 이와 같이 식 (3), (4)로 구성된 모형은 Local Linear Trend(LLT) 모형이라고 하며, RWD 모형의 일반화된 형태의 모형으로 볼 수 있다. 최종적으로 LLT 모형은 가격지수의 이동량(drift)이 시점마다 확률적인 과정을 통해 다른 값을 가지도록 하여 각 시점에서의 국지적인 변화를 담을 수 있다.

3. 베이지안 접근법의 활용

베이지안 통계는 18세기 Thomas Bayes에 의해 제안되었고 20세기 후반 제안된 마르코프 확률과정을 이용한 사후분포의 계산 방법으로 발전되었으며 최근 컴퓨터 기술의 발전으로 적극적으로 활용되고 있다. 베이지안 통계는 식 (5)와 같은 베이스 정리에 기반하여 모수의 추정이 이루어진다. $p(\theta)$ 는 사전분포(prior distribution)를 나타내며 연구자가 사전적·경험적으로 알고있는 지식에 기반한 모수의 확률분포를 의미하며 $p(Y|\theta)$ 는 가능도(likelihood)를 나타내는 것으로 모수를 바탕으로 하는 관측결과의 확률을 의미한다. $p(Y)$ 는 주변확률(marginal probability)로 식 (5)의 우변의 분자를 적분하여 계산되며 모수와는 무관한 관측결과 자체의 확률을 나타내는 정규화 상수이다. 최종적으로는 이들의 결합을 통해 $p(\theta|Y)$ 로 나타내어지는 사후분포(posterior distribution)를 얻어내는 것이 베이스 정리의 기본적인 개념이다. 다시 말하면, 사전분포와 가능도를 결합하여 사후분포를 추정하고 이에 따른 모수의 불확실성을 수치화하는 방법이다.

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} \quad (5)$$

베이지안 확률모형을 설정한 후에는 MCMC 알고리즘을 사용하여 매개변수를 추정하고 추정 결과는 사후분포로부터 추출한 표본의 형태로 얻을 수 있다. OLS (Ordinary Least Squares)와 같은 기존의 빈도주의

방법론에서는 데이터 관측이 중심이 되며 데이터를 가장 잘 설명할 수 있는 고정된 모수를 추정하지만, 베이지안 방법론에서는 데이터의 관점에서 모수에 대한 신념의 변화를 분석한다. 따라서 빈도주의 방법론에서는 모수가 한 개의 참 값으로 고정되어 있다고 보지만 베이지안 방법론에서는 불확실성을 갖는 확률분포로 모수를 추정하게 된다. 이러한 특성으로 베이지안 방법론을 이용할 경우 유연한 모형의 적용이 가능하며, 방대한 양의 계산 과정은 빠른 시간 안에 컴퓨터가 수행하기 때문에 연구 모형 자체에 보다 집중할 수 있다. 특히 빈도주의 방법론에서는 데이터의 관측을 중심으로 모수의 추정이 이루어지기 때문에 표본 수가 많을수록 더 좋은 추정결과를 얻을 수 있지만 베이지안 방법론에서는 표본 수에 크게 영향을 받지 않는 장점이 있다.

Goetzmann(1992)는 반복매매모형에 베이지안 접근법을 적용한 방식으로 확률적 추세 모형을 통해 지수 추정의 정확성을 높이는 연구를 진행하였고, 이후에는 Francke(2010), Francke and van de Minne(2017), van de Minne et al.(2020) 등의 연구를 통해 베이지안 확률모형을 활용하는 것이 과소거래시장에 대한 지수 산정에 있어 빈도주의 방법론에 비해 더 좋은 추정결과를 얻어낸다는 것을 확인하였다. 특히 실용적 관점에서 가장 중요한 모형의 선택기준이 될 수 있는 revision을 적지 않게 줄이는 결과를 얻어냈다.

지수 추정을 위해 베이지안 확률 모형에 기반하여 MCMC(Markov Chain Monte Carlo) 알고리즘을 이용하면 제시된 모형의 매개변수 추정 계산을 수행하고, 사후분포에서 추출한 MCMC 표본의 형태로써 가격변동률 추정 결과를 얻을 수 있다. 특히 Goetzmann의 RWD모형과 같은 베이지안 확률모형을 바탕으로 부동산 가격지수를 추정할 경우에 시차 편향을 유발하지 않고 시장의 전환점에서 해당 시기의 가격 대표성을 방해하지 않는 특징을 가질 수 있다(Bokhari and Geltner, 2012). 또한 베이지안 추론을 활용함으로써 모형의 모든 모수에 대한 사후분포를 얻어낼 수 있으며 이를 통해 여러 모형들 사이에 나타나는 추정성능의 차이를 다방면으로 상세하게 비교할 수 있다는 장점이 있다.

본 연구에서는 관련 선행연구를 통해 추정성능의 검증이 이루어진 식 (2), (3), (4)의 모형으로 베이지안 접근법에 기반한 MCMC를 활용하여 국내 부동산 과소

거래시장의 대표적인 예인 서울시 오피스 시장에 대한 가격지수를 추정하고 실무에서의 적용 가능성을 검토하고자 한다.

4. 기존 방법론

확률적 추세 반복매매모형이 과소거래시장에 대한 지수 산정에서 발휘하는 성능에 대한 평가를 위해 기존에 지수 산정에 활용되어져 온 두 가지 방법론을 함께 이용하였다.

먼저 Bailey et al.(1963)가 제안한 동일가중-기하평균 방식의 일반 반복매매모형을 활용하였다. 일반 반복매매모형에 대해서는 기하평균으로 인한 지수의 저평가 문제와 함께 반복거래 사이의 기간이 길어질수록 분산이 증가하게 되는 이분산 문제가 제기되어져 왔다(Case and Shiller, 1987; Goetzmann, 1992; Goetzman and Peng, 2002 등). 그러나 이러한 측면을 고려하여 제안된 가치가중-산술평균 방식의 지수는 고가의 부동산의 가격 움직임이 크게 반영하며, 기하평균 지수가 실제로 저평가된 지수로 볼 수 있는가에 대한 의문점이 남아있다(류강민 외, 2009). 오히려 동일가중-기하평균 지수가 실제로 체감할 수 있는, 시장 전체를 포괄하는 지수로 더 적절할 수 있기 때문에 본 연구에서는 식 (2)의 모형을 더미변수 방식으로 추정하는 Bailey et al.(1963)의 일반 반복매매모형을 활용하였다.

$$\ln P_{is} - \ln P_{if} = \sum_{t=1}^T \beta_t D_{it} + \epsilon_i, \quad \beta_1 = 0 \quad (6)$$

일반 반복매매모형은 식 (6)와 같이 시간더미 D_{it} 로 구성된다. D_{it} 는 $t=s$ 인 경우 1, $t=f$ 인 경우 -1, 그 외에는 0의 값을 가지며 최소제곱법을 통해 추정된 β_t 는 기준시점($t=1$)을 기준으로 한 t 시점의 누적 로그 가격변동률을 나타낸다.

다음으로는 Geltner and Bokhari(2008), Bokhari and Geltner(2012)에서 제안한 2단계 추정법을 활용하였다. 2단계 추정법은 MIT 부동산 연구센터에서 개발되어 Moody's/Real CPPI, Moody's/RCA CPPI 작성에 이용되었다. 2단계 추정법은 한 개 분기씩 미뤄진 4개의 연지수를 산정한 후에 Moore-Penrose의 의사역행렬을 이용하여 분기지수를 추정하는 방식이다.

$$\begin{aligned}
 1+r_{CY} &= (1+r_{t,Q1})(1+r_{t,Q2})(1+r_{t,Q3})(1+r_{t,Q4}) \\
 1+r_{FYM} &= \frac{(1+r_{t,Q2})(1+r_{t,Q3})(1+r_{t,Q4})}{(1+r_{t+1,Q1})} \\
 1+r_{FYJ} &= \frac{(1+r_{t,Q3})(1+r_{t,Q4})(1+r_{t+1,Q1})}{(1+r_{t+1,Q2})} \\
 1+r_{FYS} &= \frac{(1+r_{t,Q4})(1+r_{t+1,Q1})(1+r_{t+1,Q2})}{(1+r_{t+1,Q3})} \quad (7)
 \end{aligned}$$

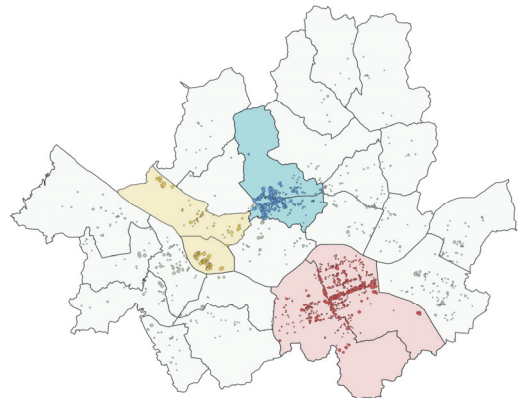
먼저 식 (7)과 같이 t 년도의 1~4분기를 한 시점으로 하는 CY 지수부터 t 년도 4분기~ $t+1$ 년도 3분기를 한 시점으로 하는 FYS 지수까지 총 4개의 set를 구성한다. 연지수를 산정할 경우 한 시점의 가격지수 추정에 이용하는 자료 수가 늘어나면서 비교적 안정적인 지수의 추정이 가능해진다는 장점이 있다.

두 번째 단계로 식 (7)을 통해 분기 가격변동률을 얻어내야 하는데, 식 (7)의 연립방정식은 식이 4개이지만 변수는 7개로 무수한 해가 존재하게 된다. 이를 위해 Moore-Penrose의 의사역행렬(Pseudo Inverse)를 이용하여 한 개의 해를 구하는 방식을 취한다.

측시 경매, 증여 혹은 토지를 제외한 건물만 거래된 경우와 재건축, 용도변경 등의 이유로 반복 대상이 다른 경우에 대해서는 제외하였고, 특히 선매입 가격은 거래시점이 아닌 미래의 잠재가격의 성격을 가지고 있기 때문에(류강민·송기욱, 2020) 준공 전 4개월 이상 선매입 거래건도 분석 대상에서 제외하였다. 이러한 방식으로 구축한 실거래 자료의 대상 오피스 빌딩 분포를 지도에 표기하면 <그림 1>와 같다.

분석대상 범위 내에서 구축한 실거래가 자료는 총 2,367건으로 기초통계량은 <표 1>과 같으며 분기 당 평균 28건의 자료가 구축되었다. 서울 오피스 실거래가 지수 산정에 관한 선행연구들과 비교하면 비교적 많은 거래 자료가 구축되었다.¹⁾ 눈에 띄는 점은 2000년부터 2005년 사이 연평균 거래자료 수는 68건인데 반해 2006년 이후 연평균 거래자료 수는 129건으로 두 배 가까이 차이가 났는데, 이는 부동산거래 신고제도 도입 이전 자료의 조사에 한계가 있는 것이 그 이유로 보인다.

<그림 1> 분석대상 오피스 빌딩 분포



주1: 구축한 실거래자료의 주소 정보를 지오코딩하여 맵핑
 주2: CBD(종로구, 중구), GBD(강남구, 서초구), YBD(마포구, 여의도동), ETC(기타지역)으로 권역 구분하여 표현

IV. 분석의 틀

1. 분석자료

1) 분석대상

지수산정을 위해 2000년 1분기부터 2021년 1분기까지의 연면적 500평 이상 오피스 빌딩 실거래가 자료를 구축하였다. 실거래가 자료의 구축을 위해 먼저 건축물대장 층별개요 자료를 활용하여 전체 연면적 중 50% 이상이 업무시설 용도로 이용되고 있는 오피스 재고를 연도별로 구축하였다. 이후 국토교통부 실거래가 공개시스템, 대법원 인터넷등기소, 금융감독원 전자공시시스템 등을 통해 주소 정보와 매칭하여 최종적으로 주소 정보, 빌딩 연면적, 거래일자, 거래가격, 거래 연면적 등 반복매매모형에 필요한 변수들을 포함한 실거래가 자료를 구축하였다. 그러나 실거래 자료 구

2) 이상치 제거

자료 기입의 오류로 인해 가격이나 면적이 잘못 기입되어 있는 경우 등 다양한 이유로 발생할 수 있는

1) 분기별로 최성호 외(2010)은 22건, 정유신·이기영(2010)은 10건, 류강민 외(2011)은 10건, 황규완·손재영(2017)은 4.5건, 류강민·송기욱(2020)은 19건으로 이들 연구에서 서울 오피스 실거래가 지수 작성을 위해 이용된 자료에 비해 본 연구에서 구축한 자료 수가 더 많은 것으로 나타났다. 앞의 2건의 연구는 전체자료를, 그 외 연구는 반복거래자료만을 포함한 자료 수이고, 연구별로 대상으로 하는 오피스 빌딩 면적, 거래 유형 제외 여부 등에 따라 차이가 있을 수 있겠으나 이를 고려하더라도 선행연구들에 비해서는 많은 자료가 구축된 것으로 판단된다. 다만, 본 연구의 분기별 28건이라는 수치는 반복거래가 이루어지지 않은 자료도 모두 포함한 전체자료의 수임을 고려할 때 여전히 기존의 방법론으로는 안정적인 지수 산정에 한계가 있어 연구 목적을 해치지 않는 자료셋인 것으로 판단된다.

이상치는 안정적인 지수의 작성을 방해하는 요소가 된다. 따라서 사전적으로 가격 이상치를 식별하여 제거할 필요가 있다. 현재 한국부동산원에서는 실거래가격 지수 산정을 위한 실거래 신고자료 이상치 처리를 위해 진단가격 대비 신고가격 비율을 시도별·월별로 그룹화 하여 그룹별 Box-Plot 방법을 통하여 상한(Q3) 및 하한(Q1)에서 사분위범위(IQR)의 3배 이상 벗어난 자료들을 제거하고 있다. 송영선 외(2020)에서는 log 단위가격을 기준으로 상한(Q3), 하한(Q1) 으로부터 사분위범위(IQR)의 1.5배 이상 벗어난 자료들을 이상치로 식별하여 제거하였다.

이외에 상·하위 2% 내외의 자료를 일괄적으로 제외하는 등의 다양한 방법이 있으나 이는 실거래 자료의 실제 분포가 고려되지 않은 방식으로 본 연구에서는 Box-plot을 이용하여 이상치를 식별하였다. 다만, 오피스 시장의 거래 자료는 기본적으로 그 수가 적고

많은 자료들을 제거할 경우 지수 안정성에 문제가 생길 수 있기 때문에 2단계로 나누어 이상치 식별 작업을 진행하였다.

Box-plot 방식을 위해 기본적으로 매 시점에서 기준이 되는 가격 대표값과 사분위범위의 산출이 필요한데 해당 시점의 가격 자료만을 이용할 경우 시점별로 가격 편차가 크게 나타날 수 있고 거래자료 수가 부족한 경우 문제가 발생할 수 있다. 따라서 서울시 오피스 시장 권역별로 그룹화하여²⁾ 가격 대표값은 과거 12개 분기의 중위값으로, 사분위범위는 과거 4개 분기의 사분위 범위 값으로 산출하였다. 또한 낮은 가격에 비해 높은 가격이 중위가격으로부터 멀리 떨어져 있는 관계로 발생 가능한 비대칭 분포의 문제를 고려하여 단위 거래가격의 log 변환값을 이용하였다.

먼저 Q3, Q1로부터 IQR의 3배 이상 벗어난 자료 17건을 사전제거하였다. 이후에 다시 가격 대표값과

<표 1> 분석자료 기초통계량

거래년도	총 거래가격 (만원)	평균 거래가격 (만원/평)					총 거래면적 (만 평)	자료수
		평균	중위	표준편차	최소	최대		
2000	23,990.6	655.1	583.3	218.4	319.1	1,161.0	34.8	62
2001	22,363.6	606.0	606.4	193.6	271.4	1,180.0	32.1	59
2002	23,772.1	630.1	597.3	197.7	198.1	1,111.9	35.7	78
2003	24,522.3	771.1	690.2	592.2	363.2	5,007.0	34.0	76
2004	30,021.7	786.5	796.1	194.4	250.4	1,446.4	31.9	54
2005	25,317.5	899.1	788.5	544.9	294.0	4,322.1	26.0	77
2006	27,501.7	965.1	906.3	451.4	138.1	2,849.5	30.9	154
2007	49,006.7	1,174.2	1,053.1	635.2	328.9	5,105.3	40.3	171
2008	50,560.0	1,301.3	1,182.9	692.8	463.8	5,237.8	36.7	104
2009	57,168.8	1,247.4	1,200.1	673.0	421.1	6,261.5	42.1	146
2010	41,885.8	1,290.5	1,160.1	571.3	499.6	3,048.6	29.6	92
2011	56,013.8	1,456.5	1,320.2	749.4	313.5	5,049.2	36.5	121
2012	60,066.8	1,387.9	1,344.9	614.7	383.7	4,130.0	43.0	108
2013	61,702.1	1,396.5	1,248.8	790.0	243.4	5,743.5	40.2	119
2014	67,010.4	1,455.0	1,328.2	789.7	63.0	6,208.4	40.5	116
2015	50,507.7	1,485.9	1,263.8	941.7	152.8	7,601.8	32.2	121
2016	102,437.9	1,793.5	1,641.1	1,266.5	344.4	13,499.7	56.4	132
2017	74,133.4	1,678.7	1,525.9	839.5	597.1	4,640.9	43.5	135
2018	105,314.7	1,782.1	1,632.7	855.9	275.0	6,292.8	53.8	136
2019	108,233.8	2,032.4	1,961.9	998.4	275.0	7,563.5	51.9	141
2020	130,170.6	2,231.0	2,136.1	1,072.2	682.0	7,675.9	58.3	146
2021	18,837.2	2,014.0	1,834.1	796.0	552.2	3,639.3	8.0	19
합계	1,210,539.1						838.3	2,367

2) 권역별로 분석기간 전체 단위거래가격의 평균을 계산해보면 적게는 12%에서 많게는 34%까지 권역별 가격의 평균 및 분산에 뚜렷한 차이가 나타나는 것을 확인하였고, 이러한 점을 고려하여 권역별로 다른 기준을 적용하여 이상치 식별 기준을 산출하였다.

사분위범위를 산출하여 Q3, Q1으로부터 IQR의 1.5배 이상 벗어난 93건의 자료를 이상치 후보군으로 설정하였다. 이 중 반복 거래쌍을 구성하는 53건의 자료에 대해서는 한 건씩 자료에서 제거하여 지수를 산정해보고, 사전제거 자료의 지수와 함께 총 54개의 지수 형태 및 추정성능(표준오차)의 변화를 비교·검토했다. 최종적으로는 이상치 후보군 중 8건의 자료를 이상치로 분류하여 제거하였다.

3) 동일부동산 가정

지수 산정을 위해 반복매매모형을 활용하기 때문에 동일 부동산 가정을 통해 반복거래쌍을 구성하여야 한다. 동일한 부동산으로 가정하기 위해서는 기본적으로 동일한 지번에 위치해야 하며 층축, 철거 등으로 연면적이 10%이상 변화하거나 재건축된 경우에는 동일 부동산이 아닌 것으로 가정하였다. 지번거래의 경우에는 거래 면적이 다르더라도 동일 부동산 가정을 충족시킨다면 반복 거래쌍의 구성이 가능하다.

최종적으로 반복거래쌍을 구성하지 못한 자료는 843건으로 <표 2>와 같이 1,499건의 자료만이 지수산정에 이용되었다. 2번 거래가 되어서 1개의 거래쌍을 구성한 자료 수는 782건으로 가장 많았고, 24건의 거래는 6회 반복거래가 발생한 자료로 각각 5개의 거래쌍을 구성하였다.

<표 2> 반복거래횟수 별 이용자료 수

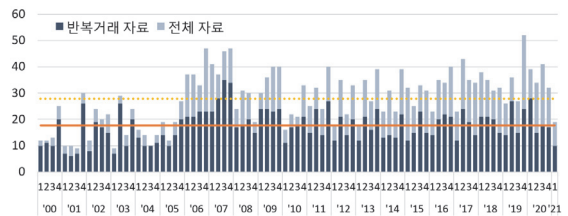
반복거래횟수	거래건 수	거래쌍 수	동일빌딩 수
1	843	-	843
2	782	391	391
3	411	274	137
4	192	144	48
5	90	72	18
6	24	90	4
합계	2,342	901	1,441

구축한 전체자료 2,367건과 지수산정에 이용된 1,499건의 분기별 거래건수는 <그림 2>와 같이 나타났다. 구축한 전체자료 2,367건 대비 자료이용률은 63.3%로 분기당 평균 약 18건의 자료가 지수산정에 이용되었다. <그림 2>의 진한 실선은 반복거래쌍을 구성한 자료의 분기별 평균 자료 수를 나타내는 것으로 2008년 금융위기 직전인 2007년에 발생한 거래건에

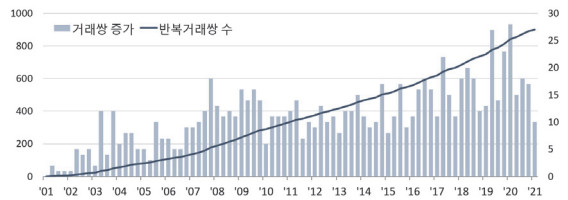
대해서는 평균적인 수준을 웃도는 거래 자료가 이용되었고, 부동산거래신고제도 도입 이전인 2000년부터 2005년 사이에는 기본적인 자료 수 부족으로 평균보다 낮은 수준의 자료 수가 이용되었다. 다만, 해당 6년간의 평균 자료이용률은 78.5%로 상대적으로 많은 거래사례들이 이후에 또다시 거래가 발생한 것으로 나타났다.

최초로 반복 거래쌍이 구성된 2001년부터 각 시점까지 구성된 총 반복 거래쌍 수와 해당 분기에 추가된 거래쌍 수는 <그림 3>과 같이 나타났다. 초기에는 누적된 거래자료가 부족한 관계로 반복 거래쌍 수도 적었으나 2007년 이후로는 매 분기 새로운 자료가 추가될 때마다 평균 14건 정도의 거래쌍이 추가되었다. 특히 기본적으로는 해당 시점의 거래자료 수에 비례하여 거래쌍 수도 추가되었는데, 최근시점으로 가까워질수록 추가되는 거래쌍 수가 조금씩 늘어나는 경향이 나타났다.

<그림 2> 분기별 이용자료 수



<그림 3> 지수작성 기간별 반복 거래쌍 수



2. 분석 틀

통계패키지 프로그램으로 기존 방법론을 활용한 지수의 추정에는 SAS, 베이지안 확률모형을 활용한 지수의 추정에는 R을 이용하였다.

베이지안 확률모형을 도입한 지수를 추정하기 위한 MCMC 시뮬레이션은 'rstan' 패키지를 활용하였다. 이를 위한 시뮬레이션 툴로 WinBUGS, JAGS, Stan

등을 이용할 수 있는데, 본 연구에서는 RStan을 이용하였다. Stan은 계산 알고리즘에 HMC(Hamiltonian Monte Carlo)의 한 가지 버전인 NUTS(No-U-Turn Sampler; Hoffman and Gelman, 2014)를 사용하고 있는데, 매개변수가 많은 경우에도 효과적으로 표본을 추출할 수 있다는 장점이 있으며 Gibbs sampler, Metropolis Hastings algorithm과 비교하였을 때 적은 횟수의 계산 반복 횟수로도 유의한 결과를 얻을 수 있고 계산에 소요되는 시간을 크게 줄여 보다 효율적으로 결과를 산출할 수 있다. 본 연구에서는 Stan 설정을 chains=4, iter=3,000, warmup=1,000으로 하여 총 8,000회의 샘플링을 통해 사후분포를 추정하였다.

3. 모형 세부사항

Stan을 통한 베이지안 추정을 위해서는 지수 추정에 이용하고자 하는 모형 식과 함께 모형의 확률분포, 추정 모수의 사전분포를 설정하게 된다.

반복매매모형의 우도 확률분포는 가장 기본적으로는 정규분포로 나타낼 수 있으나 정규분포는 평균 μ 로부터 $\pm 2\sigma$ 만큼 떨어지게 되면 확률밀도함수의 값이 급격하게 떨어져 바깥값(outlier)이 존재할 경우 추정 결과가 늘어질 수 있다. 이런 경우에 정규분포 대신에 꼬리가 긴 t-분포를 도입하는 방안을 고려할 수 있다.

표본 수가 부족할수록 바깥으로의 분산이 더 커지는 경향이 있기 때문에 t-분포와 같이 꼬리가 긴 분포를 적용하는 경우 바깥값(outlier)을 보다 잘 반영할 수 있다. 특히 베이지안 추정에서 꼬리가 긴 분포를 활용함으로써 사전분포와 데이터 간의 충돌이 처리되는 방식을 통제할 수 있고, 더 나은 추론과 표본 외 예측을 수행할 수 있다(McElreath 2020, p.234, p.370). 관련 연구들을 통해 t-분포를 적용함으로써 보다 나은 추정 결과가 산출된다는 점이 확인되었고(Francke and van de Minne, 2017; van de Minne et al., 2020), 본 연구에서도 정규분포와 t-분포를 모두 시도하여 비교한 결과 t-분포를 도입하는 경우에 더 좋은 결과를 얻어낼 수 있었다.³⁾

추정 모수들의 사전분포의 선택을 위해서는 선행 연구를 참고하고 몇 가지의 사전 분포를 직접 비교하였다. 최종적으로는 식 (2), (3), (4)의 오차항의 분산에

대해서는 코시 분포를 사전분포로 선택하였고, 그 외 모수들에 대해서는 가장 기본적인 정규분포를, t-분포의 ν 에 대해서는 지수분포를 사전분포로 선택하였다.

4. 지수 추정성능 평가

다양한 방법론으로 산출되는 지수의 성능을 평가하고 비교하기 위해 추정결과를 바탕으로 몇 가지 평가 지표를 활용하였다. 지수의 추정성능 평가를 위해 다양한 지표들이 이용되어져 왔다. 이창무 외(2013)에서는 지수화된 표준오차의 평균, 안정성 지수, 신호-잡음을 이용하였고, Guo et al.(2014)에서는 지수의 기간변동률의 표준편차로 나타내어지는 변동성과 1차 자기상관을 비교하였다. van de Minne et al.(2020)에서는 모형의 예측력을 평가할 수 있는 WAIC, 잡음(noise, σ_e), 변동성(volatility), 1차 자기상관, 그리고 지수변화(revision)에 대한 일록슨 부호-순위검정, 표준편차, 절대평균 등을 활용하여 모형별 추정성능을 평가하고 비교하였다.

본 연구에서는 통계적 관점에서의 추정결과의 정확성을 검증하고 실무적 관점에서의 활용성과 연구 목적에 부합하는 지수 안정성을 평가할 수 있는 다음과 같은 4가지의 평가 지표를 활용하였다.

1) 통계적 신뢰도

지수의 통계적인 신뢰도를 평가하기 위한 가장 기초적인 평가지표는 추정 값의 표준오차이다. 그러나 추정치의 수준에 따라 표준오차 값이 큰 차이를 나타내기 때문에 일반적으로는 표준편차를 평균으로 나누어 추정치의 평균적인 수준을 통제하고 단위에 영향을 받지 않는 변이계수(CV, Coefficient Variance)가 이용된다. 그러나 기준시점 설정에 따라 지수값이 달라질 수 있고, 지수 값이 커질수록 동일한 표준오차를 가진 경우에 변이계수가 감소하는 문제가 발생할 수 있다(이창무 외, 2013). 따라서 본 연구에서는 식 (8)과 같이 지수화된 표준오차의 평균값(Mean of Standard Error of Index, MSEI)으로 통계적 신뢰도를 평가하였다.

3) 본 논문에서 두 가지 확률분포를 시도한 결과를 비교하는 내용을 담지는 않았으나, 이어서 소개할 지수의 추정성능을 평가하는 방식으로 비교한 결과 평가 지표가 모두 t-분포를 도입하는 경우에 더 좋은 추정성능을 보이는 방향으로 분석되었다.

$$MSEI = \sum_{t=2}^T e^{se_t} \times 100 / T \approx \sum_{t=2}^T se_t \times 100 / T \quad (8)$$

2) 지수 안정성

본 연구는 과소거래시장에 대한 안정적인 지수 산정을 목적으로 하는 만큼 지수의 안정성 개선 정도에 대한 평가가 필수적이다. 지수의 안정성은 분기 변동률의 표준편차로 나타내어지는 변동성 등과 같은 지표들로 확인할 수 있으나 이창무 외(2013)은 지수 추세의 안정성을 평가하기 위해 보다 직관적으로 쉽게 판단할 수 있는 안정성 지수(SI, Stability Index)를 개발하였다. 기본적인 개념은 지수의 초기 시점과 마지막 시점 간 직선거리 대비 지수 그래프의 총 길이의 비를 통해 나타내어 시각적인 관측을 통한 판단을 계량화한 것이다. 따라서 지수가 초기 시점과 마지막 시점을 연결하는 직선의 형태인 경우 SI 값은 1이 되고 그 직선에서 많이 벗어난 형태일수록 0에 가까운 작은 값이 산출된다.

$t = 1$ 부터 $t = T$ 시점까지 작성된 지수의 t 시점 지수 값을 I_t 라고 할 때, 안정성 지수 SI는 다음과 같은 식(9)을 통해 산정된다.

$$SI = \frac{\sqrt{(T-1)^2 + (I_T - I_1)^2}}{\sum_{t=2}^T \sqrt{1^2 + (I_t - I_{t-1})^2}} \quad (9)$$

SI는 0에서 1 사이의 값을 가진다. SI 값이 클수록 지수 안정성이 높은 것으로 판단할 수 있으나 그 값이 일정 수준 이상으로 크게 나타날 경우 지나치게 평활화되었다는 것을 의미하기도 하며(송영선 외, 2020), 부동산 유형에 따라 시장 가격의 변동 추이가 다를 경우 동일한 수준의 안정성을 보여주는 SI 값이 다를 수 있기 때문에 SI 지표를 이용한 지수의 비교 및 평가에 주의할 필요가 있다.

3) 지수 변화(Index Revision)

반복매매모형의 특성상 새로운 분기에 거래된 자료가 추가되는 경우 과거에 거래가 이루어졌으나 반복거래쌍을 구성하지 못했던 자료가 새로이 추가됨으로써 과거에 추정되었던 지수 값이 변화하여 Revision이 발생한다. Revision이 발생하면서 추정된 지수의 정

확도가 더 올라가기 때문에 이를 허용하는 것이 바람직하겠지만, 그 변화의 정도가 클 경우 실무적 활용성이 떨어질 수 있다. Revision이 작은 지수 산정 모형이 더 나은 성능을 발휘한다고 볼 수 있으며 이는 지수의 실용성에 대한 평가 기준으로 활용할 수 있고, 지수 추정치의 정확도를 반영하기도 한다. 따라서 지수의 Revision은 지수 산정 모형의 중요한 선택기준이 되어야 한다.

van de Minne et al.(2020)에서는 Revision 비교를 위해 지수값과 변동률의 두 가지 기준을 활용하였고, Revision의 절대평균, 표준편차, 최대값, 임계치보다 큰 Revision 발생 횟수, 일콕순 부호-순위 검정과 같은 다양한 지표를 통해 비교하였다. 이러한 지표를 모두 이용할 경우 다양한 측면에서의 Revision 발생 경향을 판단할 수 있으나 본 연구에서는 직관적으로 간편하게 비교·평가가 가능한 절대평균을 이용하였고, 지수값을 기준으로 하는 경우 지수값의 수준에 따라 편차가 발생할 수 있기 때문에 변동률을 기준으로 하여 다음과 같은 RI(Revision Index)를 Revision에 대한 평가 지표로 활용하였다.

<그림 4> 지수추정 예시

		지수 작성 시점(s)					
		0	-1	-2	-3	-4	...
지수 시점 (n)	∴	∴	∴	∴	∴	∴	∴
	r_{04}	r_{13}	r_{22}	r_{31}	r_{40}		
	r_{03}	r_{12}	r_{21}	r_{30}			
	r_{02}	r_{11}	r_{20}				
	r_{01}	r_{10}					
	r_{00}						

<그림 4>에서 가로축의 지수 작성 시점 s 는 지수가 작성된 시점을, 세로축의 지수 시점 n 은 작성된 지수에서 지수 값이 어느 시점의 값인지를 나타낸다. 다시 말하면 2021년 1분기까지 자료가 구축되어 가장 최근의 지수가 작성되었다면 2021년 1분기 지수가 $s = 0$ 에 해당하며 2020년 4분기 말에 작성된 지수는 $s = 1$, 2020년 3분기 말에 작성된 지수는 $s = 2$ 에 해당하며 이를 최근 지수 작성 시점으로부터 몇 번째 전 시점에 작성된 지수인가를 표현하는 의미로 -1, -2와 같이 표기하였다. 지수 시점 n 은 작성된 지수 내에서 해당 지수값이 가장 최근 시점으로부터 몇 번째 전 시점의

지수값인지를 나타낸다. 2021년 1분기 말에 작성된 지수라면 2020년 4분기 지수값이 가장 최근의 지수 시점이 되며 2020년 3분기 지수값은 $n=1$, 2020년 2분기 지수값은 $n=2$ 에 해당한다. 따라서 현재 작성된 가장 최근의 지수를 기준으로 하여 s 번째 전 시점에 작성된 지수에서 지수값의 가장 최근시점으로부터 n 번째 전 시점의 지수값은 I_{sn} 과 같이 표현할 수 있다. <그림 4>는 이러한 개념을 바탕으로 변동률을 기준으로 한 RI를 산정하기 위해 지수값 I_{sn} 을 분기변동률 r_{sn} 으로 대체하여 표현한 것이다. 추정된 지수의 Revision 정도를 판단하는 지표로서 RI_{sn} 은 식 (10)과 같이 계산할 수 있다.

$$\frac{\sum_{t=1}^n \sum_{p=1}^s |r_{p-1,t} - r_{p,t-1}|}{s \times n} \quad (10)$$

<그림 5>를 예로 들어 2021년 1분기 말에 지수를 작성하고 RI_{54} 를 산출하고자 한다면, $s=5$ (노란색 박스 4개)이기 때문에 2021년 1분기로부터 5번째 전 시점인 2019년 4분기에 작성된 지수까지 이용하여 직전 시점 산정된 지수의 동일시점 지수값과의 차이를 계산하며, $n=4$ (빨간색 화살표 4개)이기 때문에 이러한 지수값 차이 각각을 해당 지수에서의 가장 최근 시점으로부터 4개 시점 이전 지수값까지를 계산한다. 이렇게 하면 지수값의 차이의 절대값은 총 $s \times n = 20$ 개가 산출되며 이들의 평균을 계산하여 RI_{54} 를 구한다.

<그림 5> Revision 지표 산출 예시

	-21.Q1	-20.Q4	-20.Q3	-20.Q2	-20.Q1	-19.Q4
18.Q3	327.2	326.2	328.0	327.2	326.5	327.7
18.Q4	328.5	328.7	329.2	327.6	326.2	329.3
19.Q1	335.1	333.8	333.4	332.3	329.8	333.0
19.Q2	349.9	348.4	348.1	347.2	345.2	345.5
19.Q3	353.2	351.6	351.2	350.0	348.6	346.6
19.Q4	356.9	354.8	354.6	352.6	350.8	346.1
20.Q1	369.1	366.8	366.0	361.9	358.0	
20.Q2	380.4	376.6	375.1	366.3		
20.Q3	392.9	387.2	384.9			
20.Q4	400.6	389.4				
21.Q1	411.8					

4) 모형 예측력

Widely Applicable (혹은 Watanabe Akaike) Information Criterion이라고 하는 WAIC 정보기준은 기존의 DIC(Deviance Information Criterion)을 베이지안 모형에 맞추어 발전시킨 정보기준으로, 모수 값의 사후 시뮬레이션에서 평가된 가능도를 사용하여 적합한 베이지안 모형에서 점별 표본 외 예측의 정확도를 추정하는 방법의 정보기준이다(Watanabe, 2010; Vehtari et al., 2017). WAIC는 사후분포의 로그-가능도의 평균값(pWAIC)에서 과적합 페널티 항으로서의 로그-가능도의 분산(dWAIC)을 뺀 값에 -2를 곱하여 계산되며 식 (11)⁴⁾과 같다.

$$WAIC(y|\theta) = -2(lppd - p_{WAIC}) \quad (11)$$

WAIC는 좋은 예측을 하는 모형을 식별하기 위한 기준으로 산출된 WAIC 값이 작을수록 모형의 예측력이 더 좋다고 볼 수 있다. 페널티 항이 포함되어 WAIC는 예측값의 과적합 경향을 측정해 모형의 설계나 통계적 추론의 작동을 이해하는데 도움이 된다. 다만, 인과관계를 잘 설명한다는 것을 의미하지는 않는다는 점에 유의해야 한다.

V. 분석 결과

1. 지수산정 결과

MCMC 샘플링 결과 모든 모수에 대하여 $\hat{R} < 1.1$ 으로 MCMC가 RW, RWD, LLT의 세 개 모형에서 모두 수렴한 것으로 나타났고, 효과적인 MCMC 표본 수의 거의 대부분이 샘플링 표본 수보다 높게 나타나 샘플링이 잘 이루어진 것으로 판단할 수 있었다. <그림 4>는 세 가지 모형을 통해 산정한 지수와 각각의 직전분기 대비 변동률을 그래프로 나타낸 것이다.

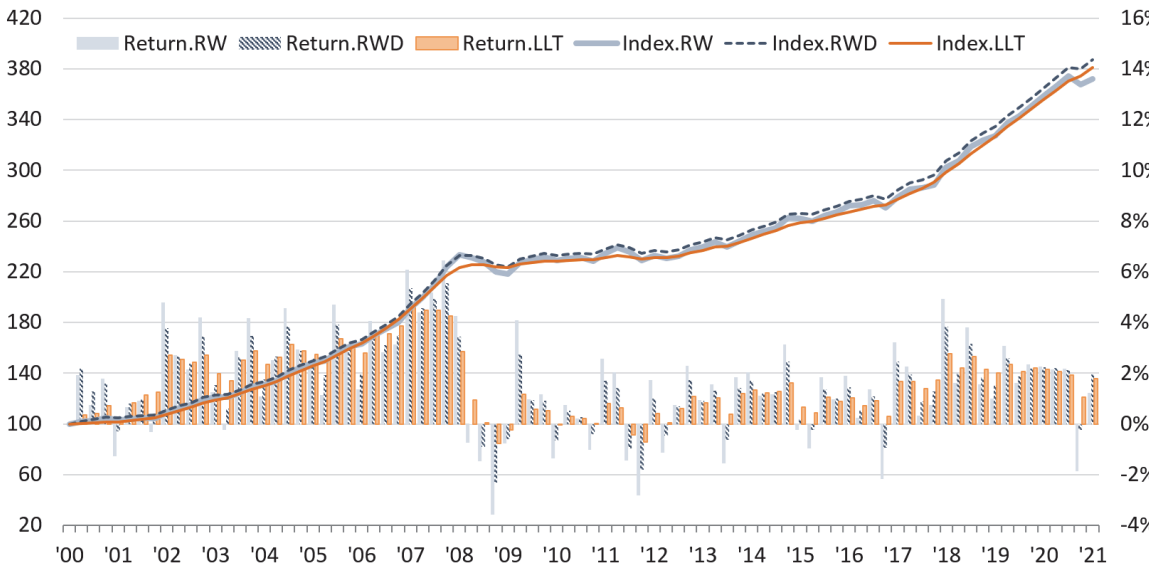
<그림 6>의 전반적인 그래프의 형태를 보면 세 가지 모형 모두 유사한 장기적인 추세를 보여주지만, 모형

4) 식 (11)에서 lppd는 로그점적예측밀도(log pointwise predictive density)로 $lppd(y, \theta) = \sum_i \log \frac{1}{S} \sum_s p(y_i | \theta_s)$ 의 식을 통해 표본 s 의 i 번째 관측치의 확률밀도를 계산하고 평균을 구하여 로그를 취하는 방식으로 계산되며 lppd 값이 높을수록 정확한 모형임을 의미한다. p_{WAIC} 는 사후예측의 분산에 비례하는 페널티 항으로 유효한 모수의 숫자의 의미를 가지며 $p_{WAIC} = \sum_i \text{var}_\theta \log p(y_i | \theta)$ 의 식을 통해 계산된다. 이외에 관련된 상세한 개념은 Watanabe(2010), Vehtari et al.(2017)을 참조.

의 특성이 잘 반영되어 RW모형은 다른 모형의 추세에 비해 분기별로 다소 독립적인 경향성을 보이고 있다. 반면에 LLT 모형은 이전 시점 지수와의 연속성이 강하여 평활화가 강하게 이루어진 지수 추이를 나타낸다.

이러한 점은 변동률을 나타낸 막대그래프에서도 잘 드러나는데 RW 모형의 변동성이 가장 크고 LLT 모형의 변동성이 가장 작은 것으로 나타났다.

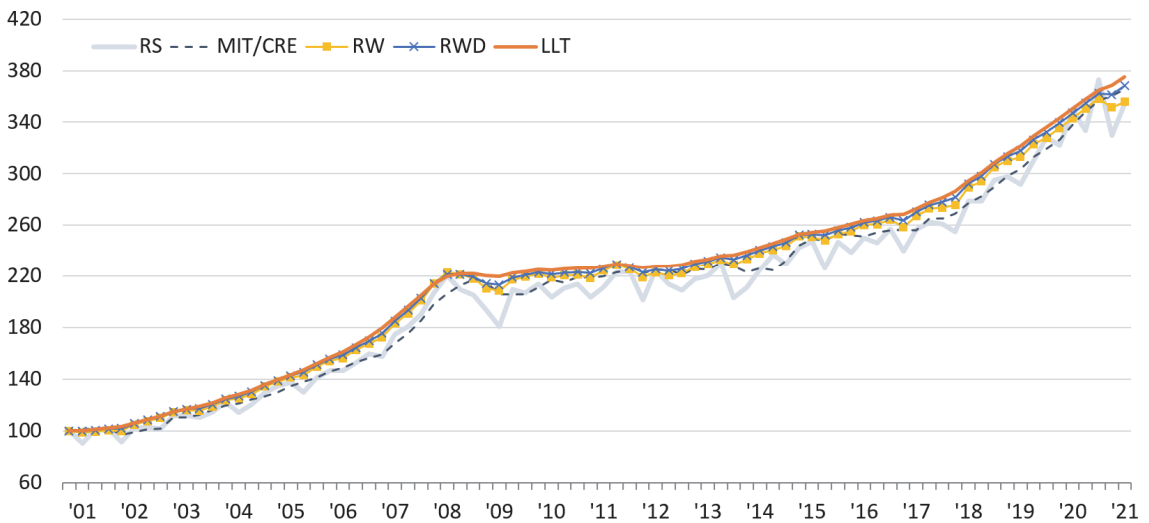
<그림 6> RW·RWD·LLT 지수 및 분기 변동률



주1: 2000.1Q=100.0

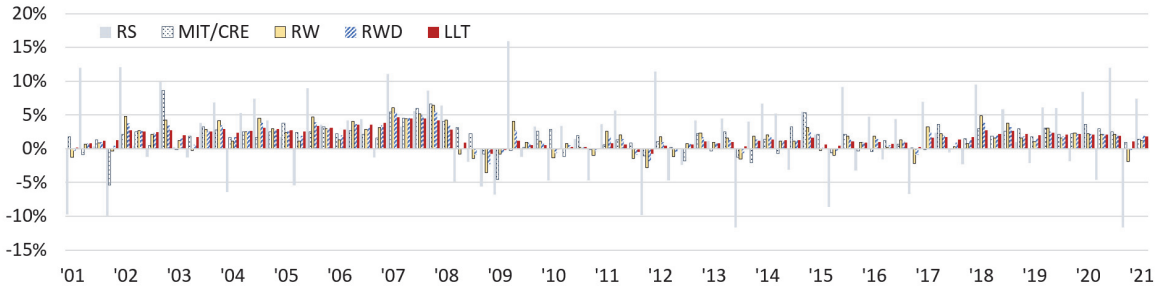
주2: Index는 기준시점을 100으로 한 가격지수를, Return은 가격지수의 직전분기 대비 변동률을 나타냄

<그림 7> 기존 방법론과의 지수 비교



주: 모형 간 직접적인 비교를 위해 기준시점을 2000년 4분기로 변경하여 나타냄(2000.4Q=100.0)

<그림 8> 기존 방법론과의 분기 변동률 비교



<그림 7>은 기존의 지수산정 방법론으로 소개한 일반반복매매모형(RS), 2단계 추정법(MIT/CRE)으로 산정한 지수와 베이지안 추세 모형 세 가지를 비교한 그래프이다. 가장 먼저 눈에 띄는 점은 일반반복매매 지수의 변동성이 눈에 띄게 크게 나타난다는 점이다. 반면에 나머지 4개의 지수는 상대적으로 안정적인 형태를 보이고 있고, 이러한 형태는 <그림 8>의 분기 변동률 막대그래프를 통해서도 뚜렷하게 확인이 가능하다. 다만, MIT/CRE 지수의 경우 가격 변동의 방향이 변화하는 극점을 중심으로 봤을 때, 약 1~2분기 정도 시차가 발생하였다. LLT 지수는 평활화가 특히 더 두드러지게 나타났는데 2008년 3분기부터 2009년 1분기, 2020년 4분기에서 다른 지수에서는 가격 하락세가 관측되는 반면에 LLT 지수에서는 드러나지 않았다는 측면에서 지수의 평활화 정도를 가늠해 볼 수 있다.

2. 지수 비교·평가

이렇게 산정한 지수들이 각각 얼마나 신뢰할 수 있고 실용성 있는지를 비교하고 새로이 시도한 모형이 얼마만큼의 성능을 보이는지를 비교하기 위해서 앞서 소개한 네 가지의 지수 추정성능 평가 지표를 통해 비교하였다.

먼저 통계적 신뢰도의 측면에서는 LLT, RWD, RW, MIT/CRE⁵⁾, RS 모형의 순서로 표준오차가 낮게 산출되어 LLT 모형의 통계적 신뢰도가 가장 높은 것으로 나타났다. 특히 최소제곱법에 의한 지수의 표준오차에

비해 베이지안 추정법에 의한 지수의 표준오차 값이 상당히 작게 나타났다.⁶⁾

지수 안정성의 경우에도 마찬가지로 LLT 지수의 안정성이 가장 높게, RS 지수의 안정성이 가장 낮게 나타났다. 구체적으로는 LLT 지수의 안정성 지수 값이 0.9 이상으로 상당히 크게 나타났으며, MIT/CRE 지수와 RW 지수의 안정성이 비슷한 수준으로, RS 지수는 이들의 절반도 안되는 수준으로 나타났다. 다만, 그래프의 형태로 살펴봤듯이 LLT 지수의 경우에는 시장 추세의 변화 포인트를 잘 식별하지 못한다고 볼 수 있을 만큼 지나친 평활화가 이루어진 것으로 판단된다. 이러한 경우 2008년의 금융위기 이후 일시적인 부동산 가격의 하락 상황과 같이 가격변동 추세가 상승 또는 하락하는 방향으로 변화하는 시장의 전환점을 제대로 식별해내지 못하는 문제가 있어 지수로서의 중요한 역할을 하지 못할 가능성이 있다.

Revision은 모형별로 조금 다른 양상이 확인되었다. RS 지수의 Revision은 과거 시점(24×24)으로 갈수록 점차 증가하였고, 베이지안 추세 모형은 이와는 반대로 과거 시점으로 갈수록 점차 감소하는 것으로 나타났다. MIT/CRE 지수의 경우에는 그 경향이 특이하게 나타났는데, 가장 최근 시점(24×1) Revision이 상당히 작게, 그리고 그 이전 시점은 일정 수준을 유지하는 경향을 보였다. 베이지안 추세 모형의 가장 최근 시점 Revision은 기존 방법론에 비해 모두 높게 나타났으나 장기적인 Revision은 훨씬 낮게 나타났다. 베이지안 추세 모형의 최근시점 Revision이 높게 나타

5) 2단계 추정법을 통해 산정된 지수는 under-determined된 연립방정식의 형태를 전제로 하기 때문에 전통적인 통계량을 산출하거나 의미를 부여하기 어렵다(황규완·손재영, 2017). 따라서 MIT/CRE 지수의 통계적 신뢰도를 보여준다고 하기에는 한계가 있지만 1단계에서 산정한 연지수들(CY, FYM, FYJ, FYS set)의 표준오차 평균값으로 대체하였다.

6) MCMC의 결과 산출되는 표준오차는 MC(Monte Carlo) 오차를 나타내는 것으로 MCMC를 통해 구해진 사후분포로부터 산출된 표본의 표준편차를 유효한 샘플링 횟수로 나누어 계산된다. 따라서 8,000번의 샘플링을 통해 표준오차의 값이 상당히 작게 산출되었는데, 최소제곱법을 통한 지수 추정 결과와 표준편차를 비교하더라도 베이지안 추세 모형의 표준편차가 더 작게 나타났다.

〈표 3〉 지수 평가 지표 비교

구 분		RS	MIT/CRE	RW	RWD	LLT
MSEI		8.8889	4.2381	0.0479	0.0458	0.0398
SI		0.306	0.759	0.714	0.842	0.942
RI (%p.)	24×24	0.606	0.283	0.212	0.155	0.188
	24×16	0.577	0.271	0.214	0.159	0.135
	24×8	0.547	0.293	0.267	0.202	0.198
	24×4	0.502	0.314	0.364	0.284	0.309
	24×1	0.404	0.028	0.678	0.486	0.481
WAIC		-	-	-211.277	-216.872	-218.750

나는 것에 관해서는 만약 2008년 금융위기와 같은 리스크 상황이 발생할 경우에 LLT 모형과 같은 형태의 구조적 시계열 모형은 가장 최근의 관측치를 의미 있는 이상치(innovative outliers)가 아닌 추가적인 이상치(additive outliers)로 식별하고 시장의 전환점이 되는 포인트를 판별하지 못하게 되는 문제가 제기된 바 있는데(Durbin and Koopman, 2012; van de Minne et al., 2020), 이러한 모형의 특성은 본 연구에서 드러난 최근 시점 Revision에 대한 문제와도 연결된다. 그럼에도 장기적인 Revision이 기존 모형에 비해 상당히 작게 나타난다는 점은 고무적이며, 따라서 차선책으로써 최근 시점의 Revision이 비교적 작게 나타나는 모형이 대안이 될 수 있을 것으로 판단된다.

모형의 예측력을 평가할 수 있는 WAIC는 말 그대로 베이저안 확률 모형이 얼마나 과적합을 피하여 좋은 예측을 할 수 있는가를 판단하는 지표이다. WAIC 값이 낮을수록 예측력이 더 높은 모형을 의미하며 LLT, RWD, RW 모형의 순으로 더 좋은 예측을 하는 모형인 것으로 나타났다. 그러나 이것이 인과관계를 잘 설명하는지를 보여주는 지표는 아니고, 세 가지 모형에서 그 차이가 적게 나타났기 때문에 참고 정도로만 활용 가능할 것으로 보인다. 다만, 기존 방법론과의 비교를 위해 일반 반복매매모형을 MCMC에 적용하여 추정해 본 결과 일반 반복매매모형의 WAIC 값은 -76.825로 베이저안 추세 모형이 일반 반복매매모형에 비해서는 과적합 경향을 피하는 비교적 좋은 예측을 하는 모형이라는 점은 확인할 수 있다.

VI. 결론

현재 국내에서는 오피스 시장에 대한 투자지수 수요에 비해 제대로 된 정보의 제공이 이루어지지 않고 있다. 특히 실거래가격에 기초한 매매가격지수는 자본환원을 산정을 위한 분모 부분의 가치 평가 시 시점 보정을 위한 기초 지수로도 활용 가능성이 있는 지표이다. 그러나 현재 가장 대표적인 상업용부동산 임대동향조사원의 자본수익률은 평활화 및 저평가된 상승률이 시장 상황을 민감하게 반영하지 못하고, 대상 표본의 평가 가치 역시 시장 가격수준을 충분히 반영하지 못하는 문제를 가진다. 민간에서 제공되고 있는 헤도닉 모형에 기반한 실거래가지수는 단기적인 불안정성으로 해당 시점에서의 시장의 가격 변동을 확신할 수 없다는 문제점이 있다. 부동산 간접투자 시장의 확대와 함께 요구되는 시장 정보 제공의 측면에서 본 연구는 상당히 의미 있는 것으로 생각된다. 더불어 과소거래시장에 대한 지수 산정과 관련된 국내의 기존 연구들과 비교하면 현재 공신력 있는 기관들에서 공표되는 지수 산정의 틀 안에서 단순한 모형을 통해 상대적으로 안정적인 지수 산정이 가능한 방법론을 서울시 오피스 시장에 적용하고 실무적 활용 가능성을 확인했다는 점에서 차별화되는 부분이 있다.

베이저안 추세 모형이 기존 방법론과 비교할 때 추정능력이 크게 개선된다는 점은 여러 가지 지표를 통해 확인하였다. 특히 LLT 모형은 거의 대부분의 지표를 바탕으로 판단하자면 가장 좋은 지수를 추정한다고 볼 수 있을 것이다. 다만, 지수 평활화가 과하게 이루어진다는 점에서 실제로 시장 상황을 잘 보여주는 지수

인가에 대해서는 추가적인 검증이 필요할 것으로 보인다. 대신에 기존 방법론에 비해 표준오차, 지수 안정성이 크게 개선되어 나타나고, 최근 시점의 Revision이 LLT 모형과 크게 차이가 나지 않는 RWD 모형이 본 연구에서 얻은 결과에 한정하여 본다면 현재로서는 더 나은 대안이 될 것으로 보인다.

다만, 몇 가지 측면에서 여전히 한계점으로 남는 부분이 있다. 먼저, 지수의 평활화가 적절하게 이루어졌는지 혹은 시장의 전환점과 가격 변동을 적절하고 보여주고 있는 지에 대한 검증이 명확하게 이루어지지 않았다. 오피스 시장의 경우 부동산 간 이질성이 주택에 비해 월등히 크고 거래가 빈번하게 일어나지 않기 때문에 시장에서 가격 변동이 실제로 어떻게 나타나고 있는 지를 정확하게 측정하는데 한계가 있다. 그럼에도 불구하고 시장에서 적극적으로 활용되기 위해서는 시장 상황을 적절하게 반영하고 있는지에 대한 검증이 요구된다.

또한 반복매매모형이 구조적으로 가지는 Revision이라는 한계에 대한 추가적인 연구가 필요하다. 본 연구의 결과에서 세 가지 베이지안 추세 모형의 장기적인 Revision은 작게 나타난 점은 고무적이거나 가장 최근 시점의 경우에는 기존 방법론에 비해 크게 나타났다. 최근 시점 Revision이 높게 나타나는 경우 투자를 결정하거나 성과를 평가하고, 정책적 선택을 하는데 있어서 문제가 될 수 있다. 특히 시장의 전환점 부근에서 Revision이 더욱더 크게 발생하는 것을 확인할 수 있었는데 이러한 점은 앞서 언급한 바와 같이 모형 자체가 가지는 구조적인 한계에서 기인한 현상일 가능성이 있기 때문에 Revision이 발생하게 되는 패턴이나 구조적 특성에 대한 추가적인 연구를 진행하고 이를 해결하기 위한 대안을 찾아나갈 필요가 있다.

마지막으로는 실무적 활용 가능성의 측면에서 지수의 시차 발생 가능성에 대한 고민이 필요하다. 하나는 잔금청산 기간에 관한 문제이다. 주택과 달리 오피스 빌딩은 거래금액의 규모가 상당히 크기 때문에 계약이 이루어진 직후 바로 부동산 매입 금액이 지불되는 것이 아니라 일정 기간 동안 나눠서 지불될 수도 있는데 그 분할 비율도 다양하게 나타날 수 있다. 본 연구에서 구축한 자료에서는 전체 거래의 약 88%가 3개월 이내에 잔금 청산이 이루어진 것으로 나타났으며, 류강민 외(2011)과 최근까지 서울 오피스 매매가격지수를 작성하여 발표하였던 ‘이시스-대신 서울 오피스 매매지

수’에서는 오피스 빌딩 매매계약과 잔금청산 시점 사이에 대략 2~3개월 정도 차이가 발생한다는 점을 고려하여 2분기 이동평균을 한 지수를 산정하였다. 그러나 거래대금 지불 방식에 따라 거래 가격이 거래계약이 이루어지는 ‘Decision Making’ 시점의 선택인지 잔금청산 스케줄을 고려한 가격인지에 대해 불분명한 측면이 있어 이에 대한 추가적인 고민이 필요할 것으로 보인다.

그럼에도 지수산정 모형의 개선을 통해 부동산 가치 평가 시 시점 보정을 위한 기초 지수로 활용될 수 있는 가능성을 보여주었다. 또한 오피스 시장뿐 아니라 아파트 동별 혹은 단지별 가격지수, 비아파트 가격지수, 또는 주 단위 가격지수 등 부동산 시장에서 세분화된 다양한 시장 지표 산출에도 활용이 가능하다. 추후에 앞서 한계점으로 지적한 문제들과 더불어 보다 세분화된 시장을 자세히 들여다 볼 수 있는 연구를 수행해나간다면, 본 연구가 오피스 시장뿐 아니라 부동산 시장 전체에 걸쳐서 신뢰할 수 있는 정보 제공을 통한 정보 비대칭 해소 등 시장 전반에 기여할 수 있는 안정적인 지수 산정이 가능하다는 것을 확인하였다는 점에서 의미가 있다고 판단된다.

논문접수일 : 2021년 9월 29일
 논문심사일 : 2021년 10월 7일
 게재확정일 : 2021년 11월 8일

참고문헌

1. 국토교통부, “공모형 부동산간접투자 활성화 방안”, 보도자료, 2019.09.11.
2. 류강민 · 박수훈 · 이창무, “부동산 파생상품 개발을 위한 오피스 가격지수 산정”, 「선물연구」 제19권 제4호, 한국파생상품학회, 2011, pp. 363-387
3. 류강민 · 박유미 · 이창무, “비선형 회귀분석을 이용한 산술평균 반복매매지수 산정방법에 관한 연구”, 「주택연구」 제17권 제4호, 한국주택학회, 2009, pp. 259-278
4. 류강민 · 송기욱, “반복매매모형을 활용한 서울시 오피스 벤치마크 가격지수 개발 및 시험적 적용 연구”, 「LHI Journal」 제11권 제2호, 한국토지주택공사 토지주택연구원, 2020, pp. 33-46
5. 류강민 · 이상영, “S&P/Case-Shiller 반복매매모형을 이용한 주택가격지수 산정에 관한 연구”, 「주택연구」 제18권 제2호, 한국주택학회, 2010, pp. 183-204
6. 박연우 · 방두완, “평가기반 아파트가격지수에서의 비대칭 평활화 현상에 관한 연구”, 「주택연구」 제19권 제2호, 한국주택학회, 2011, pp. 23-46
7. 박헌수, “거래빈도가 낮은 시장에서의 실거래 부동산 가격지수 작성에 관한 연구 : 강남구를 대상으로”, 「부동산학연구」 제13집 제3호, 한국부동산분석학회, 2007, pp. 187-200
8. 박헌수 · 유은영, “상태공간모형을 활용한 부동산실거래가격지수 추정에 관한 연구 : 거래빈도가 낮은 지역을 중심으로”, 「부동산학연구」 제20권 제1호, 한국부동산분석학회, 2014, pp. 5-17
9. 손재영 · 김경환, “서울시 오피스 임대료의 횡단면 분석”, 「국토계획」 제35권 제5호, 대한국토·도시계획학회, 2000, pp. 279-295
10. 손진수 · 김병욱, “서울 오피스시장의 임대료지수 개발에 관한 연구”, 「국토계획」 제37권 제4호, 대한국토·도시계획학회, 2002, pp. 109-122
11. 송영선 · 윤명탁 · 이창무, “아파트 하위시장 실거래가 지수 산정방식 비교 연구”, 「부동산분석」 제6권 제3호, 한국감정원, 2020, pp. 1-19
12. 이상경, “서울시 오피스 매매가격지수 개발에 관한 연구”, 「서울도시연구」 제6권 제4호, 서울연구원, 2005, pp. 121-134
13. 이상경, “시변모수법에 의한 오피스 매매가격지수 구축에 관한 연구”, 「국토계획」 제42권 제5호, 대한국토·도시계획학회, 2007, pp. 233-245
14. 이상경, “오피스 투자 행태의 시공간적 특성에 관한 연구 - 투자 결정요인과 자본수익률을 중심으로”, 「서울도시연구」 제10권 제1호, 서울연구원, 2009, pp. 47-59
15. 이영유 · 이상경, “표본선택편의를 고려한 오피스 매매가격 결정요인 분석 및 매매가격지수 산정”, 「부동산학연구」 제19집 제1호, 한국부동산분석학회, 2013, pp. 83-96
16. 이용만, “특성가격함수를 이용한 주택가격지수 개발에 관한 연구 - 시간변동계수모형에 의한 연쇄지수”, 「부동산학연구」 제13집 제1호, 한국부동산분석학회, 2007, pp. 103-125
17. 이용만 · 이상한, “국민은행 주택가격지수의 평활화 현상에 관한 연구”, 「주택연구」 제16권 제4호, 한국주택학회, 2008, pp. 27-47
18. 이창무 · 김병욱 · 이현, “반복매매모형을 활용한 아파트 매매가격지수”, 「부동산학연구」 제8집 제2호, 한국부동산분석학회, 2002, pp.1-19
19. 이창무 · 김진유 · 이상영, “공동주택 실거래가 지수 산정에 관한 연구”, 「국토계획」 제40권 제4호, 대한국토·도시계획학회, 2005, pp. 121-134
20. 이창무 · 류강민 · 김지연, “Quantile Regression을 이용한 반복매매지수 산정에 관한 연구”, 「부동산학연구」 제19집 제4호, 한국부동산분석학회, 2013, pp. 27-40
21. 이창무 · 배익민, “시세가격을 활용한 아파트 실거래가 반복매매지수 산정”, 「부동산학연구」 제14집 제2호, 한국부동산분석학회, 2008, pp. 21-37
22. 정유신 · 이기영, “오피스 가격지수의 개발 및 거시경제변수와의 관계에 관한 연구”, 「부동산학보」 제43집, 한국부동산학회, 2010, pp. 184-199
23. 최성호 · 류강민 · 이건우 · 이창무, “반복매매모형을 활용한 오피스 매매가격지수에 관한 연구”, 「국토계획」 제45권 제7호, 대한국토·도시계획학회, 2010, pp. 119-131
24. 황규완 · 손재영, “MIT/CRE 2단계 추정법을 활용한 서울 오피스 가격지수 산출에 관한 연구”, 「주택연구」 제25권 제1호, 한국주택학회, 2017, pp. 151-175
25. A Division of S&P Global, S&P CoreLogic Case-Shiller Home Price Indices Methodology, 2020
26. Bailey, M. J., R. F. Muth, and H. O. Nourse, “A regression method for real estimate price index construction,” Journal of the American Statistical Association, Vol. 58, 1963, pp. 933-942
27. Bokhari, S. and D. Geltner, “Estimating real estate price movements for high frequency tradable indexes in a scarce data environment,” The Journal of Real Estate Finance and Economics, Vol. 45 No. 2, 2012, pp. 522-543
28. Case, K. E. and R. J. Shiller, “Price of Single Family Homes since 1970: New Indexes of Four Cities,” New England Economic Review, September/October, 1987, pp. 46-56
29. Case, K. E. and R. J. Shiller, “The Efficiency of the Market for Single Family Homes,” The American Economic Review, Vol. 79 No. 1, 1989, pp. 125-137
30. Clapp, J. M. and C. Giaccotto, “Estimating price indices for residential property: A comparison of repeat sales and assessed value methods,” Journal of the American Statistical Association, Vol. 87, 1992, pp. 300-306
31. Coulson, N. E. and D. P. McMillen, “The dynamics of intraurban quantile house price indexes,” Urban Studies, Vol. 44 No. 8, 2007, pp. 1517-1537

32. de Wit, E. R., P. Englund, and M. K. Francke, "Price and transaction volume in the Dutch housing market," *Regional Science and Urban Economics*, Vol. 43 No. 2, 2013, pp. 220-241
33. Durbin, J. and S. J. Koopman, *Time series analysis by state space methods Vol. 2*. Oxford: Oxford Univ Press, 2012
34. Fisher, J., D. Gatzlaff, D. Geltner, and D. Haurin, "Controlling for the Impact of Variable Liquidity in Commercial Real Estate Price Indices," *Real Estate Economics*, Vol. 31 No. 2, 2003, pp. 269-303
35. Francke, M. K., "Repeat Sales Index for Thin Markets: A Structural Time Series Approach," *The Journal of Real Estate Finance and Economics*, Vol. 41 No. 1, 2010, pp. 24-52
36. Francke, M. K. and A. van de Minne, "The Hierarchical Repeat Sales Model for Granular Commercial Real Estate and Residential Price Indices," *The Journal of Real Estate Finance and Economics*, Vol. 55 No. 4, 2017, pp. 511-532
37. Geltner, D. and S. Bokhari, *A Technical Note on Index Methodology Enhancement by Two-Stage Regression Estimation*, MIT Center for Real Estate LLC. Report, 2008
38. Goetzmann, W. N. and Liang Peng, "The Bias of the RSR Estimator and the Accuracy of Some Alternatives," *Real Estate Economics*, Vol. 30 No. 1, 2002, pp. 13-39
39. Goetzmann, W. N., "The Accuracy of Real Estate Indices: Repeat Sales Estimators," *Journal of Real Estate Finances and Economics*, Vol. 5, 1992, pp. 5-53
40. Guo, X., S. Zheng, D. Geltner, and H. Liu, "A New Approach for Constructing Home Price Indices: the Pseudo Repeat Sales Model and its Application in China," *Journal of Housing Economics*, Vol. 25, 2014, pp. 20-38
41. Hoffman, M. D. and A. Gelman, "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo," *Journal of Machine Learning Research*, Vol. 15 No. 1, 2014, pp. 1593-1623
42. Lancaster, K. J., "A New Approach to Consumer Theory," *Journal of Political Economy*, Vol. 74 No. 2, 1966, pp. 132-157
43. McElreath, R., *Statistical Rethinking: A Bayesian course with examples in R and Stan*, CRC Press, 2020
44. McMillen, D. P. and P. Thorsnes, "Housing Renovations and the Quantile Repeat-Sales Price Index," *Real Estate Economics*, Vol. 34 No. 4, 2006, pp. 567-584
45. Rosen, S., "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, Vol. 82 No. 1, 1974, pp. 34-55
46. Shiller, R. J., "Arithmetic repeat sales price estimators," *Journal of Housing Economics*, Vol. 1 No.1, 1991, pp. 110-126
47. van de Minne, A., M. Francke, D. Geltner, and R. White, "Using Revisions as a Measure of Price Index Quality in Repeat-Sales Models," *The Journal of Real Estate Finance and Economics*, Vol. 60 No. 4, 2020, pp. 514-553
48. Vehtari, A., A. Gelman, and J. Gabry, "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC," *Statistics and Computing*, Vol. 27 No.5, 2017, pp. 1413-1432
49. Watanabe, S., "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory," *Journal of Machine Learning Research*, Vol. 11, 2010, pp. 3571-3594
50. Yeon, Kyupil, "A Study on Repeat Sales House Price Index Based on Penalized Quantile Regression," *Procedia Computer Science*, Vol. 91, No. 6, 2016, pp. 260-267
51. 건축데이터 민간개방 시스템, <https://open.eais.go.kr>
52. 국토교통부 실거래가공개시스템, <https://rt.molit.go.kr>
53. 금융감독원 전자공시시스템, <https://dart.fss.or.kr>
54. 대법원 인터넷등기소, www.iros.go.kr

<국문요약>

부동산 과소거래시장에 대한 안정적인 실거래가 지수 산정에 관한 연구

송 영 선 (Song, Young-Sun)
신 혜 영 (Shin, Hye-Young)
이 창 무 (Lee, Chang-Moo)

최근들어 국내 오피스 시장의 투자 규모가 급격히 성장해왔고 부동산 간접투자 활성화를 위한 움직임이 활발한 상황에 따라 시장 정보의 투명성에 대한 요구가 증가하고 있으나, 정보 공개가 상당히 제한적이고 시장 상황을 세밀하게 진단할 수 있는 지표가 불충분한 상황이다. 특히 자본환원율의 분모로서 중요한 한 요소인 자산가치의 평가에 시점 보정을 위한 기초지수로 활용될 수 있는 매매가격지수는 민간 기업에서 발표가 되었으나 현재는 일부에서만 간신히 유지되고 있다.

오피스 시장은 주택시장과 비교하여 거래단위의 재고가 현저히 적고 거래 빈도 또한 적어 실거래 사례를 이용하여 안정적인 지수의 산정이 쉽지 않다. 따라서 본 연구에서는 서울시 오피스 매매가격지수 산정을 위하여 기존 보다 발전된 방법론을 제시하고자 한다. 이를 위해 기존에 널리 이용되어져 온 OLS 반복매매모형을 대신하여 베이지안 접근법을 도입하였고, 세 가지 베이지안 추세 반복매매모형을 이용하여 서울시 오피스 매매가격지수를 산정하였다.

지수를 산정한 후 기존 방법론을 이용한 지수와 비교한 결과 지수의 안정성, 추정결과의 통계적 신뢰도, 지수 변화(Revision), 모형의 예측력 등의 다양한 측면에서 크게 개선된 지수 추정 성능을 보이는 것을 확인하였다. 그러나 지수의 활용성 측면에서 중요한 Revision에 관한 추가적인 고민이 필요하다. 그럼에도 거래 빈도가 적어서 안정적인 지수 산정이 어려운 대상에 대해서는 그간 어려움을 겪어온 문제들을 해결할 수 있는 방안을 제시하였고, 향후에는 권역 혹은 규모별 등의 하부지수를 산정하여 보다 세분화된 시장 정보의 제공이 가능할 것으로 기대한다.

주 제 어 : 과소거래시장, 오피스 가격지수, 반복매매지수, 베이지안 추세 모형, 마르코프 연쇄 몬테 카를로