

아파트 매매 실거래 가격지수에 관한 다각적 분석

A Study on the Apartment Sales Price Index by
Applying Various Methods of Time Series Analysis

작성자 통계학과 2019310508 김수빈
통계학과 2019310157 박지성

목 차

초록	3
키워드	3
I 서론	4
1. 주제 선정 배경 및 연구 목적	4
2. 선행연구 검토 및 차별성	4
3. 분석모형 및 분석자료	5
II 기간별 분석	6
1. 기간 구분 기준	6
2. 기간별 시계열 분석	7
3. 두 모델 간 시계열 특징 비교	9
III 단변량 분석	9
1. ARIMA 모형	9
2. ARIMA + 변동성 모형	10
IV 다변량 분석	12
1. 변수 선정	12
2. VAR 모형	14
V 결론	16
< 참고문헌 >	17

아파트 매매 실거래 가격지수에 관한 다각적 분석

A Study on the Apartment Sales Price Index by Applying Various Methods of Time Series Analysis

통계학과 2019310508 김수빈

통계학과 2019310157 박지성

초 록

본 연구는 2006년부터 2021년까지의 아파트 매매 실거래가격지수를 예측하는 모형을 구축하고자 한다.

다각적 인사이트를 도출하기 위해 다양한 시계열 모형이 사용되었다. 우선 기간을 시계열 특성이 다른 두 기간으로 나누어 (2006~2012, 2013~2021), 이후 ARIMA모형을 적용했다. 전 범위 기간에 대한 ARIMA모형, 그리고 모형의 변동성을 고려하기 위한 ARIMA와 GARCH를 결합한 모형도 적용했다. 마지막으로, 아파트 매매 실거래가격지수에 영향 미치는 중요 변수들을 선정하여 다변량 분석을 VAR모형을 통해 진행했다.

예측한 결과, 전범위 기간에 대해 2차 차분하여 ARIMA모형을 적합한 모델의 정확도가 제일 높게 나왔다. 2006년부터 2021년까지 서울시 내 아파트 매매 실거래 가격의 주요 변동은 금융위기와 부동산 투기 제한 정책에 의한 아파트 수요가 감소됨에 따른 아파트 실거래 매매 가격의 하락으로 인해 발생하였음을 깨달을 수 있었다.

키워드

아파트 매매 실거래가격지수 예측, 시계열 모형, ARIMA, GARCH, VAR

Abstract

This study aims to construct a model that forecasts the household price using the apartment price index from year 2006 to 2021 issued by the Korea Real Estate Board.

To retrieve collective insights, various time series models were used. First, the time period was divided according to time series trends (2006~2012, 2013~2021) and ARIMA model was applied. Then, ARIMA model was applied to the whole time period. Additionally, ARIMA with GARCH model variation was applied to deal with the residuals in the time series. Finally, important variables that affected the household prices were included to apply the VAR model which deals with multivariate cases.

According to the forecast result, ARIMA model that was differenced twice resulted as the model with highest accuracy. It is assumed that the variances in household prices were caused by financial crisis and real estate policies which decreased demands for apartments.

Keywords

Apartment price index forecasts, time series analysis, ARIMA, GARCH, VAR

I. 서론

1. 주제 선정 배경 및 연구 목적

2008년도에 미국의 서브프라임모기지 금융위기로 인해서 우리나라 금융시장은 충격을 받아 주택시장의 가격하락이 발생했다. 이에 따라서 전세가격은 상승하여 주택시장에 여러 가지 문제점이 발생했다. 여러 가지 부동산제도의 도입과 경제 사건 등으로 부동산 시장에 많은 변화를 가져왔다. 특히, 아파트 거래는 큰 폭의 가격변동으로 인해 주택시장은 앞으로 계속 많은 변화를 겪을 것이며 불규칙적일 것으로 예상된다.

따라서, 본 논문의 연구 목적은 변동이 잦은 부동산 시장 파악의 필요성을 느껴 여러 가지 시계열 방법을 이용하여 아파트 가격을 예측하고자 한다. 그리하여 시계열 특성이 다른 두 기간을 구분하여 기간별 특성을 파악하고 단변량 분석 및 예측을 진행하고자 한다. 추가로, 주택가격에 영향을 주는 여러 설명변수를 사용하여 다변량 분석 및 예측을 진행하고자 한다. 그리하여 마지막에 모델들 간 비교 분석을 통해 인사이트를 도출하고자 한다.

2. 선행연구 검토 및 차별성

한국의 주택시장에 대한 시계열적 연구는 대부분 서울시를 대상으로 한 아파트 매매가격 혹은 아파트 매매가격 지수를 통해 진행되었다. 더불어 예측을 위해선 ARIMA모형을, 다변량 시계열 예측을 위해 VAR모형을 대다수 사용하였다. 본 연구와 관련된 주요 논문들을 살펴보면 다음과 같다.

김근용(1988)은 ARIMA모형과 state-space모형을 통해 전국 32개 도시의 평균 월별 아파트 매매 및 전세가격지수의 변동을 예측하고자 하였다. State-space모형은 모형 내에서 추세와 사이클에 대한 정보를 반영한다는 점에서 차분으로 인해 추세와 사이클이 상실되는 ARIMA모형의 한계를 보완한 모형이라고 할 수 있다. 분석 결과, 매매가격지수에 대해선 ARIMA모형의 성능이 높았던 반면, 추세가 매매가격지수에 비해서 분명한 전세가격지수에선 state-space모형의 성능이 높음을 확인하였다. 하지만 state-space 모형의 경우 장기 전망 및 시장구조 변화에 대한 예측에는 한계점이 있음을 시사했다.

김동환(2015)는 여러 거시경제 변수 지표와 VECM모형을 활용하여 토지시장에 대한 단기 예측모형을 형성하고자 하였다. 거시경제 변수 지표에 대해 그래인저 인과관계 검정과 공적분 검정을 실시하였고 이를 통해 VECM모형을 최종 모델로 선정하여 향후 8분기에 대한 지가변동률을 예측하였다. 그 결과 전국평균지가 변동률과 회사채 수익률, 실질 GDP성장률이 인과관계가 있는 것으로 파악하였고 수요자의 기대심리와 같은 계량화되기 어려운 요인들을 반영하지 못했다는 점을 한계로 지적하였다.

조보근(2020)은 ARIMA, LSTM, 랜덤포레스트 그리고 LIME모형을 활용하여 서울, 부산, 인천의 아파트 실거래 가격지수를 예측하고자 하였다. 이들은 LIME모형을 활용해 지역 별 부동산 가격변동에 영향을 주는 주요 변수를 파악하고자 하였다. 해당 연구에선 통화량, 지가지수와 같은 거시경제지표뿐만 아니라 미시경제지표로써 구글 트렌드 검색지수도 활용하였고 해당 검색지수는 서울, 부산의 부동산 가격변동의 중요한 지표로써 파악하였다. 전반적인 모델의 성능은 LSTM, ARIMA, 랜덤포레스트 순으로 나타났고, ARIMA모형은 전반적 추세를 예측하는 데에는 적합하나, 실제 가격이 상승/하락 정도를 예측하는데에는 한계가 있음을 명시했다.

앞선 여러 선행연구와의 차별성으로 해당 연구는 서울시 아파트 매매 실거래가 지수를 지표가 지니는 시각적인 특성에 따라 2013년을 기준으로 두 기간으로 나누어 기간별 시계열적 특성을 파악해보고 한

다. 더불어 전 기간에 대하여 ARIMA모형 뿐만 아니라 잔차에 대해 변동성 모형을 결합하여 여러 형식의 시계열 모형에 대한 성능을 비교해보고자 한다. 최종적으로 거시경제 지표를 활용하여 다변량 시계열 분석을 진행해보므로써 각 거시경제 변수의 충격에 따른 서울시 아파트 매매 실거래 가격 지수의 영향 정도를 살펴볼 것이다.

3. 분석모형 및 분석자료

(1) 분석모형

1) ARIMA 모형

Auto Regressive Integrated Moving Average (ARIMA) 모형 혹은 Box-Jenkins 모형은 단기예측에 많이 이용되는 시계열 모형이다. 비정상 시계열을 $ARIMA(p, d, q)$ 로 표시되는데 d 는 몇 회차 차분을 적용시켜 정상 시계열로 전환시킨 모형을 $ARMA(p, q)$ 로 표시한다 (박유성, 허명희, 1998).

2) ARCH, GARCH 모형

Auto-Regressive Conditional Heteroskedasticity (ARCH) 모형은 가격의 변동성에서 흔히 관찰되는 시계열 의존성을 모형화한 모형이다. 이 모형은 분산이 시점에 의존하므로 해당 시점의 전시점까지의 정보를 이용하여 변동성을 측정하는 조건부 분산이 예측 오차를 줄일 수 있다.

Generalized ARCH (GARCH) 모형은 $ARCH(p)$ 모형을 추정할 때 시차 (p)를 크게 설정하는 경향을 보완하는 모형이다. GARCH 모형은 ARCH 모형을 일반화하여 변동성의 시계열 의존성, 즉 자기상관을 표현하는 데 있어서 모수의 수를 줄일 수 있다는 장점을 가지고 있다 (고봉현, 2007).

3) VAR 모형

Vector Autoregressive (VAR) 벡터자기회귀모형은 단변량 자기회귀모형을 다변량 자기회귀모형으로 확장시킨 모형으로 예측 및 내생변수의 변화에 따른 효과분석 등과 관련하여 자주 활용되는 모형이다. 단변량 자기회귀모형은 예측하고자 하는 변수에 영향을 미칠 수 있는 다른 중요변수와 그런 변수들과의 관계를 고려하지 않아 일차원적인 분석이 진행될 수 있다. 이러한 한계점을 VAR 모형을 통해서 보완하고자 한다. VAR 모형에서 충격반응분석 (impulse response analysis)을 통하여 어떠한 한 변수의 변화가 내생변수에 미치는 동태적 효과를 파악할 수 있고, 분산분해 (variance decomposition)을 통하여 각각의 내생변수의 변동 중에서 이들이 전체 변동에 기여한 부분의 상대적 크기를 분석할 수 있다 (문권순, 1997).

(2) 분석자료

본 논문은 주택가격을 예측하기 위해 한국부동산원 부동산통계정보시스템에서 제공한 아파트 매매 실거래가격수를 사용했다. 이는 전국의 아파트 및 연립과 다세대를 대상으로 하여 실제 거래되어 신고된 아파트 및 연립과 다세대의 거래 가격수준이다. 자료는 2006년 1월 1일부터 2012년 12월 1일까지 데이터 수집했다. 다변량 분석을 진행하기 위해 아파트 매매 실거래가에 영향을 줄 수 있는 변수를 선정하였다. 해당 변수들은 소비자 물가지수, 서울시 지가지수, 통화량 (M2), 가격대출 금리, 그리고 주가지수 (KOSPI)를 사용했다.

<표 1> 변수 목록

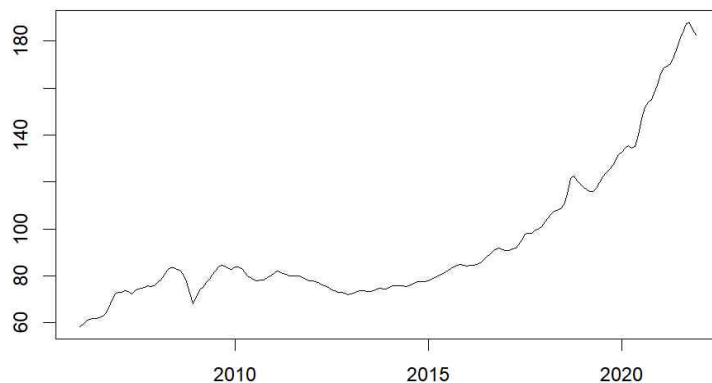
구분	변수	변수명
종속변수	아파트 매매 실거래가격지수	Y
경제지표	소비자 물가지수	consumer
	지가지수 (서울시)	land
	통화량 (M2)	M2
	가격대출 금리	loan
	주가지수 (KOSPI)	KOSPI

출처: 아파트 매매 실거래가격지수, 서울시 지가지수, 가격대출 금리는 한국감정원 부동산통계정보 (<http://r-one.co.kr>), 소비자물가지수, 통화량 (M2), 주가지수 (KOSPI)는 통계청 (<http://kosis.kr>)

<표 2> 데이터 기초 통계량

	최소값	최대값	평균	중앙값
아파트 매매 실거래가격지수	58.50	187.80	94.02	81.25
소비자 물가지수	75.23	104.04	91.51	93.57
지가지수 (서울시)	64.28	107.02	81.41	77.42
통화량 (M2)	10272336	3619578	2070473	1933582
가격대출 금리	2.740	7.560	4.673	4.525
주가지수 (KOSPI)	1063	3297	1987	1985

II. 기간별 분석



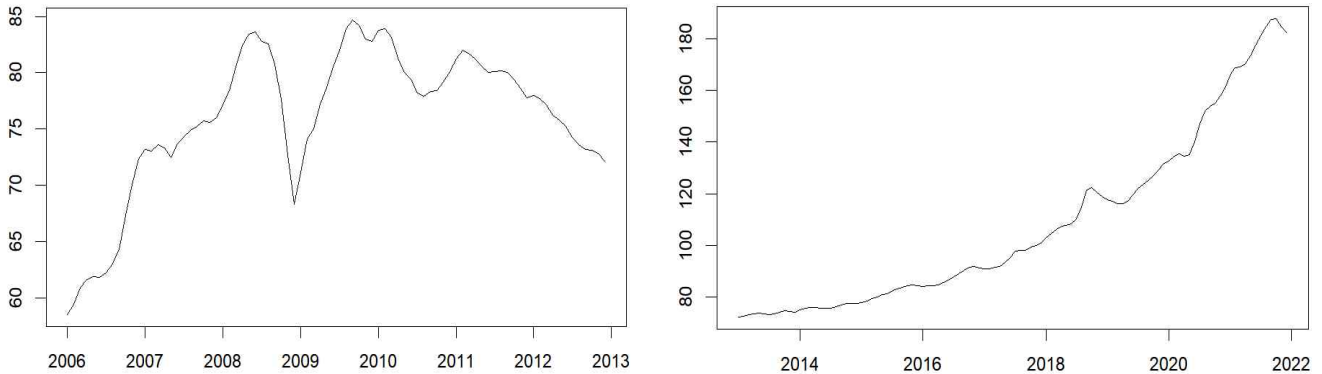
<그림 1> 연도별 아파트 매매 실거래가격지수 그래프

1. 기간 구분 기준

분석 기간을 기간 1 (2006년~2012년) 그리고 기간 2 (2013년~2021년) 기준으로 나눴다. 시계열 그래프를 그렸을 때 그래프의 추세가 초반에는 안정적이었다가 2008년에 금융위기를 겪은 후 다시 안정성을 되찾으며 아파트 매매가격은 점점 올라가는 추세를 가지게 된다. 기간 1 (2006년~2012년)을 안정적인 기간으로 보고, 기간 2 (2013년~2021년)을 불안정적인 기간으로 대략 간단하게 정의할 수 있다. 더 구체적으로 정의하면, 기간 1은 2008년에 금융위기로 인하여 아파트 매매가격지수가 급하락한 것을 확인할 수 있다. 그렇지만 금방 경향성을 회복하였다는 점에서 정상성 특징을 지닐 수 있을 것으로 예상

한다. 반면 기간 2는 지속적으로 상승하는 경향성을 나타낸다. 따라서, 이런 다른 특징을 가지고 있는 두 기간을 시장상황에 따라 구분하여 분석하여 모형의 예측력을 비교하고자 한다.

2. 기간별 시계열 분석



<그림 2> 기간 1와 기간 2의 시계열 그래프

(1) 단위근 검정

시계열 분석에 들어가기 앞서, 본 자료가 정상적인 시계열인지 검정하고, 만약 아니라면, 차분을 통한 정상성 변환까지 하는 것이 첫 번째 단계다. 이를 하기 위해서는 단위근 검정을 실시한다. 단위근 검정은 단위근 존재 여부를 이용하여 시계열의 정상성을 확인하는 방법이라고 할 수 있다. 단위근 검정 방법 중 Augmented Dickey-Fuller (ADF) 검정과 Kwiatkowski-Phillips-Schmidt-Shin (KPSS) 검정을 사용했다. ADF 검정 방법의 귀무가설은 ‘시계열은 단위근을 포함한다’ 즉, ‘정상 시계열이 아니다’를 의미한다. 따라서, p-value가 유의수준 0.05보다 값이 크면 귀무가설을 기각하지 못하고, 해당 시계열은 비정상 시계열이라고 정의할 수 있다. 반면, KPSS 검정 방법은 ADF와 반대로 귀무가설은 ‘시계열은 단위근을 포함하지 않는다’ 즉, ‘해당 시계열은 정상 시계열이다’를 의미한다 (Kwiatkowski, Phillips, Schmidt, & Shin, 1992). 따라서, p-value가 유의수준 0.05보다 값이 작으면 귀무가설을 기각하여, 해당 시계열은 비정상 시계열이라고 정의할 수 있다.

<표 3> 기간별 단위근 검정 결과

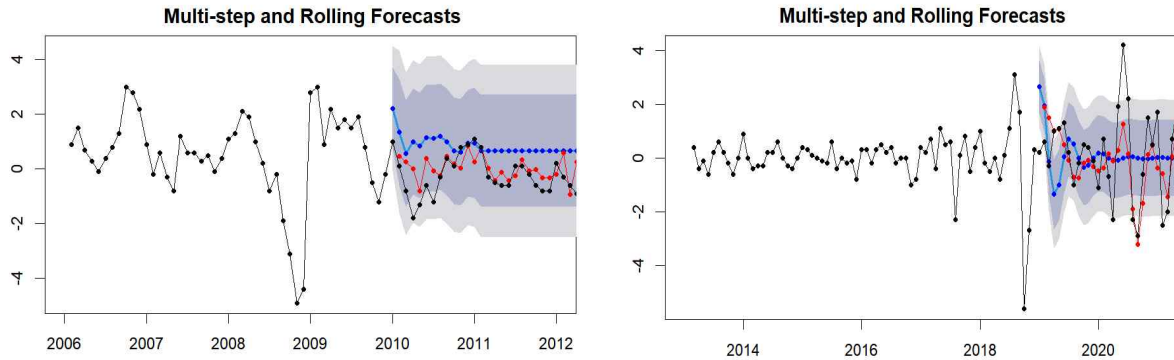
	기간 1		기간 2	
	검정통계량	p-value	검정통계량	p-value
ADF	-2.0533	0.554	-0.46023	0.9821
KPSS	0.8976	0.01	2.0362	0.01

위 결과에 따르면 두 시계열은 비정상 시계열인 것을 확인했으며, 정상성 변환이 필요하여 차분을 진행했다. R 패키지에 있는 ndiffs함수를 통해서 적당한 차분 횟수를 정할 수 있었다 (R Documentation, 2022). 확인 결과, 기간 1 시계열은 1차 차분이 제일 적당한 차분 횟수로 나왔다. 반면, 기간 2 시계열은 2차 차분이 제일 적당한 차분 횟수로 나왔다. 이후 차분된 시계열을 ADF와 KPSS 검정을 통해 정상적인 시계열인 것을 확인하고 모델링을 진행할 수 있었다.

(2) 시계열 분석 모델링

최종 모델은 auto.arima함수를 통해서 선정됐다. 해당 함수는 AIC, AICc, BIC값을 바탕으로 최적의

ARIMA 모델을 출력하는 함수다 (R Documentation, 2022). 이 방법을 통해서 기간 1 시계열은 AIC값 155.32, AICc값 156.27, 그리고 BIC값 162.72을 가지고 있는 ARIMA(0,0,1)(0,0,1)[12]을 추정하였다. 그리고 기간 2 시계열은 AIC값 169.6, AICc값 170.21, 그리고 BIC값 178.59을 가지고 있는 ARIMA(2,0,1)을 추정하였다. 그리고 추가적으로 잔차의 계열 상관성 여부를 ggtsdiag함수를 통해서 검정하였으며 해당 모델이 잔차의 계열 상관성이 없다는 것을 확인하였다. 따라서, 기간 1은 ARIMA(0,0,1)(0,0,1)[12] 그리고 기간 2는 ARIMA(2,0,1)을 최종 모형으로 선정하였다.



빨간 선: multi-step forecast

파란 선: rolling forecast

<그림 3> 기간 1과 기간 2의 Forecast 그래프 결과

<표 4> 기간별 모델링 결과

	기간 1	기간 2
AIC	155.32	169.6
AICc	156.27	170.21
BIC	162.72	178.59
최종 모형	ARIMA(0,0,1)(0,0,1)[12]	ARIMA(2,0,1)

(3) 예측 결과

여러 time-step을 예측하는 multi-step forecast와 rolling window를 사용하여 예측하는 rolling forecast 두 가지 방법을 사용하여 기간별 아파트 매매실거래가격지수 예측을 진행했다. 우선 기간 1에는 2009년까지 학습 데이터로 정해주고, 2010년 이후 데이터를 시험 데이터로 정해줬다. 그리고 기간 2에는 2018년까지 학습 데이터로 정해주고, 2019년 이후 데이터를 시험 데이터로 정해줬다. 기간별로 multi-step forecast와 rolling forecast 진행했을 때 예측 정확도를 비교한 결과는 다음과 같다.

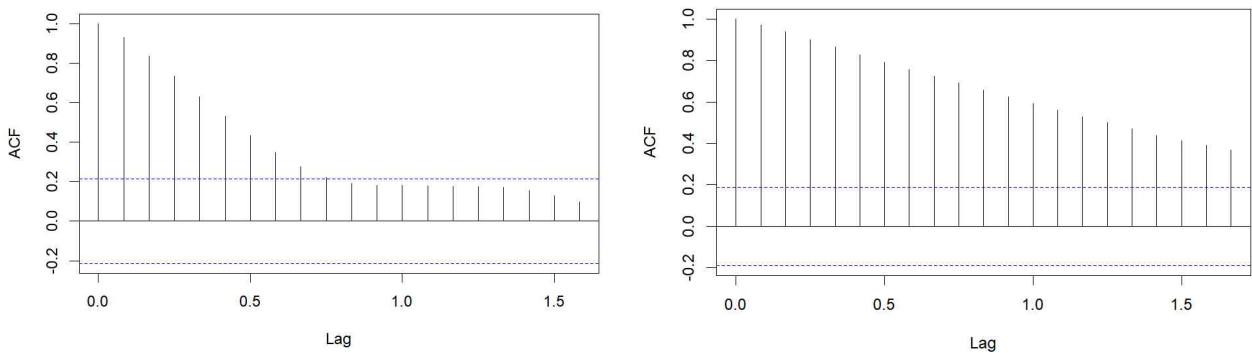
<표 5> 기간별 모델 예측 정확도 결과 비교

	기간 1		기간 2	
	Multi-Step	Rolling	Multi-Step	Rolling
RMSE	1.27316	0.6627419	1.27316	0.6178813

Root-Mean-Square Error (RMSE)는 평균 제곱근 오차이고 이는 잔차의 표준 오차이다. 이는 잔차가 회귀선 데이터 포인트에서 얼마나 멀리 떨어져있는지 측정하며 이러한 잔차들이 얼마나 분산되어 있는지 측정한다. 따라서, RMSE 값이 낮을수록 모델의 예측력이 높다는 의미가 있다. 두 기간 모두 rolling forecast RMSE가 multi-step RMSE 값보다 작다. 즉, 두 기간 모두 rolling forecast 예측 정확도 multi-step forecast 예측 정확도가 높다는 의미다.

3. 두 모델 간 시계열 특징 비교

기간을 나누는 두 모델 간 시계열 특징을 파악하고자 한다. 기간 1 (2006년~2012년)은 shock이 존재하나 회복하는 양상을 띈다. 반면 기간 2 (2013년~2021년)은 지속적인 상승 추세를 띄고 있다. 두 기간은 계절성이 없으므로 이를 따로 고려할 필요가 없었다. 그리고 <그림 4> Autocorrelation Function (ACF) 자기상관함수 그래프를 확인했을 때 기간 1은 0으로 느리게 감소하며, 기간 2는 0으로 더 느리게 감소하는 것을 확인할 수 있다. 이는 정상성을 의미하는데 자기상관함수가 0으로 느리게 감소할수록 이는 비정상성 특징을 강하게 띈다는 의미를 가진다. 이를 바탕으로, 기간 2가 기간 1보다 더 강한 비정상성을 가진다고 할 수 있다. 따라서, 기간 1은 1차 차분하여 정상성을 확보했지만, 기간 2는 2차 차분을 추가적으로 진행하여 정상성 확보 후 예측을 진행할 수 있었다.



<그림 4> 기간 1와 기간 2 ACF 그래프

III. 단변량 분석

주어진 데이터의 전 기간에 대하여 단변량 예측을 실시해보고자 한다. 이때 2006년 1월~2018년 12월(156개월)을 학습 데이터, 2019년 1월~2021년 12월(36개월)을 시험 데이터로 설정하였다.¹⁾ 주어진 학습 데이터 범위에 대하여 정상성을 확인 후 두 가지의 방식으로 분석을 진행해보고자 한다.

1. ARIMA 모형

(1) 단위근 검정

예측을 진행하기 전에 정상성 검정을 통해 시계열이 정상적인 시계열인지 검정했다. 이전에 기간별 분석과 같이 ADF 검정과 KPSS 검정을 통해서 이를 진행했다.

<표 6> 단위근 검정 결과

	검정통계량	p-value
ADF	0.75281	0.99
KPSS	2.6949	0.01

위 결과에 따르면 두 시계열은 비정상 시계열인 것을 확인했으며, 정상성 변환이 필요하여 차분을

1) ARIMA 모형을 적용한 시계열 분석은 본질적으로 동태적 분석방식에 속한다는 특징을 지니고 있어, 적어도 50개 이상의 관측치가 존재하여야 만족할 만한 분석이 가능해진다고 명시하여 해당 연구의 학습 데이터가 분석을 위한 충분한 관측치임을 확인하였다.(김동환, 2014)

진행했다. 이전과 같이 R 패키지에 있는 `ndiffs` 함수를 통해서 시계열은 2차 차분이 제일 적당한 차분 횟수로 나왔다. 이후 차분된 시계열을 ADF와 KPSS 검정을 통해 정상적인 시계열인 것을 확인하고 모델링을 진행할 수 있었다.

(2) 시계열 분석 모델링

최적 모델을 선정하기 위해 두 가지 방법을 사용했다. 우선, 첫 번째 방법은 `pmax`와 `qmax`를 직접 설정하여 제일 작은 AIC값을 가지는 모델을 최종 모델로 선정할 수 있다. `p`값과 `q`값을 직접 설정하고 바뀌가면서 AIC값이 제일 작게 나오는 모델을 최종 모델로 선정하였다. 모델링 결과, AIC값 413.2를 가지고 있는 ARIMA(1,0,5)를 추정했고, 이를 `LjungBox-Test`를 통해 잔여 값 간 상관관계가 없는 것을 확인 후 ARIMA(1,0,5)를 최종 모델로 선정했다. 두 번째 방법은 기간별 분석에서 사용했던 방법처럼 `auto.arima` 함수를 이용해서 자동적으로 최적 모델을 추정하고 선정했다. 코드를 돌린 결과, AIC값 414.21, AICc값 414.79, BIC값 432.33를 가지고 있는 ARIMA(0,0,4)(0,0,1)[12] 모델을 추정했다. 마찬가지로, `ggetdiag` 함수를 통해서 모델이 잔차의 계열 상관성이 없다는 것을 확인하였고 해당 모델을 최종 모델로 선정하였다.

(3) 예측 결과

`p`값과 `q`값을 직접 설정하여 선정한 모델을 가지고 예측을 진행하는 것을 더불어 `multi-step forecast`와 `rolling forecast` 두 가지 방법을 사용하여 전범위 기간 아파트 매매실거래가격지수 예측을 진행했다. 전범위 기간에 대한 `multi-step forecast`, `rolling forecast`, 그리고 `p`값과 `q`값을 직접 설정하여 선정한 모델을 가지고 진행했을 때 예측 정확도를 비교한 결과는 다음과 같다.

<표 7> 모델 예측 정확도 결과 비교

	Multi-Step	Rolling	(p,q) Model
RMSE	1.27316	0.6627419	1.61728

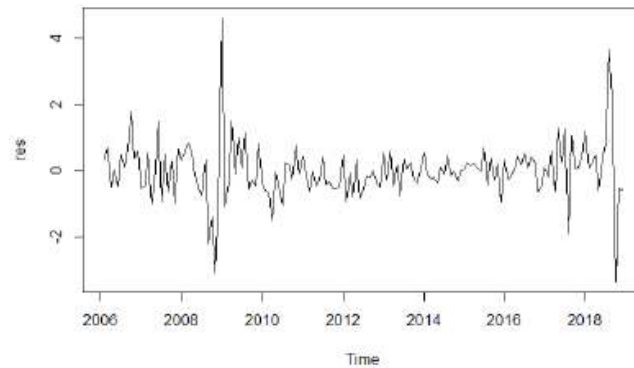
모델의 예측 정확도를 봤을 때 기간별 모델 예측 정확도 결과와 마찬가지로 `rolling forecast` RMSE 값이 제일 작다. 그다음 `multi-step forecast` RMSE 값이 다음으로 작고, `p`값과 `q`값을 직접 설정하여 선정한 모델 RMSE 값이 제일 크다. 따라서, `rolling forecast` 방법으로 모델을 예측했을 때 정확도가 제일 높다고 할 수 있다.

2. ARIMA + 변동성 모형

앞서 2차 차분을 통하여 ARIMA모형에 적합하기 전 `auto.arima` 함수에 1차 차분한 학습 데이터를 적합한 후 잔차를 확인해 보았다. 아래 <그림 5>는 학습 데이터를 1차 차분 후 ARIMA(0,0,3)(1,0,0)[12]에 적합한 후의 잔차이다.

<그림 5>를 통해 두 기간에서 잔차의 변동성이 급증함을 확인할 수 있었다. 첫 번째 변동폭이 커진 기간은 2008년 4분기부터 2009년 1분기에 해당하며 2008년 미국에서 발생한 금융위기에 의한 영향으로 추정된다. 두 번째 기간은 2018년 4분기에서 2019년 1분기에 해당하며 이는 9.13정책의 영향으로 인한 변동폭 상승으로 추정되어진다. 2008년 금융위기로 인해 수도권 주택 매매가격 하락에 대한 금리 상승 기조와 가격하락의 우려로 주택 매수 심리가 위축되어 실거래 가격이 하향조정 되었다. 이와 유사하게 부동산 시장에서 투기를 목적을 한 수요를 차단하고 실수요자를 보호하기 위한 9.13정책이 시행된 이후로 아파트 매매수요가 감소함에 따라 아파트 매매 실거래 가격이 하락하였다. 즉, 잔차의 변동성이 커진 두 기간 모두 외부의 충격 혹은 국내 부동산 정책으로 인한 급격한 실거래 매매가격 하락에서 비롯된

것임을 알 수 있다.



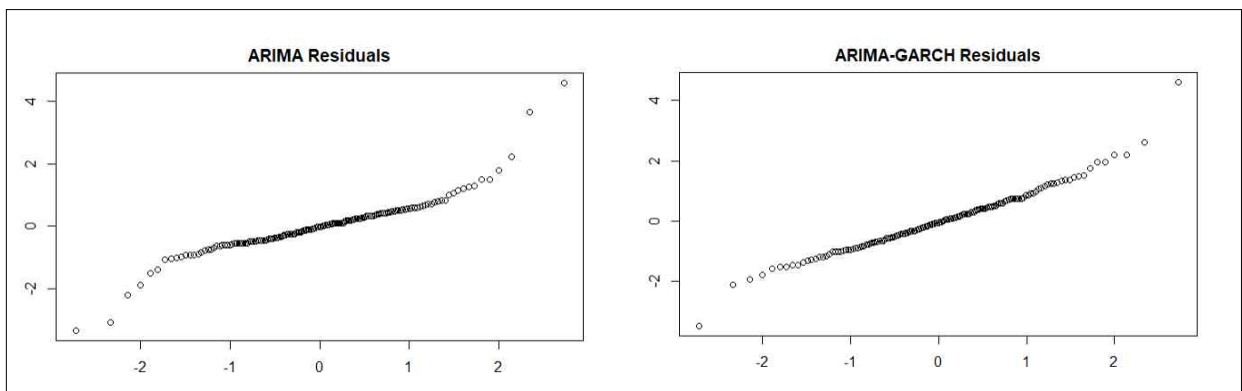
<그림 5> 모델 잔차 플랏

이에 해당 잔차를 변동성 모형에 적합해 보아, 시험 데이터 기간의 예측 신뢰구간을 보다 정확하게 추정해보고자 하였다. 우선 주어진 잔차에 가장 적합한 모델을 파악하기 위해 ugarchspec함수를 활용하였다. ugarchspec함수는 garchOrder로 형성되는 variance.model과 armaOrder로 형성되는mean.model로 구성되어 있다. 주어진 잔차에 적합한 모형을 탐색할 것이므로 armaOrder(p,q)에는 (0,0)을 garchOrder(p,q)에는 각각 0~3, 1~3까지의 범위를 지정하여 AIC값이 최소가 되는 p,q조합을 찾고자 하였다. 총 12가지의 경우의 모형을 적합한 후 각 모델의 AIC값은 다음과 같다.

<표 7> 모델 예측 정확도 결과 비교

	1	2	3
q \ p			
0	2.6023	2.6158	2.6287
1	2.1939	2.2092	2.2229
2	2.2040	2.2150	2.2263
3	2.2188	2.2298	2.2388

상단의 표를 통해 AIC값이 가장 최소가 되는 GARCH(1,1)를 최적모형으로 결정하였으며, 모델의 적합성을 진단하기 위해 GARCH모형의 잔차가 정규성을 따른다는 가정 하에 변동성 모형 적합 전과 후의 잔차의 qqplot을 비교해보았다. 이때 GARCH모형의 적합 후 잔차(e_t)는 $\epsilon_t / \sqrt{h_t}$ 를 통해 도출하였다.



<그림 6> 잔차들의 정규성 비교

2) 여기서 h_t 란 시간에 따른 조건부 분산(conditional variance)을 의미한다.

<그림 6>을 통해 변동성 모형을 적합하기 이전보다 적합한 이후의 잔차들이 비교적 정규성을 충족하였다는 사실을 파악할 수 있으므로, GARCH(1,1)이 주어진 잔차에 적합한 모델임을 확인할 수 있다. 적합한 모델에 대해 ugarchfit과 ugarchforecast 함수를 활용하여 시험 데이터에 해당하는 36개월의 분산을 추정하여 신뢰구간을 형성한 후 GARCH모형을 적용하기 전과 비교해보고자 하였다.

<표 8> 모델 예측 정확도 결과 비교

Model	RMSE	95% Confidence Interval (2020.5)		
		Forecast	Lower	Upper
ARIMA(0,2,4)	3.683	133.4927	131.7268	135.2587
ARIMA(0,1,3)	4.131	133.7241	130.5575	135.8255
ARIMA(0,1,3+ GARCH(1,1))	3.727	133.1915	131.4239	135.1992

상단의 <표 8>은 3가지 모델에 대한 전반적인 RMSE와 시험데이터 구간에서 타 시점 대비 아파트 실거래 매매가격지수가 다소 감소하는 2020년 5월³⁾을 대상으로 해당 시점의 점 추정과 신뢰구간을 추정해보았다. 전반적인 성능에 있어서는 2차 차분 후 ARIMA모형을 사용한 첫 번째 모형이 우수함을 알 수 있다. 하지만, 타 시점 대비 변동폭이 큰 지점(2020년 5월의 경우)에서의 경우 신뢰구간의 길이를 비교해보았을 때 세 모델 중 가장 짧았다는 점에서 시간에 따른 분산의 변동성을 모형이 반영해 주고 있음을 알 수 있다.

IV. 다변량 분석

현재 다양한 경제정책 및 부동산 정책이 시행되고 있는 바, 아파트 실거래 가격 지수에 영향을 주는 주요 거시 경제 변수들을 조사하고 해당 변수들과 아파트 실거래 매매가격지수가 어떠한 영향을 주고받는지를 파악해보고자 VAR모형을 통한 다변량 시계열 분석을 진행해보고자 한다.

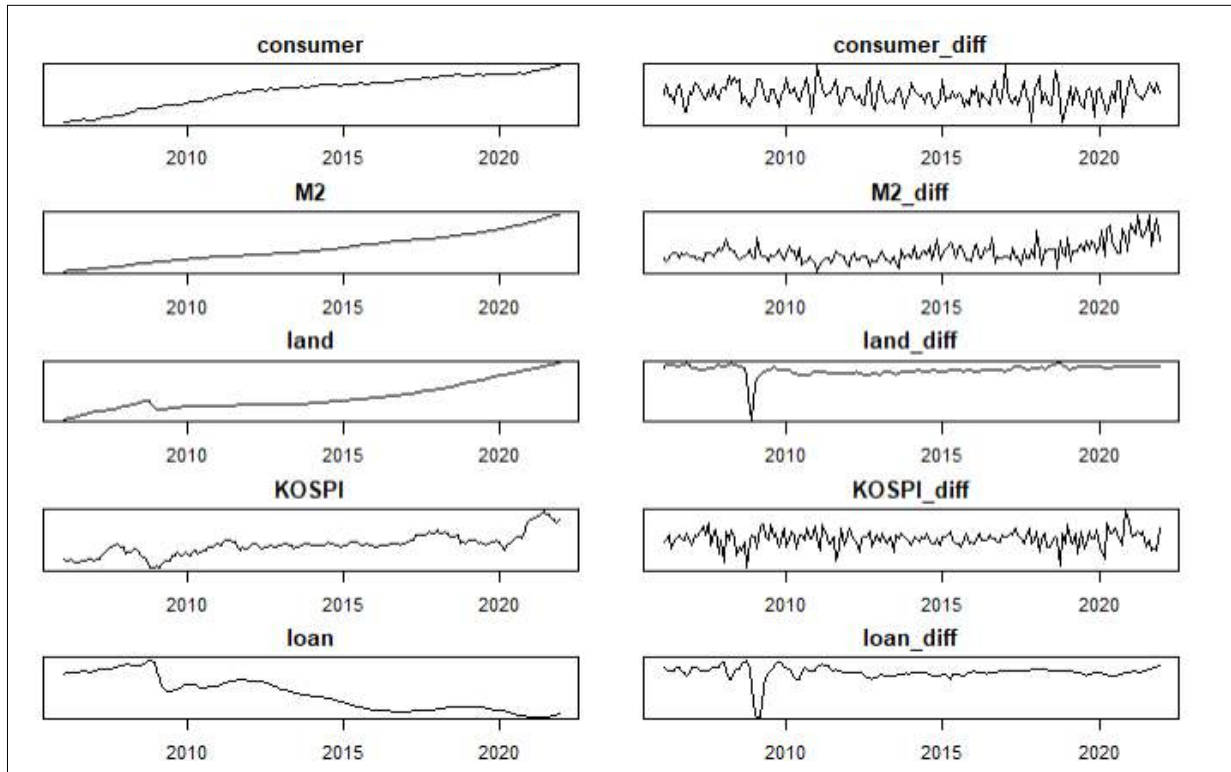
1. 변수 선정

우선, 아파트 실거래 매매가격 지수 형성에 영향을 줄 것이라고 판단한 소비자물가지수, 지가지수, 주가지수, 통화량(M2), 주택담보대출금리를 후보 변수로 선정하였다. 주택담보대출금리의 경우, 분석기간에 대응하는 데이터가 존재하지 않았기에 타 대출금리 중 가장 비슷한 추세를 보이는 가계대출금리를 주택담보대출 금리를 대리하는 지표로 활용하게 되었다.

<표 9> 후보 변수 단위근 검정 결과

	단위근 검정	차분 후 단위근 검정
consumer	0.6663	<0.01
M2	0.99	0.032
land	0.99	<0.01
KOSPI	0.069	<0.01
loan	0.3533	<0.01

3) 코로나 19로 인해 사업체들이 현금을 확보하기 위하여 자산을 매각하게 되었고 그 결과 부동산의 수요 대비 공급이 증가하여 아파트 매매 가격이 하락하는 기조를 보인 것으로 추정된다.



<그림 7> 설명변수 차분 전과 차분 후 그래프

<그림 7>의 왼쪽은 선정된 5개의 변수에 대한 2006년 1월부터 2019년 12월까지의 변화 양상이며 오른쪽은 각 5개의 변수에 대해 1차 차분을 진행한 것이다. 이후 최종적으로 선정된 변수와 아파트 매매 실거래 가격 지수를 VAR모형에 적합할 것이기에 주어진 변수들이 정상성을 만족하는지 확인해 보았다. 주어진 5개의 변수가 모두 추세가 존재한다는 가정 하에 type 3($y_t = c + \delta_t + \alpha y_{t-1} + u_t$)에 대한 adf.test를 시행해 보았고 모두 귀무가설을 기각함으로써 모든 변수에 대해 단위근이 존재함을 파악하였다. 이후 1차 차분 후 단위근 검정을 다시 한번 더 진행하였고, [표 c]를 통해 모든 변수에 대해 정상성이 충족됨을 확인 할 수 있다.

다음으로는 아파트 실거래 매매 가격 지수와 5개의 변수 간의 그래인저 인과관계의 유무를 파악해 보고자 한다. 그래인저 인과관계가 변수 간 일반적인 인과관계를 의미하는 것은 아니지만 그래인저 인과관계가 존재하지 않는다는 건 두 변수 간 어떠한 인과관계도 없다는 것을 의미하기에, 아파트 실거래 가격 매매 지수를 예측하는데 불필요한 변수가 무엇인지 일차적으로 구분할 수 있다고 여겼기 때문이다. 우선 VARselect()함수를 활용하여 그래인저 인과검정을 위한 최적의 lag를 파악하고자 하였다. 시차를 산정하는 여러 기준 중 전반적으로 작은 시차를 선택한 SC를 기준으로 삼았고 각 시차를 활용해 그래인저 인과검정을 진행한 결과는 다음과 같다.

<표 10> 그래인저 인과성 결과

	consumer	M2	land	KOSPI	loan
SC(n)	2	3	3	2	4
Granger	0.1668	0.064	0.001869	0.9331	0.001049

<표 10>를 통해서 파악할 수 있듯이, 지가지수(land)와 가계대출금리(loan) 두 변수가 아파트 매매 실거래 가격지수와 그래인저 인과성을 띠를 확인하여 두 변수를 통해 예측을 진행하기 위한 최적의 모형을 탐색하고자 한다.

모형 탐색 과정은 GAM(General Additive Model)과 왈드(Wald)검정을 통해 이루어졌다. GAM은 변수에 대해 선형 모형만이 아닌 평활함수를 사용하여 변수 간 비선형적 관계를 파악할 수 있기에 시계열 변수 간 적합한 모형을 판별함에 있어 OLS와 같은 선형모델보다 적합할 것이라고 판단하였다.

Model 1 : GAM(아파트 매매 실거래 가격 지수 ~ 지가지수 + 가계대출 금리)	(Full Model)
Model 2 : GAM(아파트 매매 실거래 가격 지수 ~ 지가지수)	(Reduced Model)
Model 3 : GAM(아파트 매매 실거래 가격 지수 ~ 가계대출 금리)	(Reduced Model)

GAM을 활용하여 위와 같이 3가지 모형을 형성한 후 왈드 검정을 통해 최적의 모형을 선택하였다. 왈드 검정이란, 각각 두 모델의 최대 가능도 추정량의 차이를 이용한 검정법으로 귀무가설을 '두 모델 간 최대 가능도 추정량의 차이가 같다'이다. 해당 검정법은 매개변수가 모두 존재하는 Full model과 특정 매개변수가 존재하지 않는 Reduced Model을 비교한다. 만약 두 모델 간 추정량의 차이가 0에 가깝다면, 두 모델의 설명력이 유사하다, 즉 Full model에만 존재하는 매개변수는 모델을 설명하는데 큰 역할을 하지 못한다는 의미 귀무가설을 기각하지 못하고 Reduced model을 택하게 된다. 해당 연구에서도 Model1과 Model2, Model3를 각각 왈드 검정을 진행하였고 결과는 다음과 같다.

<표 11> 왈드 검정 결과

	Model 1	Result
Model 2	8.424e-09	H_0 rejected Model 1 selected
Model 3	< 2.2e-16	

<표 11>에서와 같이 2번의 왈드 검정은 모두 귀무가설을 기각하였으며 지가지수(land)와 가계대출금리(loan) 변수들을 아파트 매매 실거래 가격 지수를 예측하는데 활용하고자 한다.

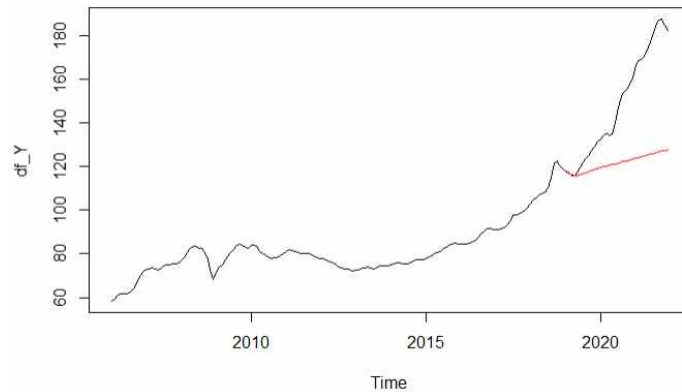
2. VAR 모형

앞서 최종적으로 선정된 두 변수(지가지수, 가계대출 금리)와 아파트 실거래 매매 가격 지수를 VAR모형에 적합한 후 다시 한번 아파트 실거래 매매 가격 지수와 두 변수 간 그래인저 인과성을 살펴보았다.

<표 12> Y와 두 변수 간 그래인저 인과성 검정

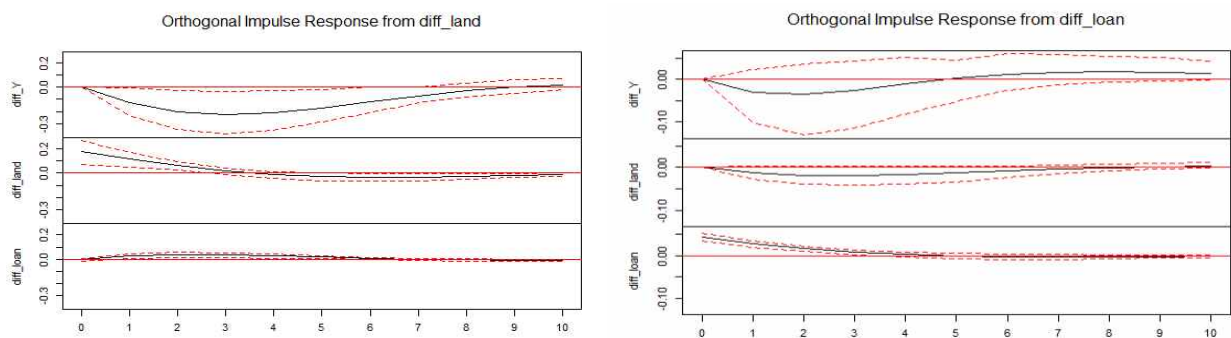
인과관계	lag	p-value
Y → land	3	0.001375
land → Y		0.05144
Y → loan	4	0.0009551
loan → Y		0.01251

<표 12>를 통해서 확인할 수 있듯이, 아파트 실거래 매매 가격 지수는 지가지수와 가계 대출 금리에 모두 강한 그래인저 인과성을 지니고 있다. 하지만 두 변수는 그에 대비 아파트 실거래 매매 가격 지수에 비교적 약한 그래인저 인과성을 지니는 것을 확인할 수 있었다. 적합한 VAR모형을 바탕으로 앞선 ARIMA모형과 동일하게 아파트 실거래 매매 가격 지수에 대한 36개월에 대한 예측을 진행하였고 결과는 하단의 그림과 같다.



<그림 8> VAR모형의 Forecast 결과

상단의 그림을 통해서 확인할 수 있듯이, 앞선 ARIMA를 활용하여 예측한 것에 비해 성능이 떨어짐을 확인할 수 있다. 이는 그래인저 인과검정, 왈드 검정 과정을 거쳐 선정된 두 변수(지가지수, 가계대출금리)가 아파트 실거래 매매 가격지수의 상승 추세를 충분히 설명하지 못한다는 것을 의미한다. 이는 앞선 <표 12>에서 아파트 실거래 매매 가격지수에 대한 두 변수의 인과성이 충분치 못하다는 점에서도 추정할 수 있다. 본 연구는 추가적으로 적합한 VAR모형에 대해 충격반응함수를 그려 세 변수 간 관계를 더 면밀히 살펴보고자 했다.



<그림 9> VAR모형의 충격반응함수 그래프

<그림 9>의 왼쪽은 지가지수에 대한 충격반응함수를 시각화 한 것이며 오른쪽은 가계 대출 금리에 대한 세 변수의 충격반응변수를 시각화한 것이다. 이를 통해 지가지수에 충격이 발생할 시 아파트 매매 실거래 가격 지수에 대해 negative한 방향으로 영향을 미침을 파악할 수 있었다. 이에 반해, 가계 대출금리의 경우 모든 시차에서 신뢰구간이 0의 값을 포함하고 있어 해당 변수가 아파트 매매 실거래 가격 지수에 영향이 있다고 판단하기 어렵다. 더불어 두 변수는 충격이 발생할 시 시차가 3일 때까지 positive한 방향으로 영향을 미침을 파악할 수 있었다.

V. 결론 및 한계

기간에 따라 아파트 매매 실거래 가격지수를 살펴보았을 때, 2013년 이전은 2008년 미국 금융위기 영향으로 인한 가파른 서울시 아파트 매매가격 하락으로 인해, 2013년 이후로는 지속적인 아파트 가격 상승으로 인해 정상성을 충족하지 않았다. 각 기간에 대해 차분을 진행한 후 ARIMA모형을 통해 모형을 적합하였으며, 모두 rolling window를 통한 예측 성능이 모두 높게 나왔음을 확인하였다. 아파트 매매 가격 지수는 국내 부동산 정책 및 금융동향 과 같은 외부변수들에 의한 변동성이 큰 지표이기에 이전 시점의 추정값도 예측에 포함하였을 때 비교적 높은 성능을 보였다고 판단하였다.

두 번째로 전 기간에 대해 아파트 매매 실거래 가격지수를 예측해보고자 2차 차분 후 ARIMA 모형, 1차 차분 후 ARIMA+GARCH모형을 적합해 보았다. ARIMA모형을 활용하였을 때, 전반적으로 높은 예측 성능을 보였으나 아파트 매매 실거래 가격 지수의 변동성이 큰 지점에선 ARIMA와 GARCH를 결합한 모형이 추정구간을 보다 정밀하게 예측하였음을 확인하였다. 더불어, 2006년부터 2021년까지 서울시내 아파트 매매 실거래 가격의 주요 변동은 금융위기와 부동산 투기 제한 정책에 의한 아파트 수요가 감소됨에 따른 아파트 실거래 매매 가격의 하락으로 인해 발생하였음을 깨달을 수 있었다.

마지막으로 서울시내 아파트 매매 실거래 가격 지수에 영향을 줄 것이라고 판단한 후보 변수들로 그래inger 인과검정, 월드 검정을 통해 최적의 변수 조합을 선정하였다. 이를 VAR모델이 적합하여 이후 36개월의 아파트 매매 실거래 가격 지수를 예측해보았다. 3개의 변수에 대한 그래inger 인과성과 예측 성능을 바탕으로 최종적으로 설명변수로써 선택된 지가지수와 가계대출금리가 아파트 매매 실거래 가격 지수의 추세를 설명하는데 큰 역할 하지 못함을 파악하였다.

해당 연구는 다양한 시각에서 서울시 아파트 실거래 매매 가격 지수를 하였고 ARIMA모형과 rolling window기법을 통해 높은 성능을 보이는 모델을 형성할 수 있었다. 더불어 서울시 내 아파트 매매 실거래가 가격 지수가 큰 폭으로 변동하게 되는 주요 원인을 파악하여 이후 아파트 매매 수요를 감소하게 할 외부 요인 발생 시 3~4개월 이후 그 충격이 아파트 매매 실거래가 지수에 반영됨을 확인할 수 있었다. 하지만, 다변량 시계열 예측을 진행하는 과정에서 후보 변수를 임의적으로 선정하였다는 점에서 한계를 지닌다. 만약 보다 선행연구를 통해 면밀히 진행하였다면, 현 연구결과보다 아파트 매매 실거래 가격 지수에 영향을 주는 변수들과의 관계를 보다 정확하게 파악할 수 있을 것이다.

참고문헌

- 김동환, “VAR 모형을 이용한 토지시장 가격예측”, 제 33권, 대한부동산학회지, 2015
- 조보근, 박경배, 하성호, “기계학습 알고리즘을 활용한 지역 별 아파트 실거래가격지수 예측모델 비교 : LIME 해석력 검증”, 제 29권 제 3호, 한국정보시스템학회, 2020
- 김근용, “주택가격 예측을 위한 모형설정과 검정”, 통권 197호, 54-61, 국토, 1988
- 고봉현, “GARCH 모형을 이용한 수산물 가격변동성에 관한 연구”, 제22권, 해양정책연구, 2007.
- 박유성, 허명희, “시계열 자료분석”, 제3강, 제4강, 자유아카데미, 1998.
- 문권순, “벡터자기회귀(VAR)모형의 이해”, 제2권, 통계분석연구, 1997.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1-3), 159-178.
- R Documentation. (2022). “Number of differences required for a stationary series”, accessed Dec 10, 2022, <https://search.r-project.org/CRAN/refmans/forecast/html/ndiffs.html>.
- R Documentation. (2022). “Fit best ARIMA model to univariate time series”, accessed Dec 10, 2022, <https://search.r-project.org/CRAN/refmans/forecast/html/auto.arima.html>.