

계층적 베이지안 추론을 통한 아파트 단지별 실거래 기반 시세 개발*

Study of Apartment Complex Market Price Based on Hierarchical Bayesian Inference

권민성 (Kwon, Min-Sung)**
최우현 (Choi, Woo-Hyun)***
송영선 (Song, Young-Sun)****
이창무 (Lee, Chang-Moo)*****

< Abstract >

The real estate price index is an indicator for monitoring the price appreciation of an asset compared to the base point. Real estate is not only a consumption good, it is also an investment asset, so its appreciation can derive decision-making on the part of market participants. Residential assets, which can be mainly represented by “apartments” in South Korea, can be explained by regional or apartment complex-level submarkets. However, the traditional way of estimating price indices by using OLS (Ordinary Least Squares) has some limitations. It cannot estimate price coefficients when there are no transactions at the specific time point, so the traditional index cannot precisely capture the ups and downs of individual submarkets. The repeat-sales model, which is widely-used for constructing housing price indices, uses transaction pairs, so the lack of samples can be a serious obstacle in accurate estimation. In order to overcome the limitations of the OLS methods, we estimate stable submarket-specific indices with hierarchical Bayesian inference by MCMC (Markov Chain Monte Carlo) sampling. This study initially focuses on estimating sub-market indices for apartment-complexes in Mangwoo-dong and Junghwa-dong in Jungrang-gu. We subsequently compare the performance of the newly constructed indices by this research and the existing indices developed by R114.

Keyword : Repeat Sales Model, Bayesian Inference, MCMC Sampling, Price Indices

I. 서론

부동산 가격지수는 해당자산의 기준시점 대비 가격 변화율을 확인할 수 있는 지표이다. 부동산은 소비재

이면서 동시에 투자재의 성격을 가지며 그 영향력을 살펴보면, 대한민국의 경우, 국내 가구의 총자산 대비 그 비중이 60% 이상을 차지할 만큼 가격의 변화는 수 많은 시장참여자들의 행동 변화를 이끌어낸다(한국은행, 2021). 따라서 자산이 상업용이든 주거용이든, 또

* 이 논문은 2022년 한국부동산분석학회 상반기 학술대회 발표 연구를 수정, 보완한 연구임

** 한양대학교 도시공학과 석사과정, kwonminsung95@gmail.com, 주저자

*** 한양대학교 도시공학과 박사과정, eter0212@naver.com, 교신저자

**** 한양대학교 도시공학과 석박사 통합과정, youngsun_song@naver.com, 공동저자

***** 본 학회 정회원, 한양대학교 도시공학과 교수, changmoo@hanyang.ac.kr, 공동저자

한 시장참여자가 포트폴리오에 해당 자산을 편입했거나, 편입하고자 하는 상황 모두 그 움직임에 민감하게 반응한다.

아파트를 투자대상으로 살펴보면 해당 자산을 매입한 시점 이후의 임대수익과 금융에 들어간 비용, 그리고 현 시점에서의 가격을 신뢰성 있게 추정할 수 있어야 자산을 매각할 것인지, 일정기간 추가로 보유할 것인지 등의 행동을 취할 수 있다.

이처럼 시장참여자들의 합리적 판단을 바탕으로 한 행동을 이끌어내기 위해서는 비교적 세분화된 지수산정에 관한 연구가 필요하다. 상업용의 경우 권역별, 면적별 등의 하부시장으로 나눌 수 있고, 주거용의 경우 시군구별에서 동 단위, 그리고 단지 단위로 하부시장을 만들어 낼 수 있다. 그러나 OLS(Ordinary Least Squares) 방식으로 통계적 추정치를 산출하여 지수를 산정하게 되면 특정 하부시장의 특정 시점에 거래사례가 없을 시 추정치를 계산하지 못한다. 또한 거래사례가 적을 경우 한두 가지 사례의 움직임이 해당 시점 자체를 대표하는 상황을 만들어낸다는 단점이 존재한다.

특히 지수 산정에 널리 쓰이는 반복매매모형은 동일하게 인식되는 자산의 2회 이상 거래사례를 바탕으로 거래쌍을 만들어 활용하기에 거래사례의 한계는 더욱 부각된다. 나아가 단일시장 내 하부시장의 가격변동을 담을 수 있는 지수 산출에는 위의 한계가 극명히 드러난다. 하부시장의 개별적인 자료를 이용하여 지수를 산출할 경우 자료수의 부족 문제와 함께 해당 하부시장간 유사성이 보장되지 않을 수 있다. 특정 지역에서 나뉘는 하부시장의 경우 가격움직임이 완전히 독립적으로 나타나기 보다는 하부시장들 간 공유하는 가격움직임이 존재할 것이라고 판단된다.

이와 같이 자료 부족의 문제를 보완하고 하부시장간 가격변동의 유사성을 담을 수 있는 분석 구조가 필요하다. 최근 베이지안 추론을 바탕으로 하는 MCMC(Markov Chain Monte Carlo) 샘플링 기법과 계층모형을 활용하여 전체 시장은 물론 하부시장의 신뢰성 있고, 안정적인 지수산출에 관한 연구가 지속되고 있다. 한편 이를 아파트 시장에서 단지별로, 평형대별로 실거래가격의 움직임은 일반적으로 민간 부동산 플랫폼에서 확인할 수 있다. 그러나 거래사례가 적은 단지의 경우 시세가 실거래가격을 대표하지 못하는 상황이 존재하며, 그 움직임 또한 실제 거래가격의 움직임을

정확히 포착하지 못하는 양상을 보인다.

이에 본 연구에서는 위의 계층모형을 도입한 베이지안 추론을 활용하여 중랑구 내 중화동, 망우동의 가격 지수를 산정하고, 나아가 각 동의 아파트 단지별 지수를 산출한다. 또한 해당 지수들의 성능을 안정성, 통계적 신뢰도로 측정, 비교하고 부동산114에서 제공하는 아파트 단지별 평형대별 시세를 소형, 중형, 대형 평형으로 나눈 후 세대수에 따라 가중평균한 시세지수와 비교하여 개발된 단지별 평형대별 실거래가 기반 시세의 적합성을 비교 평가하는 것을 목적으로 한다. 다시 말해 아파트 단지별 각 평형대별 지수화라는 작업을 통해 보다 실거래 가격을 정확히 포착하는 시세를 개발해낼 수 있음을 확인하는 것에 목적을 둔다. 이를 바탕으로 산출한 아파트 단지별 평형대별 실거래가 기반 시세는 특히 거래사례가 적은 하부시장에서 기존 활용되는 조사 기반의 시세 대비 상대적으로 정확한 시세를 제공할 수 있다는 점에서 기존 시세의 한계점을 개선할 수 있다.

II. 선행연구

1. 실거래가격지수

실거래가격을 활용한 가격지수는 그 산정 방식에 따라 SPAR(Sale Price to Appraisal Ratio) 지수, 헤도닉 가격지수, 그리고 활용도가 가장 높은 반복매매모형 가격지수 등 다양한 형태를 가진다. SPAR 지수는 실거래가격을 기준시점 감정평가 가격으로 나눈 값으로 지수를 산출하는 방식이다. 헤도닉 가격지수는 헤도닉 특성함수를 통해 가격의 결정요인에 대한 통제 후 시점에 따른 계수를 바탕으로 지수를 산출하는 방식이다. 자산의 질적인 특성을 모두 통제하여 시점에 따른 가격의 움직임을 포착할 수 있다는 장점을 지닌다. 그러나 현실적으로 매 시점마다의 가격 및 질적 요소 자료를 구축해야한다는 단점이 있다. 또한 시간의 흐름에 따라 변화하는 질적인 요소를 적시에 통제할 수 없다는 현실적인 어려움이 존재한다.

반복매매모형 지수는 동일하게 인식되는 자산의 2회 이상 거래된 사례를 바탕으로 가격의 변화율을 활용하여 지수를 산정한다. 반복매매모형은 헤도닉 가격

함수에서 시작되어 동일자산의 특성은 가격의 차(difference)를 구하는 산식에서 소거되고 시간더미에 따른 변화를 포착한다. 따라서 자산의 특성 자료가 불필요하다는 장점이 존재한다. 이에 반해 해당 시점에 동일자산(또는 동일하게 인식되는 자산)의 반복적인 거래가 존재해야 거래쌍(Pairs)을 생성할 수 있다는 단점이 존재한다. 또한 현실적인 관점에서 동일한 자산의 리모델링 전후에 대한 질적 요소 통제 여부에 대한 논의도 필요하다.

본 연구에서는 반복매매모형 가격지수를 추정방식으로 하여 지수를 산정한다. 반복매매모형은 Bailey · Muth and Nourse(1963)에 의해 개발되었는데 Bailey · Muth and Nourse는 자산별로 동일한 가중치를 바탕으로 한 기하평균 지수를 산출하였다. 이는 Case and Shiller(1987, 1989), Shiller(1991)의 연구를 바탕으로 수정되어 동일가중 기하평균 지수부터 가치가중 산술평균 지수까지 개발되었다. 그럼에도 여전히 반복매매모형, 그리고 실거래가격지수가 갖는 문제점이 제시되었다. Clapp and Giacotto(1992)는 2회 이상 반복 거래된 자료만을 활용하여 지수를 작성하는 반복매매모형의 특성으로 인해 자료이용의 효율성이 떨어지고 표본 선택의 편의에 관한 문제가 있음을 확인하였고, 이창무 · 김병욱 · 이현(2002)은 관측되는 반복거래 시점 간 차이가 길어짐에 따라 발생하는 이분산(Heteroscedasticity) 문제를 제시하였으나 실용적인 수준에서 큰 문제를 초래하지는 않는 것으로 보고하였다.

류강민 · 이상영(2010), 최성호 외(2010)는 반복매매모형에서 발생하는 이분산의 문제를 해결하기 위해서 Case and Shiller(1987)가 제안한 WRS(Weighted Repeat Sale)모형에 도구변수를 이용한 지수를 산출하여 기존 지수가 갖는 단점을 일부 보완하려 했다. 이창무 외(2013)는 반복매매모형에 분위회귀 방식을 도입하여 지수의 통계적 신뢰도가 훼손되지 않고 지수의 불안정성을 보완할 수 있는 대안으로 제시하였으나 자료수가 극히 부족한 경우 중위수의 특성 상 오히려 불안정한 지수가 산출되는 문제가 발생함도 함께 보고하고 있다.

반복매매지수는 거래사례가 부족한 하부시장에서는 안정성을 갖춘 지수 산출에 한계를 보인다. 비아파트, 오피스를 필두로 한 상업용 시장의 경우 그 한계가 더욱 부각되고 있다. 아파트로 대변되는 주거용 부동

산의 경우 비교적 거래사례가 충분히 확보되어 있다고 볼 수 있다. 그러나 하부시장 별 지수를 살펴보면 이 또한 거래사례의 한계는 여전히 존재한다. 이와 같은 거래사례의 한계를 탈피하기 위해 국내외에서 다양한 방법들이 시도되고 있다.

박헌수(2007)는 전통적인 헤도닉 가격지수의 대안으로 시계열 모형을 도입해 거래빈도가 낮은 시장에 대해 지수의 안정성을 개선한 결과를 도출하였다. 황규완 · 손재영(2017), 류강민 · 송기욱(2020)은 Geltner and Bokhari(2008)의 Two-Stage Least Squares (2SLS)를 활용하여 비교적 안정된 서울 오피스 매매가격지수를 산정하였고, 송영선 외(2020)는 이동평균을 활용한 S&P/ Case-Shiller Home Price Index에서 착안하여 3개월 자료 중첩 서울시 아파트 하위시장 실거래가격지수 산정을 시도하여, 거래사례가 부족한 하부시장에서도 비교적 안정적인 지수 산정이 가능함을 확인하였다.

2. 베이지안 접근법의 활용

기존의 OLS(Ordinary Least Squares)방식을 통해 시간더미에 대한 회귀계수를 추정하면 거래사례가 존재하지 않는 경우 계수가 추정되지 않으며 거래사례가 부족한 경우 해당시점 지수 값이 부족한 사례들로 대표되는 문제점이 발생한다. 이러한 문제점에서 탈피하기 위해 사전지식(Prior)과 가능도(Likelihood)를 결합하여 사후분포(Posterior)를 산출하는 베이지안 추론(Bayesian Inference)을 바탕으로 지수를 산출하는 연구가 국내외에서 지속되고 있다.

Francke(2010)는 시간에 따른 계수는 현 시점에서 다음 시점으로 이동할 때 확률적 보행을 한다는 기본 전제하에서 RW모형(Random Walk)을 제시하였다. 또한 RW모형을 기준으로 일정한 이동량(Drift)을 가진 후 확률적 보행을 한다는 RWD모형(Random Walk with Drift), 이동량도 확률적 보행을 한다는 가정의 LLT모형(Local Linear Trend Model)을 통해 네덜란드의 주택 거래사례를 분석하여 거래사례가 적은 Thin market에서도 안정적으로 지수가 작성되는 것을 확인하였다.

Francke and Van de Minne(2017)은 Student's-t 분포를 가정한 계층 추세 모형(Hierarchical Trend Model)을 활용하여 미국 및 네덜란드 암스테르담을

배경으로 상업용, 주거용의 지역별-용도별 가격지수를 산정하였다. 반복매매모형을 공통 추세(Common Trend)와 클러스터별 추세(Cluster Specific Trends)를 추정하도록 작성하여 지수를 산정하고, 보다 세밀한 하부시장의 지수도 안정적으로 산정하는 것을 확인하였다.

우남교·권범준(2016)은 P-스플라인을 이용한 준모수적 접근방법을 활용하여 서울 25개 자치구의 종합유형 매매가격지수와 지가지수자료를 연구하였다. 계층적 베이지안 모형을 통해 주택매매가격 지수와 지가지수를 소지역 모형에 결합하고, 지가지수를 이용하여 주택매매가격 지수를 추정하였으며, 공간적 평활화의 문제가 발생하지만 추정 계수들이 실제 조사가격에 비해 안정적임을 확인하였다.

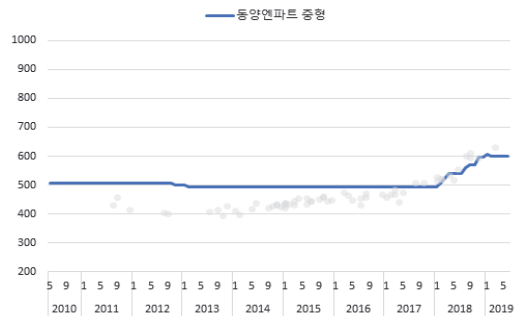
송영선·이창무(2021)는 계층모형을 도입한 베이지안 추론을 바탕으로 오피스 하부시장 별 반복매매모형 지수를 산출하였다. Francke et al.(2017)의 연구에서 착안하여 모형을 RW(Random Walk), RWD(Random Walk with Drift), LLT(Local Linear Trend)로 구성하여 안정적인 지수 산출이 가능함을 보였으며 각 모형별 지수의 성능을 비교하였다. 지수의 통계적 신뢰도(Mean of Standard Error of Index), 지수의 안정성(Stability Index), 지수변화(Revision Index), 그리고 베이지안 모형의 예측력(WAIC)을 통해 비교하였다.

3. 선행연구와의 차별성

이와 같이 베이지안 접근법을 이용한 선행연구들이 자료 수가 충분하지 않은 상황에서도 안정적으로 지수를 산출할 수 있는 것을 볼 수 있다. 본 연구에서는 선행연구들을 바탕으로 베이지안 추세 반복매매모형과 계층 추세 반복매매모형을 활용하여 거래 빈도가 낮은 하부시장의 아파트 실거래가격지수를 산정하고 이에 실거래가격을 입힌 시세를 개발하고자한다. 구체적으로 본 연구는 추정 기법 상 Francke and Van de Minne(2017), 송영선·이창무(2021)의 연구에서 착안하여 계층모형과 베이지안 추론을 통해 지수를 구축하고 시세를 개발한다. 그러나 모형의 적용대상을 아파트로 하고 아파트의 단지별, 그리고 평형대별로 적용한다는 점과 최종적으로 지수를 통해 시세를 개발한다는 점에서 차별성을 가진다.

한편 아파트 단지, 그리고 평형대별로 실거래가 기반 시세가 필요한지에 대한 논의가 나올 수 있는데 기존 민간 부동산 플랫폼이 제공하는 시세의 경우 거래 사례가 적은 지역에서는 그 시세 값이 실거래가격을 정확히 포착하지 못하는 양상을 보인다. 아래의 <그림 1>은 중랑구 중화동의 동양엔파트 중형 평형대(전용면적 60㎡ 초과 85㎡ 이하)를 대상으로 한 실거래가격(회색 산점도)과 부동산114에서 제공하는 시세(㎡당 가격)를 비교한 그림이다. 이를 통해 살펴보면 거래사례가 적은 하부시장에서는 현재 부동산114에서 제공되는 시세가 실거래가격을 정확히 반영하지 못하고 그 변동의 양상을 포착하지 못하는 것을 보이며 동시에 이러한 한계점을 개선할 실거래가 기반 시세의 필요성을 나타낸다.

<그림 1> 동양엔파트 시세 - 실거래 비교



III. 분석의 틀

1. OLS 반복매매모형

반복매매지수모형은 Bailey et al.(1963)에 의해 개발되었다. Bailey et al.(1963)의 반복매매모형은 모든 거래된 자산에 동일한 가중치를 부여하는 방식으로 지수를 산출하며 이를 산식으로 나타내면 식(1)과 같다. 이는 Case and Shiller(1989)에 의해 수정 및 발전되어왔는데 Case and Shiller는 거래되는 자산의 가치를 변동률에 가중하여 보다 시장참여자들이 체감하는 가격의 등락을 나타내고자하였다.

Bailey et al.(1963)가 개발한 반복매매모형의 산식은 아래의 식(1)에 나타나 있는데 여기서 두 번째 거래시점(s) 자산 i의 가격($p_{i(s)}$)과 첫 번째 거래시점(f) 자

산 i 의 가격($p_{i(f)}$)에 각각 로그를 취하여 그 차이를 산출한다. 해당 로그가격의 차이는 두 번의 매매시점 간 누적 로그가격변동을 의미하고 다시 두 번째 거래시점 로그 가격지수 값인 $\mu_{(s)}$ 에서 첫 번째 거래시점 로그 가격지수 값 $\mu_{(f)}$ 를 빼는 형태로 계산된다. 이는 누적 변동을 방식으로 계산하는 형태이며 모형을 식(2)와 같은 형태로 나타내면 직전 시점 대비 로그 가격변동을 나타내는 기간변동을 표현할 수 있다. 식(2)의 구성은 첫 번째 거래시점과 두 번째 거래시점 사이의 기간에 대해 모든 기간간 변동을 포착한다. 이를 통해 첫 거래시점과 이후 1시점 사이의 변화를 β_{f+1} , 두 번째 거래시점과 해당 시점 직전 1시점 사이의 변화를 β_s 로 나타내어 그 사이의 값을 모두 더하여 자연지수를 취하면 해당 시점의 자산 i 의 지수 값이 계산된다(송영선 · 이창무, 2021).

$$\ln p_{i(s)} - \ln p_{i(f)} = \mu_{(s)} - \mu_{(f)} + e_{i,s} - e_{i,f},$$

$$e_{i,t} \sim Normal(0, \sigma_e^2)$$

식(1)

$$\ln p_{i(s)} - \ln p_{i(f)} = \beta_{f+1} + \dots + \beta_s + e_{i,sf}$$

$$e_{i,sf} \sim Normal(0, 2\sigma_e^2)$$

식(2)

- $p_{i(s)}$: 두 번째(Second) 거래시점 자산 i 의 가격
- $p_{i(f)}$: 첫 번째(First) 거래시점 자산 i 의 가격
- $\mu_{(t)}$: 기준시점 대비 t 시점까지 누적 로그 가격변화
- β_t : 자산 i 의 $t-1$ 시점 대비 t 시점 로그 가격변화
- $e_{i,t}$: 오차항

본 연구에서 사용하는 OLS 반복매매지수모형은 Bailey et al.,(1963)의 동일가중-기하평균 방식의 반복매매모형을 활용한다. 반복거래쌍의 동일자산, 동일거래시점으로 중복되는 가격자료는 평균가격을 대표 가격으로 이용하며, 중복 가격자료 수를 회귀분석에서 가중치로 부여한다.

일반적으로 누적변동률 방식은 기준시점 대비 가격의 변동을 측정하기에 시점이 뒤로 갈수록 가격변동률의 평균값의 변동폭이 넓어지는 경향을 가진다. 이에

따라 지수의 통계적 신뢰도(MSEI)가 기간변동률 방식 대비 낮아지는 경향을 가지며 지수의 안정성 또한 기간변동률 대비 다소 낮은 경향을 가진다. 본 연구에서는 중랑구-망우동-중화동 지수 산정은 누적변동률 방식으로 산정하고, 망우동 단지별, 중화동 단지별 지수와 지수별, 지수간 비교에서는 보다 유연한 방식인 기간변동률 방식만을 채택하여 산출한다. 기간변동률 방식의 지수는 이전시점까지 산출된 추정계수들과 해당 시점 추정계수의 합에 자연지수를 취한 $\exp(\beta_1 + \dots + \beta_{t-1} + \beta_t) \times 100$ 의 값으로 산정한다.

2. 베이지안 추세 반복매매모형

거래사례가 충분치 않은 시장에 대한 지수 산정에 있어 OLS 방식을 통한 반복매매모형을 이용하는 경우, 지수의 안정성이 결여되고 지수 값이 산정되지 않는 한계가 있다. 이러한 문제점에서 탈피하기 위해 최근에는 베이지안 추론을 바탕으로 한 반복매매모형이 거래사례가 적은 시장의 지수 산출에 활용되고 있다.

베이지안 추론 또는 베이지안 추세 모형은 다음의 식(3)과 같은 베이즈 정리에 기반하여 현상이나 모수에 대한 연구자의 주관적인 사전적 지식이나 정보를 사전분포(Prior distribution, $p(\theta)$)로 정의하는 것으로 시작한다. 자료 자체에 대한 사전적 정보가 없거나 불확실할 경우 무정보적 사전분포를 가정하거나 넓은 사전분포(Flat Prior)를 정의할 수 있다. 이를 자료의 가능도(Likelihood, $p(Y|\theta)$)와 결합하여 시뮬레이션을 통해 사후분포(Posterior distribution, $p(\theta|Y)$)를 얻을 수 있다(Goetzmann, 1992; Francke, 2010; Francke and Van de Minne, 2017; 송영선 · 이창무, 2021). 시뮬레이션에 사용되는 MCMC(Markov Chain Monte Carlo)의 세부 샘플링 방식은 여러 가지가 있으나 본 연구에서는 Stan을 활용한 NUTS(No U-Turn Sampler)를 이용한다.

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

식(3)

OLS(Ordinary Least Squares)방식으로 산출하는 반복매매모형에서는 각 시점의 기준시점 대비 누적 가격변동률 또는 직전시점 대비 가격변동률을 나타내

는 추정계수(β_t)가 시점 간에 독립적인 관계를 가정했다. 그러나 베이지안 추세 반복매매모형에서는 전 시점의 가격이 갖는 확률적 분포 내에서 다음 시점의 가격수준이 결정되는 관계를 산식에 포함하고 있다. 즉, 한 시점의 가격수준은 이전 시점의 가격수준을 기준으로 확률분포 내에서 변동하여 결정됨을 가정한다.

본 연구에서는 Francke(2010), Francke and Van de Minne(2017)의 연구에서 활용한 RWD(Random Walk with Drift)모형을 이용하여 지수를 산정하고 분석을 실시한다. 식(4)는 RW(Random Walk)모형을 수식으로 나타낸 것이고 식(5)는 RWD(Random Walk with Drift)모형이다.

$$\begin{aligned}
 p_{i,s} - p_{i,f} &= \mu_{t(s)} - \mu_{t(f)} + e_{i,sf} \\
 e_{i,sf} &\sim Normal(0, 2\sigma_e^2) \\
 \mu_{t+1} &= \mu_t + \xi_t, \\
 \xi_t &\sim N(0, \sigma_\xi^2)
 \end{aligned}$$

식(4)

$$\begin{aligned}
 \mu_{t+1} &= \mu_t + \kappa + \xi_t, \\
 \xi_t &\sim Normal(0, \sigma_\xi^2) \\
 \kappa &\sim Normal(0, \sigma_\kappa^2)
 \end{aligned}$$

식(5)

RW모형은 한 시점의 가격수준(μ_t)은 이전시점 가격수준(μ_{t-1})에서의 확률적인 보행의 결과로 결정된다고 가정한다. 나아가 RWD모형은 이전 시점 가격수준에서 확률적 분포를 갖는 이동량(k)만큼 이동 후 확률적 보행의 결과로 한 시점의 가격수준이 결정된다고 가정한다(Francke, 2010).

Francke(2017)의 연구에서는 샘플의 개수가 충분치 않을 경우 오차항 e_t 의 분포를 정규분포로 가정하는 것 대비 t(student's-t)분포로 가정하는 것이 극단값(Outlier)에 대해 보다 높은 유의성을 제공한다는 점에서 착안하여 오차항의 분포를 t분포로 가정하였다. Francke(2017)의 연구는 분기당 평균 3개의 거래쌍으로 지수를 산정하면서 오차항의 분포를 t분포로 가정하여 진행하였다. 본 연구에서는 망우동, 중화동의 경우 월평균 10개 내외의 거래쌍이 존재하고 단지별 지수의 경우 164개월에 대해 약 1000개의 거래쌍이 사용되어 월 평균 6개의 자료가 활용된다. 따라서 본

연구에서는 오차항의 분포를 선행연구와 같이 t-분포로 가정하여 지수를 산출하였다.

3. 계층 추세 반복매매모형

표본의 수가 적으면 일반적으로 큰 데이터 집합을 정확하게 추정하는데 한계가 있고, 기존 데이터에는 잘 맞지만 새로운 데이터에 대해 올바른 추정치를 제시하지 못하는 과적합(Overfitting)의 경향이 존재한다. 이러한 경향은 거래사례 수가 부족한 부동산 시장 내 하부시장에서 빈번하게 발생된다. 예를 들어 본 연구에서 공간적 배경으로 삼는 중랑구-망우동, 중화동 내의 각 아파트 단지는 그 가격의 움직임이 완전히 상호 독립적인 구조를 가진다고 판단하기에는 무리가 있다. 이에 착안하여 하부시장의 개별 자산들은 하나의 큰 공통적 추세를 공유하고 그 추세에서부터 개별적인 편차(Deviance)만큼 이탈한 것이 해당 자산의 지수로 대표될 수 있다. 이처럼 계층모형은 그룹 간, 그룹 내의 편차를 나타내는 매개변수는 서로 유사하고 특정한 분포에서 시작된다고 가정한다(마쓰우라 겐타로, 2019).

Francke(2017)는 계층모형을 적용하여 필라델피아/볼티모어 지역의 상업용 부동산(업무용, 산업용, 상업용 등)지수와 네덜란드 암스테르담 지역의 주거용 부동산(단독주택, 공동주택)지수를 산출하였다.

결과적으로, 계층 추세 반복매매모형은 베이지안 추세 반복매매모형에 하부시장 간의 구조적인 관계를 결합함으로써 시장 전체 자료를 모두 활용하여 보다 세분화된 하부시장에서도 효율적이고 안정적인 하부시장 지수를 산정할 수 있다. 식(6)은 Francke and Van de Minne(2017)이 설명한 계층구조를 도입한 반복매매모형의 산식이다.

$$\begin{aligned}
 p_{i,t2} - p_{i,t1} &= \mu_{t2} - \mu_{t1} + \sum_{j=1}^J v_i^j (\lambda_{t2}^j - \lambda_{t1}^j) \\
 &\quad + e_{i,t2} - e_{i,t1}, \quad e_{i,t} \sim N(0, \sigma_e^2) \\
 \lambda_{t+1}^j &= \lambda_t^j + \zeta_t^j, \quad \zeta_t^j \sim Normal(0, \sigma_\zeta^2 I_{n_j})
 \end{aligned}$$

식(6)

$p_{i,t2(t1)}$: 자산 i의 두 번째 거래 시점 로그 가격(첫 번째 거래(t1))
 μ_t : t시점의 공통 추세

I_{n_j} : 단위행렬

v_i^j : 자산 i 가 속하는 하부시장 j 를 나타내는 행벡터

λ_t^j : 하부시장 j 의 t 시점 추세

$e_{i,t}$: 오차항

베이저안 추세 반복매매모형에 하부시장 간의 구조적인 관계를 결합하기 위해 하부시장 별 추세를 나타내는 항을 추가함으로써 시장 전체의 공통적인 추세(Common Trend)를 나타내는 $\mu_{t2} - \mu_{t1}$ 와 이로부터 벗어난 편차로 개별 하부시장의 추세(Cluster Trend)를 나타내는 $\sum_{j=1}^J v_i^j (\lambda_{t2}^j - \lambda_{t1}^j)$ 의 합을 통해 하부시장 각각의 가격변동을 나타낼 수 있는 형태로 모형이 구성된다. 모형의 특성상 특정 시점에서 거래사례가 없는 하부시장의 경우에도 시장 전체의 공통적인 추세를 통해 지수를 산정할 수 있다(Francke and Van de Minne, 2017; 송영선 · 이창무, 2021).

IV. 자료구성 및 기초분석

1. 자료구성

본 연구에 사용한 자료는 국토교통부 실거래가 공개시스템 신고 자료로 분석의 시간적 범위는 2006년 1월부터 2019년 9월까지이다. 공간적 범위는 단지의 수는 많되, 해당 단지별로 관측되는 거래사례는 상대적으로 적은 중량구를 설정하였는데 이는 계층모형을 통한 베이저안 접근법이 전체 거래사례가 적고 많은 계층으로 나뉘는 상황에서도 신뢰성 있는 지수 추정이 가능함을 나타내기 위해서이다. 이를 바탕으로 서울시 전체 995,292건의 실거래가 신고자료 중 중량구 자료 수 32,070건으로 분석을 진행하였다.

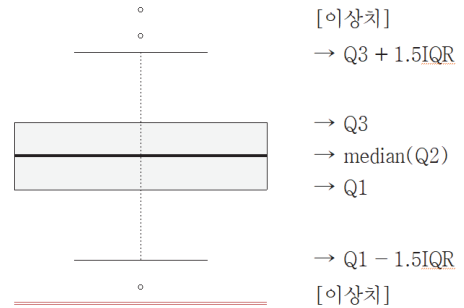
2. 가격 이상치 제거

국토교통부 실거래가 자료에 신고된 거래자료 중 신고상 오기재 또는 거래방식의 차이로 인한 시장의 거래가격과 차이가 크게 발생하는 경우 등 이상치 또는 극단치가 존재한다. 안정적인, 시장의 흐름을 반영하

는 지수를 산정하기 위해 해당하는 이상치 및 극단치를 제거한다. 이를 제거하기 위해서 단위가격에 로그를 취한 값으로 <그림 2>와 같은 상자도표식 이상치 제거법(Box-plot)을 이용한다.

상자도표식 이상치 제거 과정에서 단일 시점을 기준으로 삼을 시 매 시점마다 로그 단위가격 편차가 크게 나타날 수 있다. 위의 방식은 1,3사분위수, 중위수를 산정하여 이상치를 판별하고, 특히 해당 시점 거래사례가 부족한 경우에는 분석기간 전체를 관점으로 볼 때 이상치로 간주되지 않을 거래사례들이 이상치로 간주되는 상황이 발생할 수 있다. 따라서 이상치 판별 기준에 활용되는 중위수, 1,3사분위수의 계산은 각각 과거 1년, 4개월 간의 자료를 병합한 자료(N-months Pooling)를 이용하여 산출한다. 산출된 1,3사분위수(Q1, Q3)를 바탕으로 둘의 차이를 IQR(Inter-Quartile Range)이라 칭한다. 계산한 IQR 값과 Q3, Q1 값의 합 또는 차의 방식으로 각각 정상 거래사례의 상한(Q3 + 1.5IQR) 및 하한(Q1 - 1.5IQR)을 산정한다. 상한선을 넘거나 하한선보다 못 미치는 거래가격을 이상치로 판별하여 제거하고 분석을 진행한다.

<그림 2> 상자도표식 이상치 제거 개념도



가격 이상치 제거 결과, 중량구의 총 32,070건 중 450건, 약 1.4%가 이상치로 분류되어 제거한다.

<표 1> 이상치 개수 및 탈락율

	총 자료 수	이상치 수	탈락율
중량구	32,070	450	1.4%

3. 동일주택 가정

2회 이상 거래된 자산의 가격의 차를 이용하는 반복 매매모형의 특성상 동일한 주택의 거래사례를 확보해야 한다. 그러나 완벽하게 동일한 주택을 식별하는 데 어려움이 있으며, 완전히 동일한 주택을 거래쌍으로 선정한다고 하더라도 그 거래의 관측빈도가 낮아 자료 이용의 비효율성이 대두될 수 있다. 이러한 문제점을 보완하기 위해 유사한 특성을 가진 주택을 동일주택으로 가정하고, 각각의 가정에서 2회 이상 거래되는 사례의 전체 자료 수 대비 자료이용률을 살펴본 후 합리적인 가정을 선택할 수 있다. 동일주택에 대한 가정을 아래와 같이 3가지로 분류하고 각각의 자료이용률을 비교한다. 첫 번째로 단지+동+층그룹(저층/중층/고층), 두 번째로 단지+동+층그룹+면적, 마지막 가정으로는 단지+동+층+면적을 동일주택으로 가정하여 이용률을 산출하였다.

<표 2> 동일주택 가정 및 자료이용률

동일주택 가정	원자료	이상치 제거	2회 이상 거래자료	자료 이용률
가정 1	32,070	31,620	31,465	98.1%
가정 2			30,875	96.3%
가정 3			26,458	82.5%

세 번째 가정인 단지+동+층+면적의 동일주택 가정을 제외한 나머지 두 가정의 자료이용률이 95%이상으로 높게 나타났다. 이에 단지 + 동 + 층 + 면적의 동일주택 가정을 중량구, 동 단위 월별 지수 및 단지 단위 분기 지수 산출에 활용하였다. 나아가 단지 단위의 지수를 월별로 안정적으로 산출하기 위해 동일주택 가정을 단지 + 동 + 층그룹 + 면적으로 보다 완화하였다.

최종적으로 구축된 자료는 2006년 1월부터 2019년 9월까지 165개월간 총 26,458건이며, 분기별로는 평균 481건, 월별로는 평균 160건의 자료를 지수 산정에 이용한다.

V. 분석결과

본 절에서는 최종적으로 구축한 자료를 활용하여 각종 지수를 산출하고 해당 지수의 성능 비교과정을 제

시하며 최종적으로 실거래가지수의 시세지수로의 활용가능성에 대해 검토한다. 지수산출의 순서는 다음과 같다. 먼저 중량구, 망우동, 그리고 중화동의 가격지수를 OLS방식과 RWD방식으로 산출하여 두 방식의 지수 성능을 비교한다. 다음으로 범위를 중화동 내 단지로 좁혀 중화동 자료를 단일 계층(단지)으로 구축하여 아파트 단지별 가격지수를 기간변동률을 통해 산출한다. 네 번째로는 중화동의 자료를 평형대별로 이루어진 단일 계층을 활용하여 지수를 산출한다. 이후 위의 평형대와 단지로 이루어진 계층을 동일한 위계로 두고 복수의 계층을 구축하여 월 단위 지수를 산출한다. 이후 비교적 거래사례가 많은 특정 아파트 단지의 실거래가격과 시세지수를 비교한다. 시세지수는 부동산 114에서 제공하는 평형대별 시세를 소형, 중형, 대형으로 나누어 각 평형대별로 세대수를 가중하여 산출한다. 시세지수와외의 비교를 통해 실거래가지수가 가지는 장점을 부각하고 최종적으로 실거래가지수에 특정 시점 실거래 가격을 입혀 시세를 개발한다.

계층적 추세를 담은 베이지안 추론을 위해 R Studio에서 rstan 패키지를 활용한다. rstan 패키지 활용 시 분석을 위한 Markov Chain 설정은 chains=3, warm up=1,000회, iteration=2,000회로 총 3,000회 샘플링 결과로부터 사후분포를 도출한다. Markov Chain의 세부설정은 Stan에서 제공하는 rstan 매뉴얼에서 기본 값을 확인할 수 있다. 자료를 전반적으로 읽고 구축된 모형을 통해 계수 값을 산출하는 역할을 맡는 chain은 기본 값이 4로 지정되어 있고, 무수히 많은 샘플링 시도를 한다는 뜻의 Monte Carlo의 실행 횟수(iteration)는 2,000회, 실제 샘플링 전에 chain이 자료를 훑어보면서 그 구조를 파악하게끔 돕는 warm-up 구간은 iteration의 절반으로 기본 값이 설정되어 있다. 또한 베이지안 통계의 적용에 관한 내용을 기술한 「Statistical Rethinking」(Richard McElreath, 2020)에서는 chain 개수를 3으로 하는 것을 추천하고 있다. 이에 본 연구에서는 iteration=2,000, warm-up=1,000, 그리고 chain 개수를 3과 4로 각 한 번씩 샘플링 하였으나 지수 값, 지수의 성능은 동일하나 3개의 chain을 구동하는 것이 시간 효율적으로 나타났다.

지수 산정 후 지수의 성능을 평가함에 있어 통계적 신뢰도(MSEI)와 안정성 지수(SI) 두 가지를 기준으로 한다. 두 기준은 이창무·류강민·김지연(2013)에 의해 개발되었는데 MSEI(Mean Standard Error of

Index)는 지수 산출시 계수와 함께 산출되는 표준오차를 바탕으로 만들어지는데 기준시점을 제외한 표준오차의 평균값들의 합에 100을 곱하고 전체 시점 개수로 나눈 값이다. 통계적 신뢰도 값은 0에 가까울수록 우수하다고 평가한다(이창무·류강민·김지연, 2013; 송영선·이창무, 2021). 아래의 식 (7)에 해당한다.

안정성 지수는 지수 변화의 안정성을 계량화하여 평가하기 위해 이창무·류강민·김지연(2013)에 의해 개발되었다. 첫 시점 지수 값과 마지막 시점 지수 값을 직선으로 연결한 후 지수의 형태가 직선에 가까우면 1에 가깝고, 첫 시점에서 마지막 시점의 지수 값으로 이동하는 값들의 편차가 커지면 점점 0에 가까운 형태를 가진다. 다시 말해 상방, 하방으로의 변동이 적고 평탄하게 움직일수록 1에 가깝고 변동이 심할수록 0에 가까운 값을 가진다(이창무·류강민·김지연, 2013; 송영선·이창무, 2021). 아래의 식 (8)에 해당한다.

$$MSEI = \sum_{t=2}^T S.E_t \times 100 / T$$

$S.E_t$: t시점 산출된 지수 값의 표준오차의 평균
식(7)

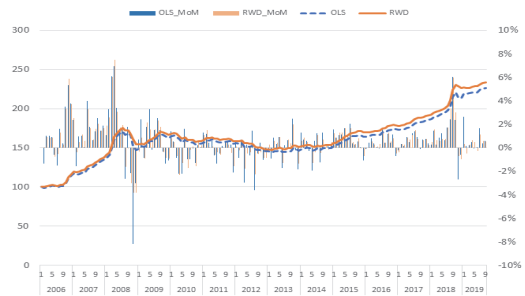
$$SI = \frac{\sqrt{(T-1)^2 + (X_T - X_1)^2}}{\sum_{t=2}^T \sqrt{1 + (X_t - X_{t-1})^2}}$$

X_t : t시점 산출된 지수 값
 T : 분석 종료 시점
식(8)

1. 중랑구 지수 산정결과

<그림 3>은 각각 OLS 방식과 RWD 방식으로 추정된 중랑구의 아파트 실거래가 지수이다. 중랑구는 분석 기간 동안 기준시점 대비 약 2.5배 아파트 실거래가 지수가 상승한 것으로 나타난다. 반복거래쌍이 충분하여 OLS 방식의 가격지수의 움직임과 RWD방식의 지수가 유사하게 산정되며 안정적인 모습을 보이나 <표 3>을 통해 지수 성능 중 통계적 신뢰도(MSEI)를 비교하면 RWD가 우수한 신뢰도를 제시함을 알 수 있다.

<그림 3> 중랑구 실거래가격 지수



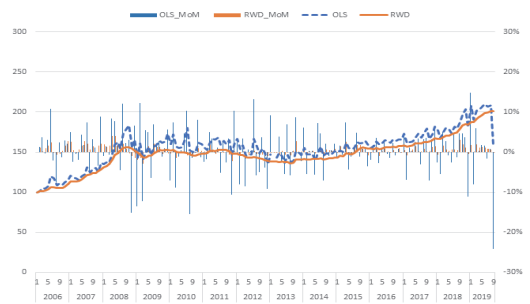
<표 3> 중랑구 지수성능 비교

구분	OLS	RWD
MSEI	1.2219	0.0136
SI	0.5392	0.6025

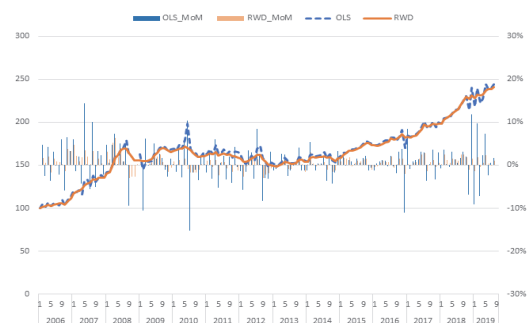
2. 망우동·중화동 지수 산정결과

<그림 4>와 <그림 5>는 중랑구의 하부시장인 망우동과 중화동의 아파트 실거래가 지수를 나타낸 그림이다. 분석 기간 동안 전체적인 실거래가 지수는 상승한 것으로 나타난다.

<그림 4> 망우동 실거래가격 지수



<그림 5> 중화동 실거래가격 지수



중랑구 하부시장인 망우동과 중화동의 경우 지수 산출에 사용된 반복 거래쌍은 각각 1,090쌍, 1,473 쌍으로 이루어졌다. OLS를 활용한 지수의 경우 거래사례가 없는 월은 지수 값이 산출되지 않으며, 거래사례가 적은 월의 경우도 불안정한 모습으로 지수 값이 산출되었다. 이에 반해 RWD 모형의 경우 망우동, 중화동 지수 모두에서 안정적인 지수 값이 산출되었고, 거래사례가 없는 월의 경우도 사전분포 및 자료의 가능성을 바탕으로 안정적으로 연결하였다.

각 동별 모형에 따른 지수의 성능을 살펴보면 RWD모형이 상대적으로 지수 추정 신뢰성이 높다는 것을 확인할 수 있다.

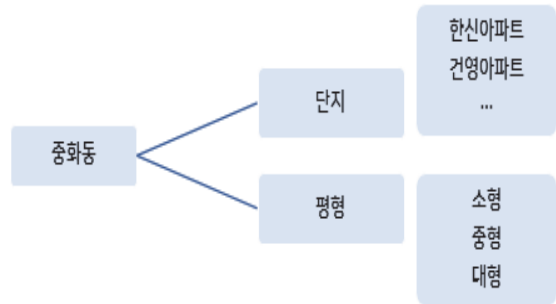
<표 4> 망우동, 중화동 지수성능 비교

구분	망우동		중화동	
	OLS	RWD	OLS	RWD
MSEI	5.990	0.046	3.538	0.035
SI	0.155	0.667	0.112	0.494

3. 중화동 단지별 및 평형대별 지수 산정결과

구별, 동별 지수에서 개별 아파트 단지로 분석을 세분화할 경우 지수 산정 시 필요한 거래쌍이 급격히 적어지는 문제점이 발생한다. 이 경우 위와 동일하게 RWD모형을 통해 산정을 시도하여도 지수의 신뢰도에 문제가 발생할 가능성이 있다. 따라서 아파트 단지별 지수 추정은 계층모형을 도입한 베이지안 추세 모형을 활용하고 자료의 이용률을 높이기 위해 기존의 동일주택 가정 상 가정 3(단지+동+층+면적)이 아닌 가정 2(단지+동+층그룹+면적)를 채택하여 실거래가지수를 산정한다. 계층모형으로 산정한 실거래가지수는 해당 아파트 단지들이 함께 공유하는 공통추세(Common Trend)와 해당 아파트 단지만이 갖고 있는 공통추세로부터 편차들로 이루어져 있다. 따라서 개별 아파트 단지의 지수는 공통추세를 나타내는 계수와 개별 편차를 나타내는 계수를 동시에 추정하여 그 둘의 합으로 이루어진다.

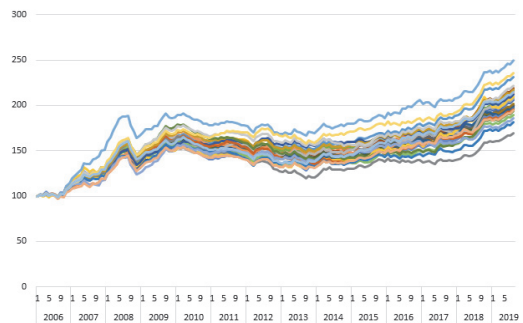
<그림 6> 계층모형의 구조도



위의 <그림 6>은 계층모형을 도식화한 구조도이다. 본 절에서는 두 개의 지수를 추정하는데 먼저 단지만을 활용하는 단일 계층으로 단지별 지수를 추정한다. 여기서 모든 거래쌍은 각각 아파트 단지별 고유한 코드를 지닌다. 이어서 평형대를 단일 계층으로 활용하는 평형대별 지수를 추정한다. 평형대별 계층은 전용면적 60㎡ 이하는 소형, 60㎡ 초과 85㎡ 이하는 중형, 그 외는 대형으로 나누어 각 평형대별 고유한 코드를 부여했다. 고유한 코드를 바탕으로 중화동 내 공통적인 가격움직임과 그로부터 벗어난 개별 단지의 편차를 추정한다.

이를 통해 중화동의 단지별 지수와 평형대별 지수를 기간변동률 추정방식을 활용하여 산출하였다. <그림 7>과 <표 5>는 각각 단지별 지수와 지수의 성능을 보여주고 있으며 <그림 8>과 <표 6>은 평형대별 지수를 나타낸다.

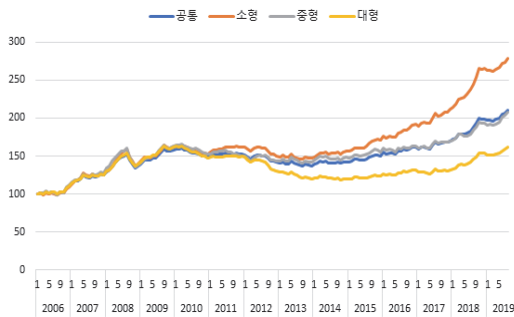
<그림 7> 중화동 단지별 기간변동률 지수



<표 5> 중화동 단지별 지수성능 비교

중화동	MSEI	SI
기간변동률	0.023	0.521

<그림 8> 중화동 평형대별 기간변동률 지수



<표 6> 중화동 평형대별 지수성능 비교

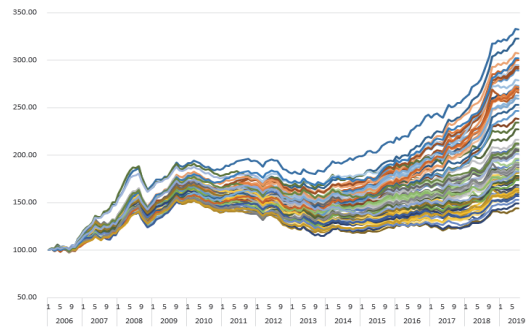
중화동	MSEI	SI
기간변동률	0.027	0.530

<표 7> 중화동 단지별/ 평형대별 지수성능 비교

중화동	MSEI	SI
단지 + 평형	0.026	0.575

<표 7>에는 식(9)의 방식으로 산출한 중화동 단지별, 평형대별로 산출된 공통지수의 성능을 제시하였다. 아래의 <그림 9>는 중화동 내 모든 단지와 평형대별로 산출된 계수를 통해 작성한 각 단지별 평형대별 지수로 공통 지수 포함 총 85개의 지수이다. 분석기간 중 가장 높은 상승률을 보인 단지와 평형대는 한솔아파트의 소형 평형대였으며 가장 낮은 상승률을 기록한 단지와 평형대는 동양엔파트 2차 아파트의 대형 평형대로 분석되었다.

<그림 9> 중화동 단지 + 평형대별 지수



4. 중화동 단지 + 평형대 지수 산정결과

본 절에서는 중화동 자료를 두 계층으로 나누어 지수를 산출한다. <그림 6>의 표현대로 단지별 계층은 중화동 자료 전체를 단지별로 구성하며, 평형대별 계층은 중화동 자료 전체를 평형대별로 나눈다. 따라서 단지와 평형대 계층이 위계상 동등한 위치를 가진 모형으로 설정된다. 단지별 계층의 경우 총 28개의 단지별로, 평형대별 계층의 경우 총 3개의 평형대별로 각 시점 계수가 산출된다. 이를 통해 개별 단지의 특정 평형대의 추세를 확인할 수 있다. 예를 들어 중화동의 한신아파트 소형 평형대의 t시점 추세는 아래의 식(9)로 표현할 수 있다. 아래의 식(9)의 방식으로 지수를 28개 단지, 3개의 평형대를 대상으로 추정하여 공통지수를 포함하여 전체 85개의 지수가 산출된다.

$$\beta_t + \gamma_t^{\text{한신}} + \delta_t^{\text{소형}} \tag{식(9)}$$

β_t : t시점 공통 추세

$\gamma_t^{\text{한신}}$: t시점의 한신아파트 단지별 추세

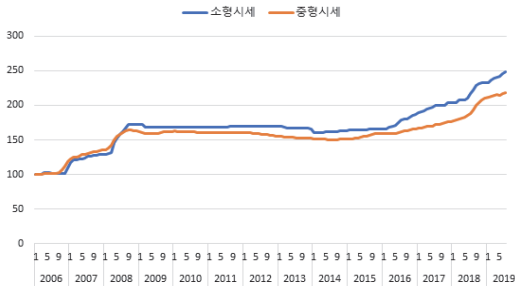
$\delta_t^{\text{소형}}$: t시점의 소형 평형대별 추세

5. 시세지수와의 비교

본 절에서는 앞 절에서 산출한 85개의 지수를 바탕으로 개별 단지의 특정 평형대 지수를 시세지수와 비교하여 실거래가지수의 시세로써의 활용가능성을 평가하고자한다. 시세지수를 활용하기 위해 부동산 114에서 제공하는 아파트 단지별 시세를 활용하였다. 부동산114는 각 아파트의 공급되는 전용면적별로 시세와 세대수를 제공한다. 이에 본 연구에서 활용한 평형대 기준(소형: 전용면적 60㎡ 이하, 중형 : 전용면적 60㎡ 초과 85㎡ 이하, 대형: 전용면적 85㎡ 초과)에 따라 각 평형별 세대수를 가중하여 개별 아파트 단지의 평형대별 시세지수를 산정하였다. 시세를 지수화하는 단계에 있어 기준 시점은 앞 절에서 산출한 실거래가 지수와 같이 2006년 1월을 기준시점(지수값: 100)으로 하여 산정하였다.

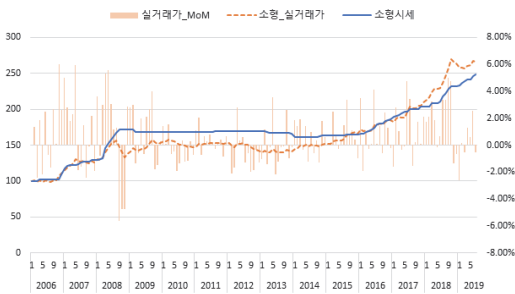
중화동 내 세대수와 거래사례가 가장 많았던 한신아파트를 대상으로 실거래가 기반 시세지수와 부동산114 시세지수를 비교하였다. 한신아파트의 경우 기준에 따라 소형과 중형 평형대가 존재하며 소형에 해당하는 세대수가 892세대, 중형이 652세대로 구성되어 있다.

<그림 10> 한신아파트 평형대별 시세지수

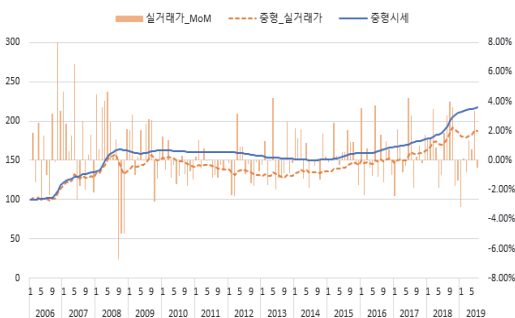


<그림 10>은 한신아파트의 평형별 세대수를 가중한 시세지수이다. 한신아파트의 경우 모든 세대가 소형(60㎡ 이하), 중형(60㎡ 초과 85㎡ 이하)로 구성되어 있다. 위의 시세지수와 각 평형대별로 산출한 실거래가 지수를 비교할 수 있다.

<그림 11> 한신아파트 소형 평형대 지수 비교



<그림 12> 한신아파트 중형 평형대 지수 비교



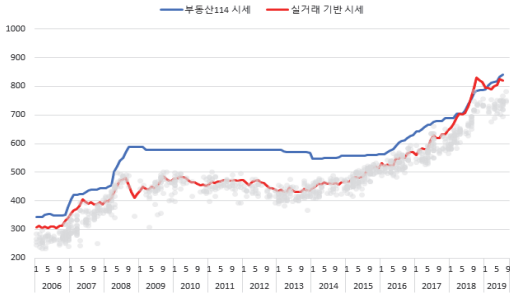
<그림 11>과 <그림 12>는 각각 한신아파트의 소형 및 중형 실거래가 기반 시세지수와 부동산114의 시세 지수를 비교한 그림이다. 종합적으로 비교 시 실거래 가지수와 시세지수가 분석기간 전반에 걸쳐 유사하게 움직이는 것을 확인할 수 있다. 또한 각각의 그림에서 실거래가 기반 시세지수가 가격변동률이 변화하는 양상을 포착한다는 점에서 시세지수 대비 갖는 장점을 확인할 수 있다. 구체적으로 살펴보면 <그림 11>에서 2008년 10월에 2008년 1분기 이후의 최저점을 기록하고 다시 상승되어 약 7년 간 횡보하는 움직임을 실거래가 지수에서 확인할 수 있다. 이에 반해 시세 지수의 경우 2008년 9월 경 부터 하락이 관측 되지 않고 횡보하는 움직임을 보인다. 또한 2018년 9월부터 보이는 2019년 4월 경 까지 나타난 하락의 움직임이 시세지수에서는 포착되지 않는 점을 확인할 수 있다. 이는 그림 13의 한신아파트 중형 평형대의 지수 비교에서도 동일하게 확인된다.

6. 실거래 기반 시세 개발

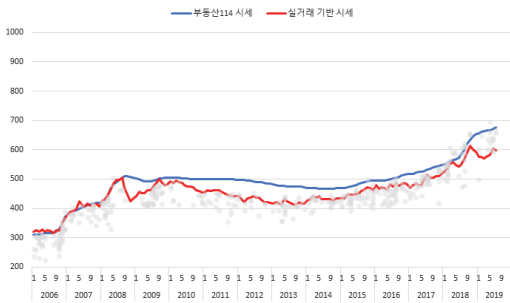
위에서 산출한 실거래가 기반 시세지수에 특정 시점의 실거래 사례의 평균가격을 대입하여 시세를 작성할 수 있다. 시점 선택은 2건 이상의 실거래 사례가 관측되고 거래가격 간 편차가 가장 작은 월을 기준으로 하여 한신아파트의 소형 평형대의 경우 2008년 8월, 중형의 경우 2013년 9월의 평균 실거래 가격을 활용하였다. 시세의 단위는 ㎡ 당 만원의 가격을 나타내며 비교에 활용된 부동산114의 시세 또한 해당하는 단위 면적당 만원의 가격으로 나타내었다.

<그림 13>과 <그림 14>는 각각 평형대별로 산출한 실거래가 기반 시세와 부동산114 시세를 비교한 그림이다. 비교를 위해 분석 월 별로 관측된 실거래사례를 산점도(회색)로 표현하였다. <그림 13>과 <그림 14>를 보면 실거래 기반 시세가 실거래 사례들의 움직임을 상대적으로 잘 포착하고 있음을 확인할 수 있다. 특히 소형 평형대의 시세를 보면 분석기간 중 대부분의 경우 부동산114 시세는 실거래 사례의 가격수준을 제대로 포착하고 있지 못하지만 실거래 기반 시세는 가격수준과 가격 움직임을 무리 없이 나타내고 있다.

<그림 13> 한신아파트 소형 평형대 시세

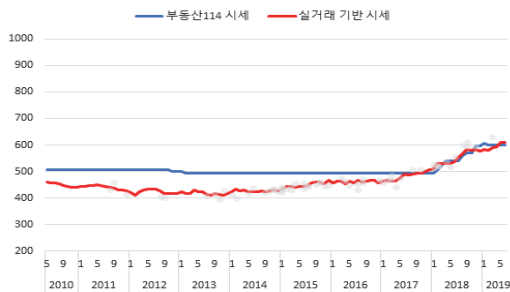


<그림 14> 한신아파트 중형 평형대 시세



추가적으로 세대수가 적고 거래사례가 적은 단지의 특정 평형대에서의 비교를 위해 동양엔파트 중형 평형대를 통해 비교하였다. 동양엔파트는 전체 103세대이며, 분석기간 중 60건의 실거래 사례가 존재하였으며 모든 세대는 전용면적 84㎡로 중형에 해당한다. 시세 지수에 적용한 평균 실거래 가격은 2015년 1월의 거래 사례를 활용하였다.

<그림 15> 동양엔파트 중형 평형대 시세



<그림 15>는 분석기간 중 가장 거래사례가 적었던 동양엔파트 중형 평형대의 시세 비교를 나타내며 위의 한신아파트 사례와 마찬가지로 부동산114의 시세가

가격 움직임이나 가격수준을 대표하지 못하는 것에 반해 실거래 기반 시세는 상대적으로 잘 표현해냄을 확인할 수 있다.

최종적으로 실거래 기반 시세지수를 바탕으로 시세를 개발하고 이의 활용 가능성을 살펴보기 위해 구 단위에서 동 단위 그리고 더 세분화하여 동 내부의 단지 와 평형대로 계층을 나누어 지수를 산출하고 시세를 개발하였다. 이렇게 산출한 실거래 기반 시세는 기존 조사 기반 시세 대비 가격변화의 양상과 가격수준을 잘 파악할 수 있다는 장점을 지닌다. 이를 통해 계층모형을 활용하여 특정 단지, 특정 평형대의 지수를 구성할 수 있고 지수를 기준으로 실거래 가격을 대입하여 효율적인 방식으로 시세를 구축할 수 있다는 시사점을 제시한다.

VI. 결론

본 연구는 국토교통부 실거래가 공개시스템에서 아파트 실거래가 자료를 이용하여 아파트 실거래가격지수를 산정하여 시세지수로의 활용가능성을 제시한다. 구, 동 단위 지수에서 세분화하여 거래 사례가 적은 동 내 단지 단위, 평형대 단위에서의 안정적인 지수 산정을 시도하였다. 이 과정에서 반복매매모형에 베이지안 추론 방법을 도입하여 RWD(Random Walk with Drift)모형을 활용하였다. 또한 상대적으로 더 거래사례가 부족한 개별 아파트 단지별, 평형대별 지수를 추정하기 위해 Francke and Van de Minne (2017)이 활용한 계층모형을 도입한 베이지안 추론으로 계층적 RWD(HRWD, Hierarchical Random Walk with Drift)모형을 적용하였다. 또한 각 지수들의 산출을 위한 가정을 변동률 산정방식 등을 통해 변경하면서 성능(MSEI, SI)을 비교하였으며 부동산114에서 제공하는 시세에 세대수를 가중한 지수를 산출하여 함께 비교해보았다. 또한 실거래 기반 시세에 실거래 가격을 적용하여 시세를 산출하여 기존 조사 기반 시세 대비 갖는 장점과 차이점을 설명하였다.

자산시장의 시장참여자들은 자산의 가격변동에 따라 각기 다른 행동을 취한다. 자산보유자는 보유기간을 늘리거나, 매각을 결정할 수 있다. 자산수요자는 현 시점에서 매입을 결정할 수 있고 현 시점 가격 대비

더 낮은 가격으로 매수하기를 기다릴 수 있다. 이처럼 효율적인 의사결정이 도출되려면 가격의 움직임에 대해 보다 신뢰성 있게 추정될 필요가 있다. 나아가 지수의 추정 범위가 시, 구, 동 단위의 넓은 범위가 아닌 세분화된, 자산군 별 지수(Asset-Class Specific)가 필요하다.

이와 같은 상황에서 본 연구가 갖는 의의는 다음과 같다. 먼저, 거래사례가 부족한 하부시장을 대상으로 계층적 베이지안 추론을 통해 실거래사례를 비교적 정확히 반영하는 실거래가 기반 시세를 개발하였다. 둘째, 계층의 구성을 단지별, 평형대별로 세분화하여 관측할 수 있는 작은 단위까지 분석을 시도하였다. 마지막으로 실거래가 기반 시세가 기존 조사 기반의 시세가 갖는 한계점을 개선할 수 있음을 확인하였다. 이를 통해 조사 대비 효율적인 방식으로 거래가격과 가격움직임을 상대적으로 신뢰성 있게 포착할 수 있는 시세를 산출해낼 수 있음을 확인하였다.

그럼에도 여전히 본 연구는 몇 가지 한계점을 지닌다. 첫 째, 개별 동 자료만을 활용한 단지별 월 단위 지수의 경우 상대적으로 안정성과 통계적 신뢰도가 낮게 산출된다. 개별 자료를 활용한 분기 단위 지수는 월 단위 지수 대비 비교적 안정적인 성능을 보였지만 이는 가격의 기준시점 대비 변동률이 적시성을 보장하지 못할 수 있다는 문제점을 지닌다. 둘째, 실거래가 기반 시세의 경우 실거래가 지수를 바탕으로 작성되어 그 변동성이 다소 크게 나타난다. 시세는 실거래가 지수와 달리 매 분석 단위마다의 움직임을 모두 포착해낼 필요는 없다. 셋 째, 반복매매모형 구성에 있어 완벽하게 동일한 자산의 반복매매를 관측한 것이 아닌 가정을 통해 동일 범주로 묶이는 자산의 반복매매를 관측하였다. 마지막으로 개별 아파트 단지에서도 각 동별, 각 호별 시세의 산정까지는 이루어지지 않았다. 일반적으로 같은 아파트 단지 내에서도 조망권, 접근성 등에 의해 다른 가격 수준이나 움직임을 가질 수 있다. 이에 따라 동별, 호별 시세 산정은 추후의 연구에서 진행되기를 기대한다.

논문접수일 : 2022년 11월 10일

논문심사일 : 2022년 11월 14일

게재확정일 : 2022년 12월 15일

참고문헌

- 권재욱 · 김호철, “변동모수모형을 이용한 아파트가격지수 추정에 관한 연구”, 「도시행정학보」 제19집 제1호, 한국도시행정학회, 2006, pp. 175-200
- 김경외 · 김영효, “모델의 불확실성을 반영한 아파트가격지수 예측 모형 연구: BMS, BMA를 중심으로”, 「부동산 분석」 제1권 제1호, 한국부동산원, 2015, pp. 27-49
- 김재익 · 김달호 · 서안나 · 조애정, “반복매매모형에 기반한 아파트 실거래가격지수의 산정방식 비교: 표본산정방식과 하위지수 병합방식을 중심으로”, 「주택연구」 제23권 제4호, 한국주택학회, 2015, pp. 47-71
- 류강민 · 이상영, “S&P Case-Shiller 반복매매모형을 이용한 주택가격지수 산정에 관한 연구”, 「주택연구」 제18권 제2호, 한국주택학회, 2010, pp. 183-204
- 류강민 · 송기욱, “반복매매모형을 활용한 서울시 오피스 벤치마크 가격지수 개발 및 시험적 적용 연구”, 「LHI Journal」 제11권 제2호, LH 토지주택연구원, 2020, pp. 33-46
- 박헌수, “거래빈도가 낮은 시장에서의 실거래 부동산 가격지수 작성에 관한 연구- 강남구를 대상으로”, 「부동산학연구」 제13권 제3호, 한국부동산분석학회, 2007, pp. 187-200
- 송영선 · 윤명탁 · 이창무, “아파트 하위시장 실거래가 지수 산정방식 비교 연구”, 「부동산 분석」 제6권 제3호, 한국부동산원, 2020, pp. 1-19
- 송영선 · 이창무, “베이저안 추세 모형을 활용한 서울시 오피스 매매가격지수 개발”, 2021 한국부동산분석학회 상반기 온라인 학술대회
- 우남교 · 권범준, “베이저안 추론방법을 이용한 소지역 주택 매매가격지수 추정”, 「부동산 분석」 제2권 제1호, 2016, 한국부동산원, pp. 1-16
- 이상경, “서울시 오피스 매매가격지수 개발에 관한 연구”, 「서울도시연구」 제6권 제4호, 서울연구원, 2005, pp. 121-134.
- 이성현, “3수준 위계선형모형을 통한 주택가격의 형성구조에 관한 연구”, 「한국지역개발학회지」 제27권 제3호, 한국지역개발학회, 2015, pp. 27-41
- 이창무 · 배익민, “시세가격을 활용한 아파트 실거래가 반복매매지수 산정”, 「부동산학연구」 제14권 제2호, 한국부동산분석학회 2008, pp. 21-37
- 이창무 · 김병욱 · 이현, “반복매매모형을 이용한 아파트 매매가격지수”, 「부동산학연구」 제13권 제2호, 한국부동산분석학회, 2002, pp. 1-19
- 차승호, “베이저안 추세 반복매매모형을 활용한 실거래가 기반 토지가격지수개발에 관한 연구”, 한양대학교 대학원 석사학위논문, 2020
- 최성호 · 류강민 · 이건우 · 이창무, “반복매매모형을 활용한 오피스 매매가격지수에 관한 연구”, 「국토계획」 제45권 제7호, 대한국토·도시계획학회, 2010, pp. 119-131
- 황규완 · 손재영, “MIT/CRE 2단계 추정법을 활용한 서울 오피스 가격지수 산출에 관한 연구”, 「주택연구」, 제25권 제1호, 한국주택학회, 2017, pp. 151-175
17. 마쓰우라 겐타로, 「데이터 분석을 위한 베이저안 통계 모델링 with Stan & R」, 길벗, 2019
18. M.J. Bailey, R.F. Muth and H.O. Nourse, “A Regression Method for Real Estimate Price Index Construction,” Journal of the American Statistical Association, Vol.58 No.304, 1963, pp. 933-942
19. M.K. Francke and A Van de Minne, “The hierarchical repeat sales model for granular commercial real estate and residential price indices,” Journal of Real Estate Finance and Economics, Vol. 55 No. 4, 2017, pp. 511-532
20. M.K. Francke, “Repeat Sales Index for Thin Markets,” The Journal of Real Estate Finance and Economics, Vol. 41, 2010, pp. 24-52
21. M.K. Francke and G.A. Vos, “The hierarchical trend model for property valuation and local price indices,” Journal of Real Estate Finance and Economics, Vol. 28, 2004, pp. 179-208
22. Geltner and Bokhari, “A Technical Note on Index Methodology Enhancement by Two-Stage Regression Estimation,” MIT-CRE Commercial Real Estate Data Laboratory, 2008
23. W.N. Goetzmann, “The Accuracy of real estate Indices: Repeat Sale Estimators,” Journal of Real Estate Finance and Economics, Vol. 5, 1992, pp. 5-53
24. Richard McElreath, 「Statistical Rethinking: A Bayesian Course with Examples in R and Stan, Second Edition」, 2020
25. R.J.Shiller, “Arithmetic Repeat Sales Price Estimators,” Journal of Housing Economics, Vol. 1, 1991, pp. 110-216
26. 부동산114/ <https://www.r114.com/>

<국문요약>

계층적 베이지안 추론을 통한 아파트 단지별 실거래 기반 시세 개발

권 민 성 (Kwon, Min-Sung)

최 우 현 (Choi, Woo-Hyun)

송 영 선 (Song, Young-Sun)

이 창 무 (Lee, Chang-Moo)

부동산 가격지수는 기준시점 대비 가격 상승률을 확인할 수 있는 지표이다. 부동산은 소비재이면서 동시에 투자재의 성격을 가지는 자산이기에 그 가격 상승률은 다양한 경제주체들의 의사결정을 이끌어낸다. 아파트로 대표되는 국내 주거용 부동산은 지역별, 지역 내 단지별 하부시장의 집합체로 나뉠 수 있다. 그러나 전통적인 OLS방식으로 산출되는 실거래가격 지수는 거래사례가 존재하지 않는 시점을 추정하지 못하여 하부시장의 가격동향을 정확하게 포착하지 못하는 단점이 있다. 더불어 가격지수로 가장 많이 활용되는 반복매매모형은 동일하게 인식되는 자산의 거래쌍을 만들어 활용하기에 거래사례의 한계 문제점은 더욱 부각되고 있다. 이와 같은 단점에서 탈피하여 Markov Chain Monte Carlo 샘플링을 방식으로 하는 계층적 베이지안 추론법을 활용하여 안정적인 하부시장 별 가격지수를 산출해낼 수 있다. 따라서 본 연구에서는 중랑구 망우동, 중화동의 아파트 단지별 가격지수를 일차적으로 산출하였다. 이를 바탕으로 시세를 구축하여 부동산 114에서 제공하는 시세와의 비교를 진행하였다.

주 제 어 : 반복매매모형, 실거래가지수, 베이지안 추론, MCMC 샘플링, 아파트 시세