



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

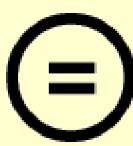
다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원 저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리와 책임은 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)



석사학위논문

서울시 아파트 매매가격지수 예측력 비교 연구

- 시계열과 딥러닝 모형 중심으로

Prediction the Trading Price Index of Apartment
Focused on the Seoul City

- Using Time Series Analysis and Deep Learning Model

조민지

한양대학교 부동산융합대학원

2023년 2월

석사학위논문

서울시 아파트 매매가격지수 예측력 비교 연구

- 시계열과 딥러닝 모형 중심으로

Prediction the Trading Price Index of Apartment Focused on the Seoul City

- Using Time Series Analysis and Deep Learning Model

지도교수 성 현 곤

1939

이 논문을 공학 석사학위논문으로 제출합니다.

2023년 2월

한양대학교 부동산융합대학원

도시·부동산개발 전공

조민지

이 논문을 조민지의 석사학위 논문으로 인준함

2023년 2월

심사위원장 : 고준호



심사위원 : 우아영



심사위원 : 성현곤



한양대학교 부동산융합대학원

목 차

제1장 서 론	1
제1절 연구의 배경 및 목적	1
제2절 연구의 범위	2
제3절 연구의 방법과 흐름	3
제2장 선행연구 검토	4
제1절 선행연구 검토	4
제2절 선행연구의 한계 및 본 연구의 차별성	8
제3장 분석의 틀	9
제1절 분석 자료	9
제2절 분석 모형	15
제3절 분석 방법 및 절차	16
제4장 실증 분석	18
제1절 주택매매가격지수(KB)의 예측력 비교	18
제2절 매매실거래가격지수의 예측력 비교	29
제5장 결론 및 시사점	40
제1절 요약 및 결론	40
제2절 시사점 및 향후 연구 방향	42
참 고 문 헌	43
부 록	45
ABSTRACT	58

표 목차

[표 1] 부동산 시장 예측력 비교 선행연구	4
[표 2] 부동산 시장 가격 결정요인 선행연구	7
[표 3] 자료구성	9
[표 4] 기초통계량	11
[표 5] 단위근 검정 - 주택매매가격지수(KB)	18
[표 6] 그랜저 인과관계 검정 결과 - 주택매매가격지수(KB)	21
[표 7] 공적분 검정 결과 - 주택매매가격지수(KB)	23
[표 8] LSTM 모형 결과 예시 - 주택매매가격지수(KB)	25
[표 9] GRU 모형 결과 예시 - 주택매매가격지수(KB)	25
[표 10] LSTM 최적 모형 결과 - 주택매매가격지수(KB)	26
[표 11] GRU 최적 모형 결과 - 주택매매가격지수(KB)	27
[표 12] 주택매매가격지수(KB) 예측력 비교	28
[표 13] 단위근 검정 - 매매실거래가격지수	29
[표 14] 그랜저 인과관계 검정 결과 - 매매실거래가격지수	32
[표 15] 공적분 검정 결과 - 매매실거래가격지수	34
[표 16] LSTM 모형 결과 예시 - 매매실거래가격지수	36
[표 17] GRU 모형 결과 예시 - 매매실거래가격지수	36
[표 18] LSTM 최적 모형 결과 - 매매실거래가격지수	37
[표 19] GRU 최적 모형 결과 - 매매실거래가격지수	38
[표 20] 매매실거래가격지수 예측력 비교	39

그림 목차

[그림 1] 변동추이	12
[그림 2] 분석방법 및 절차	17
[그림 3] VECM 예측 결과 - 주택매매가격지수(KB)	24
[그림 4] LSTM 변수 중요도 결과 - 주택매매가격지수(KB)	26
[그림 5] GRU 변수 중요도 결과 - 주택매매가격지수(KB)	27

[그림 6] LSTM 및 GRU 결과 - 주택매매가격지수(KB).....	28
[그림 7] VECM 예측 결과 - 매매실거래가격지수.....	35
[그림 8] LSTM 변수 중요도 결과 - 매매실거래가격지수.....	37
[그림 9] GRU 변수 중요도 결과 - 매매실거래가격지수.....	38
[그림 10] LSTM 및 GRU 결과 - 매매실거래가격지수.....	39



국 문 요 지

본 연구는 서울시 아파트에 대하여 KB국민은행의 주택매매가격지수(KB)와 한국부동산원의 공동주택실거래가격지수를 30개의 거시경제변수들과 다변량 데이터 분석하고자 하며, 시계열 분석 모형과 딥러닝(Deep Learning) 분석 모형을 설계하여 모형의 예측력을 비교 분석하고자 하였다. 분석에 사용된 주택 매매가격지수(KB)는 거래여부와 상관없이 모든 주택유형의 가격변동률을 확인할 수 있고, 매매실거래가격지수는 실제 거래된 주택의 가격변동률을 확인할 수 있다. 매매실거래가격지수가 주택매매가격지수(KB)보다 변동폭이 큰 경향이 있어, 전반적인 시장 파악을 위해서는 2가지의 지수를 병행해서 확인할 필요가 있기 때문에 함께 예측력을 비교 분석하고자 하였다.

시계열 분석 모형을 설계하기 위하여 30개의 거시경제변수와 주택매매가격지수(KB)와 매매실거래가격지수간의 그랜저 인과관계 분석을 통해 변수간 인과관계를 확인하였으며, 선행관계가 있는 변수를 선정하여 공적분 검정을 통해 VECM(Vector Error Correction Model) 모형으로 예측 모델을 설계하였다. 딥러닝 분석 모형을 설계하기 위하여 30개의 거시경제변수를 모두 입력변수로 사용하였고, LSTM(Long Short Term Memory)모형과 GRU(Gated Recurrent Unit) 모형으로 예측 모델을 설계하였다. 최적의 딥러닝 모형을 설계하기 위하여 하이퍼파라미터(hyperparameter)를 반복적으로 변경하면서 모형의 성능을 측정하였다. 시계열 분석 모형과 딥러닝 분석 모형의 예측력에 대한 성능 평가는 RMSE(Root Mean Square of Errors)평가지표로 비교하였다.

2006년 1월부터 2021년 11월까지 191개월을 학습하고 2021년 12월부터 2022년 5월까지 6개월을 예측하는 시계열 분석 모형 및 딥러닝 분석 모형을 설계하였고, 주택매매가격지수(KB)와 매매실거래가격지수 모두 딥러닝 분석 모형이 시계열 분석 모형 보다 예측력이 높은 것으로 나타났다. 다만, 시계열 분석 모형의 경우 단위근 검정, 그랜저 인과관계 검정, 공적분 검정 등 통계적으로 분석할 수 있다는 부분이 있기 때문에 시계열 분석 모형과 딥러닝 분석 모형을 함께 사용하는 것이 객관적으로 관계성을 확인할 수 있는 것으로 보인다.

제1장 서 론

제1절 연구의 배경 및 목적

10년간의 서울시 부동산 거래 현황 데이터를 살펴보면 주택 전체 매매 거래 중 아파트 매매는 평균적으로 50%의 비율로 거래되고 있었으나, 2022년 초부터 주택 전체 매매 거래 중 아파트는 평균 26%의 비율로 거래되고 있다. 또한 서울시 주택매매가격변동률 데이터를 살펴보면 2022년 7월 - 0.09%, 8월 - 0.24%, 9월 - 0.47%, 10월 - 0.81%로 최근 10년간 중 가장 가격 하락폭이 컸다. 이러한 원인을 살펴보면 한국은행의 기준금리가 2020년 5월 0.5% 최저를 기록하고 2022년 4월부터 6차례 연속으로 인상하면서 1.25%에서 2022년 11월 기준 기준금리는 3.25%가 되면서 짧은 기간에 상승폭이 컸다. 기준금리의 상승은 가계대출, 주택담보대출, 신용대출 등 모든 대출금리가 상승하게 되고 기업과 개인에 모두 영향을 주게 된다. 한국의 경우 가계대출 중 29%가 고정금리이며, 대다수 변동금리에 노출되어 있어 금리 인상에 취약하다.

이러한 기준 금리는 물가동향, 국내외 경제 상황, 금융시장 여건 등 다양한 거시경제 변화에 따라서 상승 또는 하락하고 있다. 이처럼 부동산 시장도 다양한 거시경제 변화에 의해서 영향 받기 때문에 부동산 가격 결정 요인에 대한 선행연구가 진행되어 오고 있으며, 또한 최근에는 부동산 시장과 거시경제 변수 데이터를 분석하여 향후 부동산 시장을 예측하여 정책적인 판단에 근거 자료로 활용하고자 하는 연구도 활발히 진행되고 있다.

본 연구는 지역의 범위를 서울시에 한정하고, 아파트 매매가격지수와 30개 거시경제변수의 관계성을 그랜저 인과관계 검정을 통해 인과관계를 확인하고, 다변량 데이터 분석을 하고자 한다. 시계열 모형과 딥러닝(Deep learning) 모형으로 향후 부동산 시장을 예측하고 어떠한 모형이 예측력이 좋은지 비교 연구하고자 한다. 부동산 가격 예측은 부동산 시장에 참여하고 있는 모든 이해관계자에게 유용한 정보를 제공할 수 있기 때문에 부동산 시장 안정화에 활용 할 수 있다.

제2절 연구의 범위

한국부동산원에서 제공하는 공동주택 매매 실거래 가격지수는 국가승인통계 자료로 2006년 1월 1일부터 시행된 부동산거래 신고제도에 의하여 「부동산거래신고에 관한 법률」 제3조에 의거 매매계약을 체결한 뒤 관할 시군구청장에게 신고한 실제 아파트 및 연립·다세대 거래가격 자료로 통계가 작성되며 작성주기는 매월 작성되어 공표된다.

전국의 재고 주택을 대상으로 실제 거래되어 신고된 주택 가격 정보로 통계를 생산하여 정확한 시장동향 정보를 국민에게 제공하고 정부 정책 수립에 참고자료로 활용되고 있다. 허위신고와 통계왜곡 문제를 해결하기 위해서 부동산 실거래신고기간은 기존 60일에서 2020년 2월 21일부터 30일로 단축되었다. 실거래신고 된 가격을 기준으로 입력오류, 특이거래 신고자료, 가격 이상치를 제외하고 지수가 산정된다.

국민은행에서 제공하는 주택매매가격지수(KB)는 전국 주택의 매매 및 전세 가격 변동 상황을 조사하여 주택시장 동향을 파악하고 분석하기 위해서 작성되었으며, 월 1회, 주 1회 자료가 조사되며, 보도자료와 KB부동산 홈페이지를 통해서 조사 익월 1일, 주간은 매주 금요일에 공표되고 있다. 표본주택이 거래가 된 경우에는 실거래가격을 거래가 되지 않은 경우에는 매매(임대)사례비교 법에 의하여 조사된 가격을 해당지역 부동산중개업소에 직접 온라인상 조사표에 입력하는 조사를 기본으로 하고, 온라인 조사가 불가능한 부동산중개업소에 한하여 조사원이 전화 또는 팩스로 조사한다. 매도자 또는 매수자의 호가는 조사되지 않는다.

국토교통부 홈페이지 정책 Q&A에서 아파트 실거래가격지수에 대해서 2012년 8월에 등록된 내용을 참고하면 2개의 지수 통계는 지수의 의미, 자료수집 방법, 통계 유형, 공표단위, 공표주기 및 시차 등에 차이가 있기 때문에 전반적인 시장상황 파악을 위해 2개의 지수를 병행 활용할 필요가 있다고 되어 있다. 따라서 본 연구에서는 서울시 아파트에 대하여 주택매매가격지수(KB)와 공동주택 매매 실거래 가격지수 2개의 지수를 종속변수로 선정하고 각 지수와 거시경제변수와의 관계성을 확인하고자 하며, 어떠한 매매가격지수가 거시경제변수와의 관계성이 높은지 확인하고, 예측 모형을 설계한다.

제3절 연구의 방법과 흐름

본 연구에서 연구의 방법은 거시경제변수를 사용하여 다변량 분석을 진행하기 때문에 시계열 모형의 경우 벡터자귀회귀모형(Vector Autoregressive Model, VAR)과 벡터오차수정모형(Vector Error Correction Model, VECM) 2 가지를 선정하였고, 딥러닝 모형의 경우 LSTM(Long Short Term Memory)과 GRU(Gated Recurrent Unit) 2 가지를 선정하였다. 시계열 분석 모형의 경우 매매가격지수와 거시경제변수의 공적분 검정 결과에 따라 공적분 관계가 없으면 VAR 모형, 공적분 관계가 있으면 VECM 모형을 사용하는 것으로 하였다. 예측하고자 하는 종속변수인 주택매매가격지수(KB)는 1986년 1월부터 통계 수집이 되어 오고 있으며, 매매실거래가격지수는 2006년 1월부터 통계 수집이 되어 오고 있다. 동일한 기간의 데이터를 기준으로 비교하기 위하여 2006년 1월부터 2022년 5월까지의 197개월의 데이터로 데이터 분석을 진행하였으며, 과거 데이터를 기반으로 6개월의 예측하는 모형을 설계 후 예측값과 실제값과의 차이를 평균제곱오차근(Root Mean Square of Errors, RMSE) 지표로 모형의 성능을 평가하였다.

본 연구는 제 5장으로 구성되어 있다. 제1장에서는 연구의 배경 및 목적, 연구의 범위, 연구의 방법과 흐름에 대해 제시한다. 제2장에서는 선행연구 검토를 진행하고, 시계열 분석 모형과 딥러닝 모형의 예측력 비교 연구에 대한 연구를 고찰하고, 거시경제변수와 부동산 시장에 미치는 관계성에 대한 연구를 고찰한다. 선행 연구들의 한계점을 확인하고 본 연구의 차이점을 도출한다. 제3장에서는 분석자료, 분석모형, 분석방법에 대해서 구체화 한다. 제4장에서는 주택매매가격지수(KB), 매매실거래가격지수에 대해서 시계열 분석 모형과 딥러닝 분석 모형을 설계하여 예측력을 비교한다. 제5장에서는 연구 결과를 요약하고 연구의 시사점과 향후 연구 방향에 대해서 고찰하며 연구를 마무리한다.

제2장 선 행 연구 검토

제1절 선 행 연구 검토

그동안 부동산 시장에서 가격 변화 추이와 가격 결정요인에 대한 다양한 연구가 진행되어 왔으며, 부동산 시장 예측을 위해서 전통적인 시계열 분석이 주로 진행 되어왔다. 최근에는 부동산 시장 예측에 다양한 인공지능(머신러닝, 딥러닝) 모형을 가지고 부동산 시장을 예측하는 연구가 진행되고 있으며, 전통적인 시계열 분석 모형과 인공지능 모형 예측력 차이를 비교하는 연구가 진행되고 있다. 다변량 분석 시에 주로 설명 변수로 주요 거시경제지표를 일부 선정하고 시계열과 인공지능 모형에 입력하여 예측하고 있다.

1. 부동산 시장 예측력 비교 선 행연구

<표1> 부동산 시장 예측력 비교 선 행연구

연구자	종속변수	독립변수	자료 범위
이주미·박성훈 조상호·김주형 (2021)	매매실거래가격지수 지가지수 전세실거래가격지수 부동산심리지수	-	2012.01 ~ 2019.12 서울
전해정 (2020)	주택매매가격지수(KB)	주택전세가격지수(KB) 회사채수익률 산업생산지수 건축허가면적	2001.01 ~ 2018.12 전국, 서울, 강남, 강북
윤만식·신성윤 (2020)	매매실거래가격지수	국고채금리 통화량 주택담보대출 소비자물가지수 산업생산지수 경제심리지수	2006.01 ~ 2019.11 서울 및 6대 광역시
배정완·유정석 (2018)	매매실거래가격지수	회사채수익률 소비자물가지수 통화량 광공업지수	2006.01 ~ 2017.08 서울

이주미·박성훈·조상호·김주형(2021) 단변량 분석을 하였고, 머신러닝 모형으로 RF(Random Forest), XGBoost(eXtreme Gradient Boosting), LSTM(Long Short-Term Memory)을 사용하였고, 모형의 예측력 비교를 위하여 평균 제곱근 오차(Root Mean Square Error, RMSE) 지표를 가지고 성능평가를 하였다. 4개의 각 부동산 지수의 특성과 데이터 형상에 따라 머신러닝 모형의 예측력의 차이가 있다고 하였다.

전해정(2020) 단변량 분석과 다변량 분석을 하였고, 시계열 분석 모형으로 자기회귀누적이동평균(Autoregressive integrated moving average, ARIMA), 벡터자기회귀 모형(Vector AutoRegressive Model, VAR), 베이지안 벡터자기회귀 모형 (Bayesian Vector AutoRegressive, BVAR) 모형을 사용하였고, 머신러닝 모형으로 LSTM, 순환 신경망(Recurrent Neural Network, RNN)을 사용하였고 모형의 예측력 비교를 위하여 평균제곱오차(Mean Squared Error, MSE)와 평균 제곱근 오차(Root Mean Square Error, RMSE) 지표를 가지고 성능 평가를 하였다. 머신러닝 모형이 시계열 분석 모형보다 예측력이 향상된다고 하였다.

윤만식·신성윤(2020) 단변량 분석과 다변량 분석을 하였고, 시계열 분석 모형으로는 ARIMA, 벡터오차수정 모형(Vector Error Correction Model, VECM) 모형을 사용하였고, 머신러닝 모형으로는 표준 LSTM, 양방향 LSTM, 게이트 순환 유닛(Gated Recurrent Unit, GRU) 모형을 사용하였고 모형의 예측력 비교를 위하여 RMSE 지표를 가지고 성능 평가를 하였다. 단변량 분석보다 다변량 분석이 예측력이 높으며, 부동산 시장 상황에 따라 변동성이 적은 경우 시계열 모형과 머신러닝 모형의 예측력 차이가 비슷하고, 변동성이 큰 경우 시계열 모형의 예측력이 낮아지는 부분이 있다고 하였다.

배성완·유정석(2018) 단변량 분석과 다변량 분석을 하였고, 시계열 분석 모형으로는 ARIMA, VAR, BVAR 모형을 사용하였고, 머신러닝 모형으로는 서포트 벡터 머신(support vector machine, SVM), RF, GBRT(Gradient boosted regression tree), 심층 신경망(Deep Neural Network, DNN), LSTM 모형을 사용하였고 모형의 예측력 비교를 위하여 MAE와 RMSE 지표를 가지고 성능 평가를 하였다. 시장 상황이 변동성이 적은 경우 시계열 모형과 머신러닝 모형 모두 의미 있는 예측력을 보여주나, 급변하는 시기에서는 시계열 모형은 예측이 어렵지만 머신러닝 모형은 의미 있는 예측이 가능하다고 하였다.

2. 부동산 시장 가격 결정요인 선행연구

최문기·이성화(2021) 거시경제변수와 서초구, 강남구, 송파구, 강동구 아파트 실거래가격지수의 연계성을 분석하였고, 단위근 검정, 공적분 검정, VECM 모형을 구축하고 상호간 인과관계를 확인하였다. 분석에 사용된 거시경제변수는 산업생산지수, 주택담보대출 이자율, 주가지수, 소비자물가지수 총 4가지에 대해서 분석이 진행되었다. 주가지수는 (+) 방향으로 영향이 있으며, 주택담보대출이자율은 (-) 방향으로 영향을 미치고, 산업생산지수와 소비자물가지수는 (-) 방향으로 영향을 미쳤다.

조한국·김영곤(2020) 거시경제변수와 서울시아파트가격지수, 서울시아파트 거래량의 연계성을 분석하였고, 단위근 검정, 공적분 검정, 그랜저 인과관계 분석, VECM 모형을 구축하고 충격반응분석, 분산분해분석을 실시하여 상호간 인과관계를 확인하였다. 설명변수로 건설수주액, 취업자수, 주택담보대출금리, 코스피지수 총 4가지의 거시경제 변수와 연관성을 확인하였다. 서울시아파트 가격지수에 서울시아파트거래량은 영향이 없으며, 영향을 주는 변수로는 건설수주액은 (+)의 관계, 취업자수와 주택담보대출금리는 (-)의 관계가 있었다.

조한국·서충원(2019) 서울시 3대 권역 오피스 임대료 및 공실률과 거시경제 변수와의 연계성을 분석하였고, 단위근 검정, 공적분 검정, VAR 모형, 그랜저 인과관계 분석, 충격반응 분석, 분산분해분석을 실시하여 상호간의 관계성을 확인하였다. 거시경제변수로는 건설투자증감률, 소비자물가지수, 경제성장률, 회사채수익률, 생산자물가지수 총 5가지의 변수를 사용하였다. 거시경제변수 중에서 건설투자증감률이 오피스시장에 영향력이 가장 높았다.

김경민(2018) 서울시아파트시장에 거시경제변수가 영향을 미치는지 실증 분석을 진행하였다. 서울시 아파트 매매/전세 가격지수(KB)를 사용하였고, 거시경제변수로는 고용률, 실업률, 소비자물가등락률, 통화량증감률, 종합주가지수, 국고채수익률, 정기예금금리, 주택담보대출금리, M2통화량을 가지고 연계성을 확인하였다. 단위금 검정, VAR 모형, 그랜저 인과관계 검정, 충격반응분석과 분산분해분석을 실시하였다. 서울시 아파트 가격지수에 영향을 주는 변수는 M2통화량과 종합주가지수가 매매 및 전세에 영향을 주는 것으로 나타났다.

박성균·이현석(2012) 주거용과 상업용 부동산에 영향을 주는 요인을 비교 분석하였다. 주거용 부동산은 서울시 아파트 매매/전세 가격지수(KB)를 사용

하였고, 상업용 부동산은 서울시 소재 대형 오피스빌딩 평균 월임대료로 선정하였다. 연구를 위해 단위근 검정, 공적분 검정, 그랜저 인과관계 검정을 통하여 연관성을 확인하였다. 주거용 부동산 시장에서 매매 가격에는 국내총생산, 소비자물가지수, 회사채수익률, CD금리(양도성예금증서), M2(광의통화)통화량, 환율, 경기선행지수가 영향이 있으며, 전세 가격에는 국내총생산, 소비자물가지수, 회사채수익률, CD금리, 통화안정증권 수익률, 국민주택채권 수익률, M2통화량, 환율, 취업자수, 고용률, 경기선행지수가 영향을 주었다. 상업용 부동산 시장에서는 취업자수, 실업률, 고용률, 건축허가면적(상업용), 국내건설수주액(사무실 및 점포), 종합주가지수가 관계성이 있었다.

<표2> 부동산 시장 가격 결정요인 선행연구

거시경제변수	논문1	논문2	논문3	논문4	논문5
	2021	2020	2019	2018	2012
국내총생산			●		●
산업생산지수	●				
생산자물가지수			●		
소비자물가지수	●		●	●	●
월평균소비지출					●
회사채수익률			●		●
CD금리수익률				●	●
통화안정증권수익률					●
국민주택채권수익률					●
주택담보대출금리	●	●		●	●
국고채수익률(3년)				●	●
정기예금금리				●	●
M2통화량				●	●
환율					●
취업자수		●			●
실업률				●	●
고용률				●	●
건축허가면적					●
국내건설수주액		●			●
건설투자증감률			●	●	●
종합주가지수	●	●			●
지가변동률					●
경기선행지수					●

제2절 선행연구의 한계 및 본 연구의 차별성

시계열 분석 모형과 딥러닝 분석 모형은 연구자의 분석 자료 설정에 따라 모형의 결과 값이 달라질 수 있기 때문에 특정 분석 모형이 우수하다고 할 수 없고, 추가 연구의 필요성이 있다(배성완·유정석, 2018). 또한 분석 모형 설계 시 공간적 범위를 세분화 하고 다양한 거시경제 변수를 활용해 예측력을 비교 연구하는 부분이 필요하다(전해정, 2020). 딥러닝 분석 모형의 경우는 모형 설계 시 하이퍼파라미터 설정 후 분석할 때마다 측정값이 바뀌는 부분이 단점이다(윤만식·신성윤, 2020).

선행연구에서 시계열 분석 모형과 딥러닝 분석 모형 설계 시 분석 자료 거시경제변수들을 다양하게 검토하지 못한 부분이 있었다. 본 연구에서는 거시경제변수를 다양하게 검토될 수 있도록 30개를 선정하고, 2006년부터 2022년 까지 최신의 데이터를 가지고 다변량 데이터 분석을 진행하고자 한다. 또한 부동산 가격지수를 나타내는 주택매매가격지수(KB)와 매매실거래가격지수 2 가지를 모두 분석하여 예측 모형 중 어느 모형이 서울시 아파트 부동산 가격지수 예측력이 높은지 확인하고자 하였다. 또한 딥러닝 분석 모형 시 측정값이 바뀌지 않도록 모형을 설계하도록 하였으며, 최신의 딥러닝 모형 LSTM과 GRU 모형을 사용하여 예측을 비교하고자 하였다.

제3장 분석의 틀

제1절 분석 자료

본 연구는 주택시장 동향을 보여주는 공식통계자료인 한국부동산원의 공동주택실거래가격지수와 민간통계자료인 KB국민은행의 주택매매가격지수(KB) 2가지의 통계자료에서 서울시 아파트 매매 가격지수에 대해서 시계열 분석 모형과 딥러닝 분석 모형으로 각각의 지수를 예측하는 최적의 모델을 찾고자 하였다.

한국은행 경제통계시스템(ECOS)와 통계청 국가통계포털(KOSIS)에서 매월 공시되는 서울시 아파트 실거래가격지수와 매매가격지수(KB) 2가지의 종속변수와 선행연구 검토를 통해 선정된 30개 거시경제변수를 독립변수로 선정하였고, 자료의 기간은 2006년 1월부터 2022년 5월까지의 197개월의 데이터를 취합하여 분석에 사용하였다. 자료의 상세 내용은 <표3>과 같다.

<표3> 자료구성

구분	지표명	변수명	단위	출처
종속변수	주택매매가격지수(KB)(서울)	KBAPI	2022.01=100	KB국민은행
	매매실거래가격지수(서울)	API	2022.01=100	한국부동산원
독립변수	행정구역별 아파트매매거래현황(서울)	APC	호수	한국부동산원
	아파트 매매가격 대비 전세가격비율(서울)	APR	%	KB국민은행
	지가변동률(서울)	FLP	%	한국부동산원
	연간소득 대비 주택가격 비율(서울)	PIR		한국주택금융 공사
	연간소득 대비 대출금액 비율(서울)	LIR		한국주택금융 공사
	소비자물가지수	CPI	2020=100	통계청
	생산자물가지수	PPI	2015=100	한국은행
	수출물가지수	EPI	2015=100	한국은행
	수입물가지수	IPI	2015=100	한국은행
	선행지수순환변동치	CLI	2015=100	통계청
	동행지수순환변동치	CCI	2015=100	통계청
	국내건설수주액	COA	백만원	통계청
	건축착공현황	CW	m ²	국토교통부

독립변수	건축허가현황	RBP	m^2	국토교통부
	경제활동인구	EAP	천명	통계청
	실업률	UR	%	통계청
	고용률	ER	%	통계청
	국내총생산	GDP	%	한국은행
	총저축률	SR	%	한국은행
	국내총투자율	GDIR	%	한국은행
	경상수지	CAB	백만달러	한국은행
	금융계정	FA	백만달러	한국은행
	환율(원/미국달러)	ECR	원	한국은행
	M2	M2	십억원	한국은행
	국고채(3년)	GB	연%	한국은행
	회사채(3년)	CB	연%	한국은행
	주택담보대출 금리	MIR	연리%	한국은행
	KOSPI	KOSPI	1980.01=100	한국거래소
	경제심리지수	ESI		한국은행
	기업경기설사지수	BSI		한국은행

분석 자료의 기초통계량은 <표4>와 같다.

서울시 아파트 주택매매가격지수(KB)의 최소값은 45.01이고, 최대값은 100.47이며, 매매실거래가격지수의 최소값은 34.89이고, 102.63으로 나타났다. 서울시 아파트매매거래현황은 평균 6,838호수 거래되었고, 최소 1,113호수 최대 25,382호수로 표준편차가 컸다. 서울시 아파트 매매가격 대비 전세가격비율은 최소 38%, 최대 75%로 나타났다. 지가변동률의 경우 최소 -3.49%, 최대 0.87%로 나타났다.

연간소득 대비 주택가격 비율(Price to Income Ratio, PIR)과 연간소득 대비 대출금액 비율(Loan to Income Ratio, LIR)은 한국주택금융공사가 금융기관으로 양수한 보금자리론 및 내집마련 디딤돌 대출 자료에 근거하여 산정하는데, PIR은 평균주택가격/평균연소득으로 계산하고, LIR은 평균대출금액/평균연소득으로 계산된다. 한국부동산원의 주택가격과 통계청의 소득 자료로 계산하지 않고 한국주택금융공사의 대출 자료에 한하여 계산된다는 점이 있어 서울시의 일부 가구를 의미한다. 서울시 PIR은 최소 4.28배, 최대 9.94배이며, 서울시 LIR은 최소 1.52배, 최대 4.8배로 나타났다. PIR 최대 9.94배가 의미하는 바는 해당 가구가 1년간의 수입으로 지출 없이 약 10년을 모아야 주택 1채를 구매할 수 있다는 의미하며, LIR 최대 4.8배는 해당 가구의 1년간 수입의 약

5배가 되는 금액을 대출 받았다는 의미한다.

국고채(3년)은 최소 연 0.83%에서 최대 5.96%로 나타났으며, 회사채(3년)은 최소 1.65%에서 최대 8.56%로 나타났다. 주택담보대출 금리는 최소 연 2.39%에서 최대 연 7.58%로 나타났다.

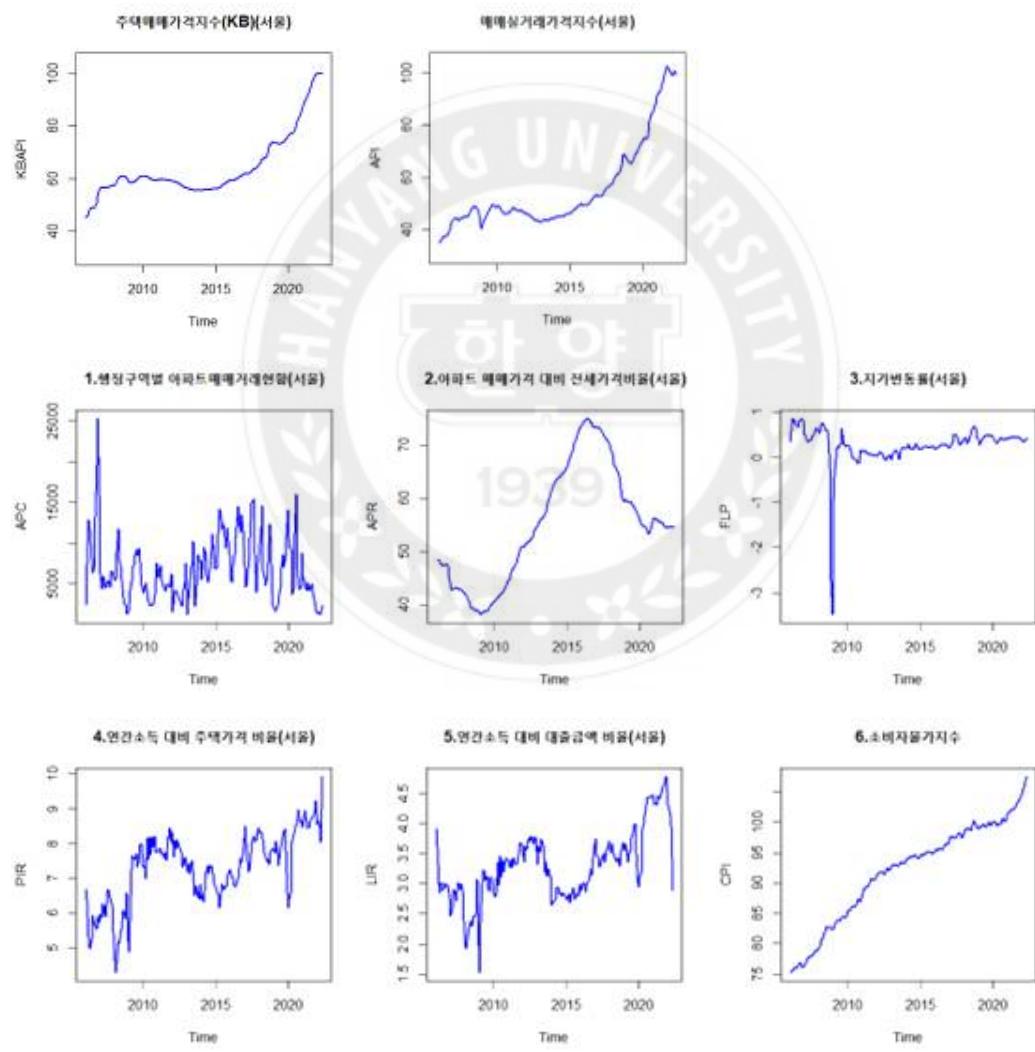
<표4> 기초통계량

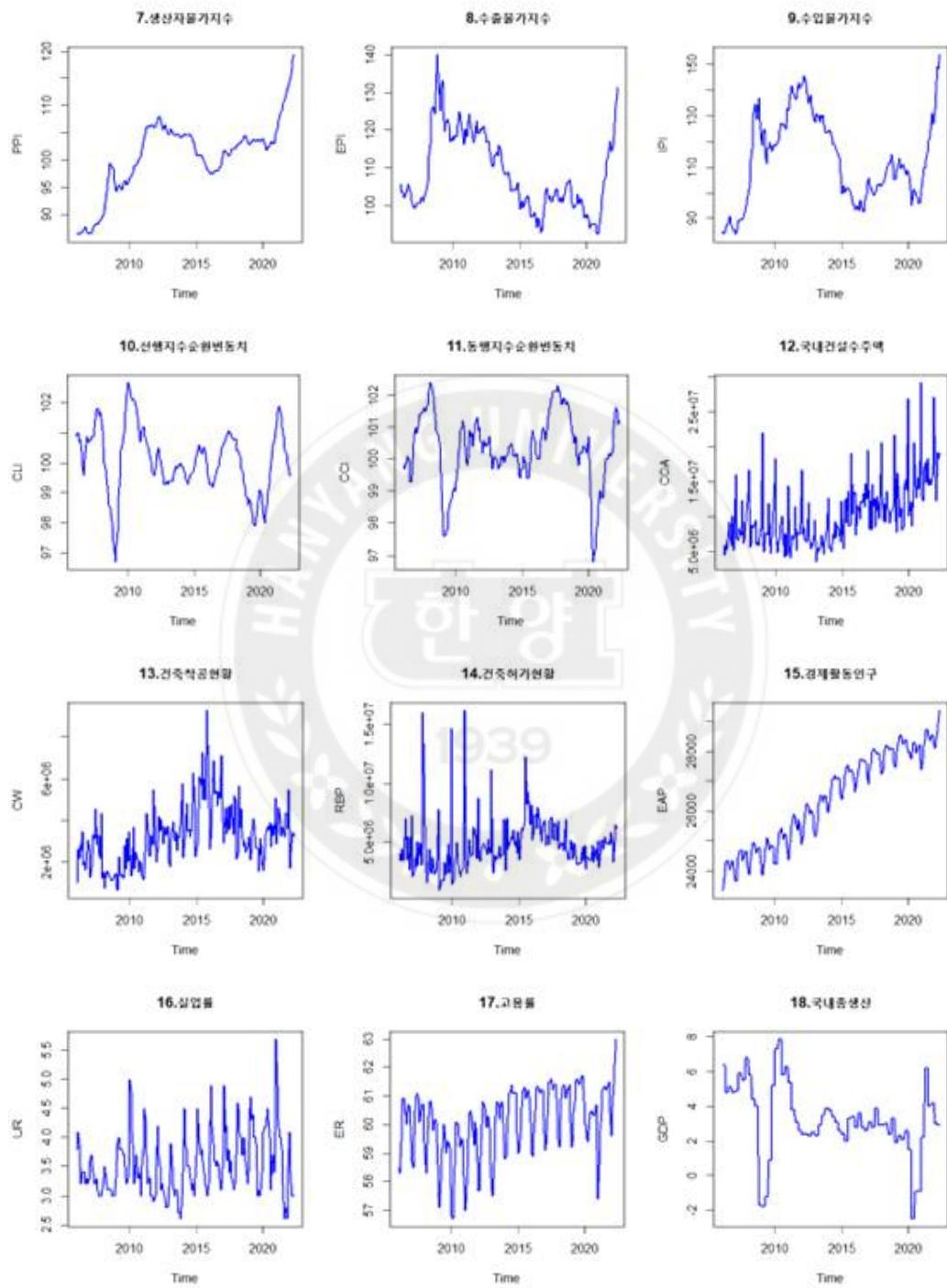
변수명	건수	평균	표준 편차	최소값	최대값
KBAPI	197	64.15	12.25	45.01	100.47
API	197	55.37	16.92	34.89	102.63
APC	197	6838.6	3876.51	1213	25382
APR	197	55.19	11.09	38.18	75.07
FLP	197	0.27	0.42	-3.49	0.87
PIR	197	7.29	1.03	4.28	9.94
LIR	197	3.3	0.59	1.52	4.8
CPI	197	91.88	8.16	75.23	107.56
PPI	197	100.76	6.72	86.42	119.43
EPI	197	108.57	10.34	92.3	140.1
IPI	197	113.23	16.84	83.84	154
CLI	197	100.09	1.11	96.7	102.7
CCI	197	100.33	1.1	96.8	102.4
COA	197	10,363,900	4,595,503	3,557,119	29,218,726
CW	197	3,250,617	1,446,308	650,093	9,279,701
RBP	197	4,799,083	2,253,240	904,537	16,276,236
EAP	197	26,390.89	1,562.52	23,347	29,374
UR	197	3.54	0.53	2.6	5.7
ER	197	60.09	1.16	56.7	63
GDP	197	3.23	2.1	-2.5	7.9
SR	197	35.09	1.28	32.8	37.7
GDIR	197	31.5	1.85	26.8	36.7
CAB	197	4,606.17	4,066.43	-7,646.1	18,750.9
FA	197	4,567.86	3,896.02	-4,023.8	12,409.7
ECR	197	1,116.15	97.03	915.86	1,461.98
M2	197	2,110,983	700,845.3	1,027,697	3,694,489
GB	197	2.88	1.39	0.83	5.96
CB	197	3.61	1.62	1.65	8.56
MIR	197	4.19	1.36	2.39	7.58
KOSPI	197	2,001.37	439.6	1,073.95	3,259.14
ESI	197	99.97	10.5	61.9	116.8
BSI	197	76.86	8.89	51	95

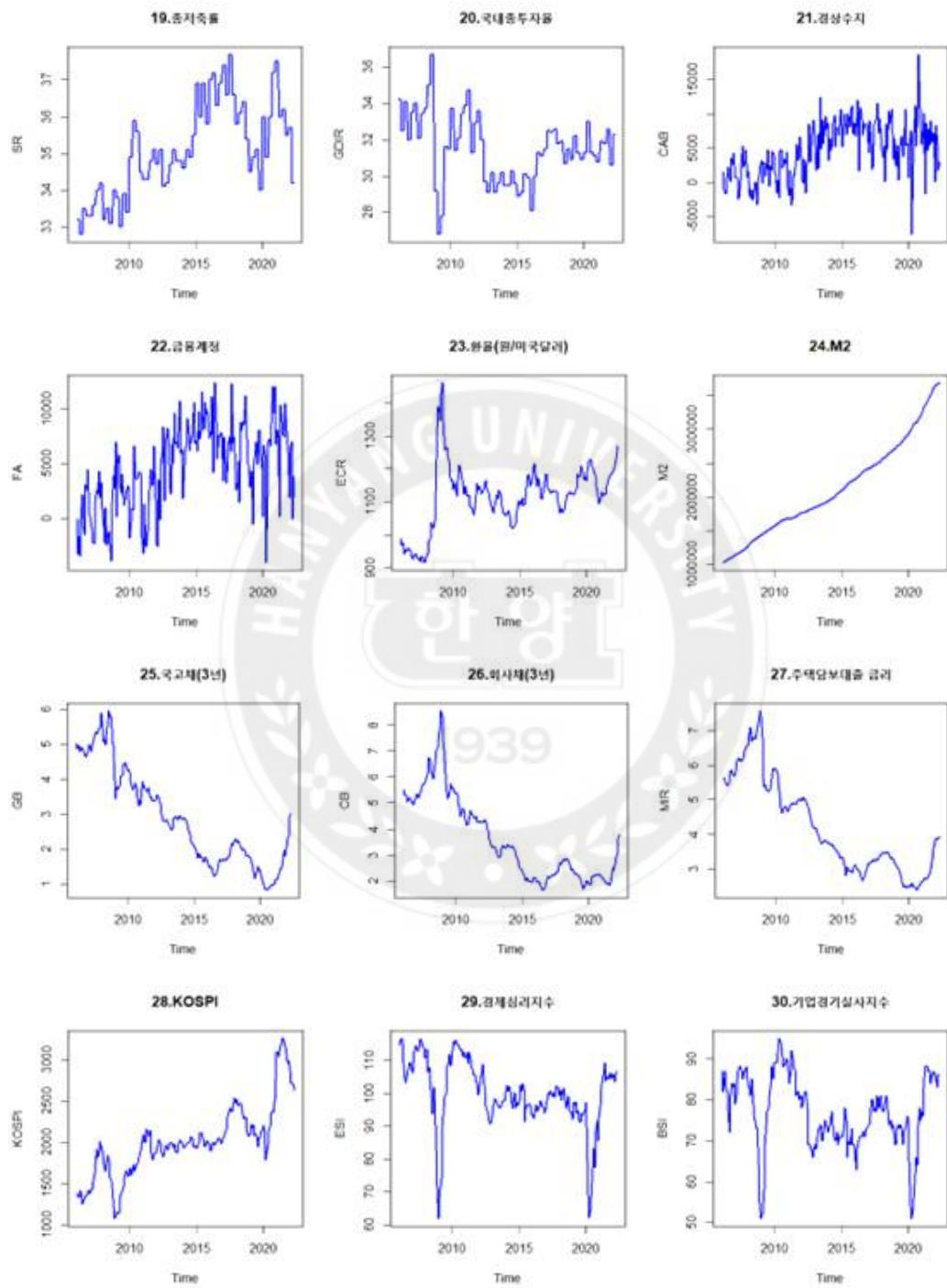
분석 자료의 변동추이는 <그림1>과 같다.

2008년 금융위기에 영향을 받아 매매실거래가격지수, 지가변동률, PIR 및 LIR, 선행지수순환변동치, 동행지수순환변동치, 경제심리지수, 기업경기실사지수등 많은 지표들이 하락하는 모습을 보였으며, 2020년 코로나19로 인해 경제심리지수, 기업경기실사지수 등 하락하는 모습도 확인할 수 있었다.

<그림1> 변동추이







제2절 분석 모형

1. 시계열 분석 모형

1) 벡터자기회귀 모형(Vector Autoregressive Model: VAR)

회귀분석과 시계열분석을 결합하여 변수 간 상관관계와 인과관계를 추정할 수 있는 다변량 시계열 모형이다. 변수 간 공적분 관계가 존재하지 않을 경우 VAR 모형을 사용하여야 한다. VAR 모형의 경우 시계열의 안정성을 위해 차분 하는 과정에서 장기적인 관계에 대한 정보를 상실하여 모형 설정에 오류가 발생할 가능성이 있다.

2) 벡터오차수정모형(Error Correction Model: VECM)

벡터오차수정모형은 VAR 모형의 확장형 모형이며, 변수 간 공적분 관계가 존재하면 장기적인 균형관계를 고려해야 하는데 균형관계를 이탈하는 부분을 조정하여 수정하여 추정하는 모형이다. 변수 간 공적분 관계가 존재하면 VECM 모형을 사용하여야 한다.

2. 딥러닝 분석 모형

1) 장단기 메모리 (Long Short-Term Memory, LSTM)

순환신경망(Recurrent Neural Network, RNN)은 단기 메모리만을 가지고 학습하기 때문에 기울기 소실 발생으로 과거 학습 결과가 사라지는 장기 의존성 문제가 있다. 이러한 문제를 해결하기 위해 1997년에 LSTM이 고안되었다. LSTM은 메모리와 같은 역할을 하기 위해 Cell State가 존재하고, Cell State를 제어하기 위해 3개의 Gate가 존재한다. Input Gate는 현재 정보를 기억하기 위한 게이트이며, Forget Gate는 어느 정도의 과거 정보를 잊게 할지를 정하는 게이트이고, Output Gate는 어떤 값을 출력할지 정하는 게이트이다. 이러한 Gate는 학습을 통해 어떤 정보를 유지하고 어떤 정보를 버릴지 정하여 다음 상태의 Cell State에 업데이트 한다.

2) 게이트 순환 유닛 (Gated Recurrent Unit, GRU)

LSTM의 경우 RNN에 비하여 복잡한 구조 때문에 파라미터가 많이 필요하고 연산이 오래 걸린다는 단점이 있는데, 이러한 점을 개선하고자 2014년에

GRU가 고안되었다. GRU는 2개의 Gate가 존재한다. Reset Gate는 이전 상태를 얼마나 반영할지 정하는 게이트이며, Update Gate는 이전 상태와 현재 상태의 반영 비율을 정하는 게이트이다. Update Gate가 LSTM의 Forget gate와 Input Gate를 의미한다. GRU에서는 Output Gate가 없다. 따라서 GRU는 LSTM 보다는 연산은 적게 하고, 성능은 유사한 결과를 낸다.

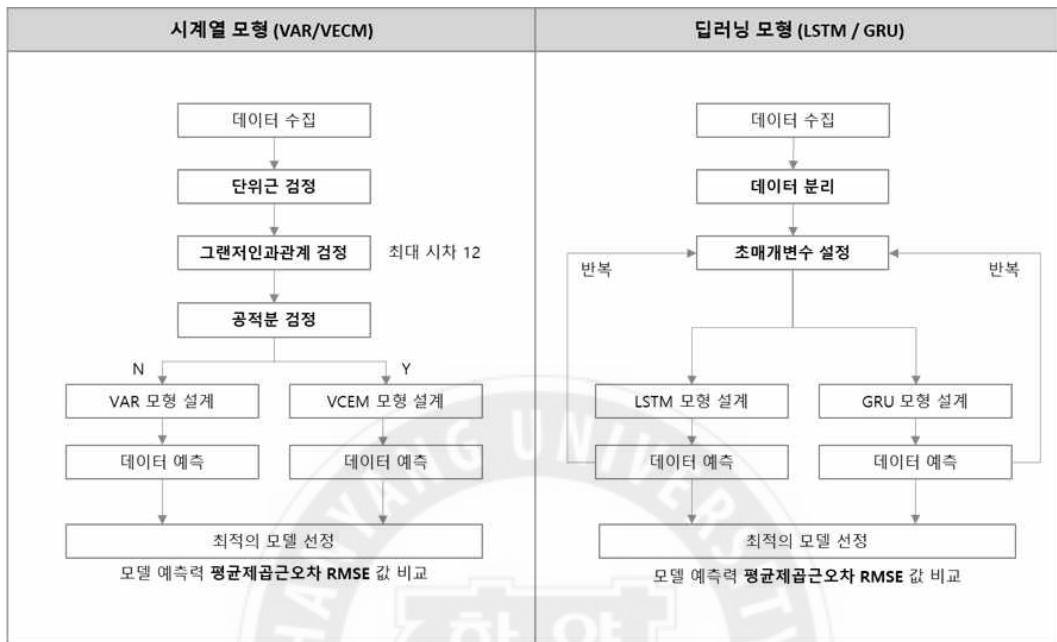
제3절 분석 방법 및 절차

분석방법 및 절차는 <그림2>의 절차로 진행되었다.

시계열 모형 분석 방법의 경우 프로그램은 R Studio을 사용하였고, 수집한 데이터를 단위로 검정을 통해 시계열 안정성을 먼저 확인한다. ADF 검정을 통해서 단위근 검정 결과값을 확인한 후, 시계열이 안정적이지 않을 경우, 데이터를 차분하여 데이터의 안정성을 확보한다. 다음으로 종속변수 주택매매가격지수(KB), 매매실거래가격지수와 독립변수 30개의 거시경제 변수간의 인과관계를 확인하기 위하여 그랜저 인과관계 검정을 진행한다. 인과관계 검정 시 최대시차를 12로 하여 1년간의 기간으로 확인하는 것으로 하였다. 그랜저 인과관계 검정을 통해 인과관계가 확인된 변수를 선정하고, 변수간 공적분 관계가 있는지 공적분 검정을 진행한다. 공적분 검정에 따라 VAR 모형 또는 VECM 모형을 선택하도록 한다. 총 197개월의 데이터 중 191개월을 기반하여 6개월을 예측하는 것으로 한다. 예측 모델의 성능 평가는 실제값과 예측값의 차이를 RMSE 지표로 비교하였다.

딥러닝 모형의 경우 LSTM과 GRU 모형을 이용하였으며, 수집된 데이터 총 197개월 중 191개월을 Train data로 6개월을 Test data로 분리하였다. 초매개 변수(하이퍼파라미터)를 변경하면서 반복적으로 분석하였으며, 프로그램은 Google Colab 환경에서 Python으로 코딩하여 개발하는 것으로 하였다. 딥러닝의 경우 분석시간이 오래 소요되어 GPU 가속기를 이용하여 속도를 개선하도록 했으며, 매번 분석할 때마다 모델의 성능지표가 변경되는 부분이 있어 Random Seed를 설정하여 값이 변경되지 않도록 하였다. 딥러닝 모형 예측 결과 해석은 shap(shapley additive explanations) 중 DeepExplainer 활용하여 모형의 변수 영향도를 확인하였다.

<그림2> 분석방법 및 절차



제4장 실증 분석

제1절 주택매매가격지수(KB)의 예측력 비교

1. 시계열 분석 모형

1) 단위근 검정 (Unit root test)

분석 자료 주택매매가격지수(KB)와 30개의 거시경제변수를 시계열 분석 모형을 설계하기에 앞서 시계열의 안정성을 확인 하였다. ADF 단위근 검정을 진행하였고, 유의수준 0.05 기준으로 검정 결과를 해석하였다.

원시계열 단위근 검정을 진행한 결과는 지가변동률(서울), 선행지수순환변동치, 건축허가현황, 경제활동인구, 실업률, 고용률, 국내총생산, 경상수지, 금융계정 9개 변수만 p-value 값이 유의수준 0.05 보다 낮게 측정되어 귀무가설을 기각하여 단위근이 존재하지 않는 것으로 나타났다. 주택매매가격지수(KB)를 포함한 그 외 다른 변수는 p-value 값이 유의수준 0.05 보다 크게 측정되어 귀무가설을 채택하여 단위근이 존재하였다. 1차 차분하여 전체 변수에 대해서 단위근 검정을 진행한 결과는 아파트 매매가격 대비 전세가비율(서울)과 M2 2개 변수만 p-value 값이 유의수준 0.05 보다 크게 측정되어 귀무가설을 채택하여 단위근이 존재하였다. 다른 변수들은 p-value 값이 유의수준 0.05보다 낮게 측정되어 귀무가설을 기각하여 단위근이 존재하지 않는 것으로 나타났다. 아파트 매매가격 대비 전세가비율(서울)과 M2 2개 변수는 2차 차분을 하여 단위근 검정을 하였으며, p-value 값이 유의수준 0.05 보다 낮게 측정되어 귀무가설을 기각하여 단위근이 존재하지 않는 것으로 나타났다.

<표5> 단위근 검정 - 주택매매가격지수(KB)

지표명	원시계열		차분	
	검정통계량	p-value	검정통계량	p-value
주택매매가격지수(KB)(서울)	-0.0777	0.99	-3.4786	0.04614
행정구역별 아파트매매거래현황(서울)	-2.9491	0.1788	-7.8006	0.01
아파트 매매가격 대비 전세가격비율(서울)	-1.5173	0.7782	-8.0039	0.01
지가변동률(서울)	-4.7661	0.01	-7.761	0.01

연간소득 대비 주택가격 비율(서울)	-2.5284	0.3549	-7.1656	0.01
연간소득 대비 대출금액 비율(서울)	-2.7381	0.2671	-5.9189	0.01
소비자물가지수	-1.7189	0.6938	-3.9673	0.01202
생산자물가지수	-1.5836	0.7504	-3.9434	0.01321
수출물가지수	-1.436	0.8123	-4.2434	0.01
수입물가지수	-1.4924	0.7887	-3.7491	0.02281
선행지수순환변동치	-4.3094	0.01	-4.9066	0.01
동행지수순환변동치	-3.4908	0.04498	-6.0889	0.01
국내건설수주액	-2.257	0.4685	-9.5847	0.01
건축착공현황	-2.0475	0.5562	-9.4419	0.01
건축허가현황	-4.2252	0.01	-9.0582	0.01
경제활동인구	-4.6367	0.01	-7.9229	0.01
실업률	-5.5027	0.01	-7.7497	0.01
고용률	-5.4229	0.01	-7.0498	0.01
국내총생산	-4.0585	0.01	-4.7452	0.01
총저축률	-2.3037	0.449	-6.873	0.01
국내총투자율	-3.7769	0.02142	-7.1632	0.01
경상수지	-5.0643	0.01	-7.7649	0.01
금융계정	-4.4585	0.01	-7.3247	0.01
환율(원/미국달러)	-3.4468	0.0491	-5.3175	0.01
M2	-0.21121	0.99	-7.8301	0.01
국고채(3년)	-0.68955	0.9699	-5.4446	0.01
회사채(3년)	-1.5271	0.7741	-5.316	0.01
주택담보대출 금리	-1.5564	0.7618	-5.5095	0.01
KOSPI	-3.9071	0.01498	-5.1034	0.01
경제심리지수	-3.5259	0.0417	-5.8698	0.01
기업경기실사지수	-3.3192	0.06967	-5.185	0.01

2) 그랜저 인과관계 검정 (Granger causality test)

그랜저 인과관계 검정을 진행하기 위해서는 시계열이 안정성을 가지고 있어야 한다. 단위근 검정 결과를 통해 1차 및 2차 차분한 시계열 데이터를 가지고 최대시차 12를 두어 그랜저 인과관계 검정을 진행하였으며, 분석 결과를 통해 시계열 분석 모형에 사용될 입력변수를 선정하고자 한다. 1차 차분한 변수는 변수명에 DIF를 사용하였으며, 2차 차분한 변수는 변수명에 DIF2를 사용하였다.

주택매매가격지수(KB)와 행정구역별 아파트매매거래현황(서울)에 대해서 그랜저 인과관계 분석한 결과, 2개 변수 간 시차 5에서 양방향으로 인과관계가 있었다. 주택매매가격지수(KB)와 동행지수순환변동치에 대해서 그랜저 인과관계 분석한 결과. 동행지수순환변동치는 주택매매가격지수(KB)에 인과관계가 없는 것으로 나타났고, 주택매매가격지수(KB)는 동행지수순환변동치에 시차 6에서 인과관계가 있는 것으로 나타났다. 주택매매가격지수(KB)와 국내건설수주액에 대해서 그랜저 인과관계 분석한 결과, 국내건설수주액은 주택매매가격지수(KB)에 인과관계가 없는 것으로 나타났고, 주택매매가격지수(KB)는 국내건설수주액에 시차 11에서 인과관계가 있는 것으로 나타났다. 주택매매가격지수(KB)와 실업률에 대해서 그랜저 인과관계 분석한 결과, 실업률이 주택매매가격지수(KB)에 시차 12에서 인과관계가 있는 것으로 나타났고, 주택매매가격주시(KB)가 실업률에 인과관계가 없는 것으로 나타났다.

주택매매가격지수(KB)와 국고채(3년)에 대해서 그랜저 인과관계 분석한 결과 국고채(3년)은 주택매매가격지수(KB)에 시차 1에서 인과관계가 있는 것으로 나타났고, 주택매매가격지수(KB)는 국고채(3년)에 인과관계가 없는 것으로 나타났다. 주택매매가격지수(KB)와 회사채(3년)에 대해서 그랜저 인과관계 분석한 결과, 회사채(3년)은 주택매매가격지수(KB)에 시차 1에서 인과관계가 있는 것으로 나타났고, 주택매매가격지수(KB)는 회사채(3년)에 시차 1에서 인과관계가 있는 것으로 나타났다. 주택매매가격지수(KB)와 주택담보 대출 금리에 대해서 그랜저 인과관계를 분석한 결과 주택담보 대출 금리는 주택매매가격지수(KB)에 시차 2에서 인과관계가 있는 것으로 나타났고, 주택매매가격지수(KB)는 주택담보 대출 금리에 시차 2에서 인과관계가 있는 것으로 나타났다.

주택매매가격지수(KB)와 KOSPI에 대해서 그랜저 인과관계 분석한 결과, KOSPI는 주택매매가격지수(KB)에 인과관계가 없는 것으로 나타났고, 주택매

매가격지수(KB)는 KOSPI에 시차 1에서 인과관계가 있는 것으로 나타났다. 주택매매가격지수(KB)와 경제심리지수에 대해서 그랜저 인과관계 분석한 결과, 경제심리지수는 주택매매가격지수(KB)에 인과관계가 없는 것으로 나타났고, 주택매매가격지수(KB)는 경제심리지수에 시차 1에서 인과관계가 있는 것으로 나타났다. 주택매매가격지수(KB)와 기업경기실사지수에 대해서 그랜저 인과관계 분석한 결과 기업경기실사지수는 주택매매가격지수(KB)에 인과관계가 없는 것으로 나타났고, 주택매매가격지수(KB)는 기업경기실사지수에 시차 1에서 인과관계가 있는 것으로 나타났다.

아파트 매매가격 대비 전세가격비율(서울), 지가변동률(서울), 연간소득 대비 주택가격 비율(서울), 연간소득 대비 대출금액 비율(서울), 소비자물가지수, 생산자물가지수, 수출물가지수, 수입물가지수, 선행지수순환변동치, 건축착공현황, 건축허가현황, 경제활동인구, 고용률, 국내총생산, 총저축률, 국내총투자율, 경상수지, 금융계정, 환율(원/미국달러), M2는 주택매매가격지수(KB)와 양방향 모두 인과관계가 없었다.

그랜저 인과관계 검정 결과를 통해 주택매매가격지수(KB)와 선행관계가 있는 변수는 행정구역별 아파트매매거래현황(서울), 실업률, 국고채(3년), 회사채(3년), 주택담보대출 금리 총 5개 변수이며, 후행관계가 있는 변수는 행정구역별 아파트매매거래현황(서울), 동행지수순환변동치, 국내건설수주액, 회사채(3년), 주택담보대출 금리, KOSPI, 경제심리지수, 기업경기실사지수 총 8개 변수로 나타났다. 다음으로 공적분 검정을 진행하는데, 입력변수 개수 제한이 있어 선행관계가 있는 변수만 사용하는 것으로 하였다.

<표6> 그랜저 인과관계 검정 결과 - 주택매매가격지수(KB)

인과관계	시차	F값	Pr(>F)	인과관계
APC_DIF→KBAPI_DIF KBAPI_DIF→APC_DIF	5	4.4069	0.0008251 ***	선행
	5	4.6111	0.0005536 ***	후행
APR_DIF2→KBAPI_DIF KBAPI_DIF→APR_DIF2	4	2.185	0.0724 .	
	4	1.9234	0.1084	
FLP_DIF→KBAPI_DIF KBAPI_DIF→FLP_DIF	4	0.5559	0.6949	
	4	2.1127	0.08098 .	
PIR_DIF→KBAPI_DIF KBAPI_DIF→PIR_DIF	3	0.6039	0.6133	
	3	0.8458	0.4704	

LIR_DIF→KBAPI_DIF	1	0.1384	0.7103	
KBAPI_DIF→LIR_DIF	1	0.9623	0.3278	
CPI_DIF→KBAPI_DIF KBAPI_DIF→CPI_DIF	4	0.3544	0.8408	
	4	0.6994	0.5933	
PPI_DIF→KBAPI_DIF KBAPI_DIF→PPI_DIF	2	0.6155	0.5414	
	2	0.4957	0.6099	
EPL_DIF→KBAPI_DIF KBAPI_DIF→EPL_DIF	2	1.4398	0.2396	
	2	0.0338	0.9668	
IPL_DIF→KBAPI_DIF KBAPI_DIF→IPL_DIF	1	2.7123	0.1012	
	1	0.2577	0.6123	
CLI_DIF→KBAPI_DIF KBAPI_DIF→CLL_DIF	4	0.5439	0.7037	
	4	2.0546	0.0886 .	
CCI_DIF→KBAPI_DIF KBAPI_DIF→CCI_DIF	6	1.0609	0.3879	
	6	2.3144	0.03551 *	후행
COA_DIF→KBAPI_DIF KBAPI_DIF→COA_DIF	11	0.9843	0.4628	
	11	2.1994	0.01676 *	후행
CW_DIF→KBAPI_DIF KBAPI_DIF→CW_DIF	5	0.575	0.7191	
	5	0.2481	0.9403	
RBP_DIF→KBAPI_DIF KBAPI_DIF→RBP_DIF	12	1.1493	0.3246	
	12	0.4746	0.9273	
EAP_DIF→KBAPI_DIF KBAPI_DIF→EAP_DIF	12	0.9615	0.4879	
	12	1.3862	0.1774	
UR_DIF→KBAPI_DIF KBAPI_DIF→UR_DIF	12	2.1363	0.01738 *	선행
	12	1.5643	0.1071	
ER_DIF→KBAPI_DIF KBAPI_DIF→ER_DIF	12	1.2756	0.2378	
	12	1.5692	0.1056	
GDP_DIF→KBAPI_DIF KBAPI_DIF→GDP_DIF	12	0.5592	0.872	
	12	1.0746	0.3849	
SR_DIF→KBAPI_DIF KBAPI_DIF→SR_DIF	3	0.1534	0.9274	
	3	0.7658	0.5145	
GDIR_DIF→KBAPI_DIF KBAPI_DIF→GDIR_DIF	1	2.1236	0.1467	
	1	1.583	0.2099	
CAB_DIF→KBAPI_DIF KBAPI_DIF→CAB_DIF	11	1.6473	0.09017	
	11	1.1396	0.3339	
FA_DIF→KBAPI_DIF KBAPI_DIF→FA_DIF	12	0.8879	0.5606	
	12	1.5213	0.1214	
ECR_DIF→KBAPI_DIF KBAPI_DIF→ECR_DIF	2	0.0604	0.9414	
	2	1.3373	0.265	

M2_DIF2→KBAPI_DIF	5	0.4965	0.7786	
KBAPI_DIF→M2_DIF2	5	1.967	0.08566 .	
GB_DIF→KBAPI_DIF	1	8.3492	0.004302 **	선행
	1	0.2216	0.6384	
CB_DIF→KBAPI_DIF	1	6.9666	0.008987 **	선행
	1	3.9933	0.04709 *	후행
MIR_DIF→KBAPI_DIF	2	8.4193	0.0003143 ***	선행
	2	4.4007	0.03723 *	후행
KOSPI_DIF→KBAPI_DIF	1	0.3811	0.5378	
	1	7.0142	0.008759 **	후행
ESI_DIF→KBAPI_DIF	1	0.1319	0.7168	
	1	12.542	0.0004993 ***	후행
BSI_DIF→KBAPI_DIF	1	0.2846	0.5943	
	1	8.5162	0.00394 **	후행

3) 공적분 검정 (Cointegration test)

분석하고자 하는 변수간의 공적분 관계가 있으면 VECM 모형으로 예측 모형을 설계해야 하므로 공적분 검정을 진행하였다. 우선 앞서 그랜저 인과관계 검정을 통해 확인한 선행관계 변수 행정구역별 아파트매매거래현황(서울), 실업률, 국고채(3년), 회사채(3년), 주택담보대출 금리 5개 변수를 가지고 VAR 모델에서 최대 시차 12로 설정하여 확인하였고, AIC가 최소가 되는 시차는 12였다.

다음으로 공적분 검정하기 위하여 요한센 검정 방법을 사용하였다. 위에서 확인한 시차 12를 입력하였고, type은 trace으로 확인하였다. 공적분 없다($r=0$) test 검정통계량 129.74 값이 유의수준 5% 102.14 값보다 크므로 공적분 관계가 있다고 할 수 있다. 적어도 공적분 1개 이하 ($r \leq 1$)에서 test 검정통계량 69.92 값이 유의수준 5% 76.07 값보다 작으므로 $r=1$ 라고 할 수 있다.

<표7> 공적분 검정 결과 - 주택매매가격지수(KB)

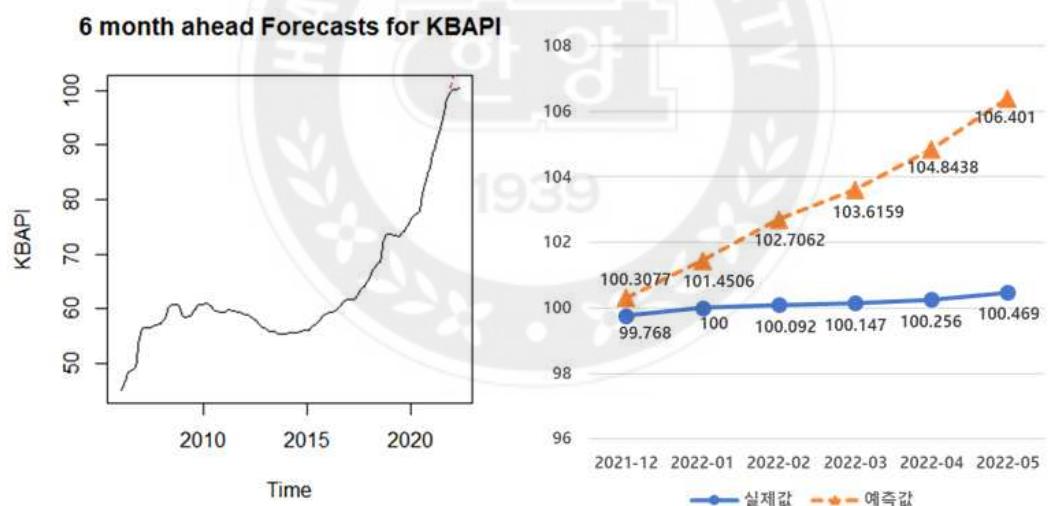
	test	10pct	5pct	1pct
$r \leq 5$	3.6	7.52	9.24	12.97
$r \leq 4$	11.13	17.85	19.96	24.60
$r \leq 3$	22.34	32.00	34.91	41.07
$r \leq 2$	42.83	49.65	53.12	60.16
$r \leq 1$	69.92	71.86	76.07	84.45
$r=0$	129.74	97.18	102.14	111.01

4) VECM 모형

공적분 검정을 통해 주택매매가격지수(KB)와 행정구역별 아파트매매거래현황(서울), 실업률, 국고채(3년), 회사채(3년), 주택담보대출 금리 5개 변수 간 장기적인 균형관계가 존재하는 것을 확인하였다. 다변량 데이터 분석에서 공적분 관계가 있기 때문에 VECM 모형을 이용하였고, 원시계열 그대로 입력하여 6개월간의 시계열 데이터를 예측하고 실제 값과의 차이를 RMSE 지표로 모형 성능평가 진행하였다.

2006년 1월부터 2021년 11월까지를 191개월 데이터를 입력하였고, VAR 모델에서 최대 시차 12로 설정하여 확인하였고, AIC가 최소가 되는 시차는 12로 입력하여 VECM 모형으로 6개월을 예측하였다. 주택매매가격지수(KB) 실제 값과 예측값 비교 RMSE 값은 3.59395로 측정되었다.

<그림3> VECM 예측 결과 - 주택매매가격지수(KB)



2. 딥러닝 분석 모형

LSTM과 GRU 모형은 VECM 모형과 동일하게 원시계열 그대로 사용했으며, 다만 모형에서 입력변수에 제한이 없기 때문에 주택매매가격지수(KB)를 예측하기 위하여 30개 거시경제변수를 모두 사용하였다. LSTM과 GRU 모형 최적화를 위해 하이퍼파라미터를 변경하면서 최적의 모형을 찾으려고 하였다. 또한 모형의 성능평가는 RMSE 지표로 측정하였다. 또한 매번 분석 할 때마다 모형의 성능 측정값이 변하지 않도록 Random seed(50)를 설정하여 분석을 진행하였다. 하이퍼파라미터는 Hidden Layer와 Hidden Units은 2(128, 64), 3(128, 256, 64), 4(128, 256, 128, 64)으로 하였고, sequence는 1 ~ 5, batch_size는 5 ~ 50, epoch는 100 ~ 1000, dropout은 0으로 설정하여 반복 테스트를 진행하였다.

<표8> LSTM 모형 결과 예시 - 주택매매가격지수(KB)

모형	Hidden Layer	Sequence	Epoch	Dropout	Batch Size	RMSE
LSTM	4	2	400	0	15	0.279538
LSTM	4	2	500	0	10	0.32007
LSTM	3	2	1000	0	15	0.33156
LSTM	2	3	700	0	10	0.347191
LSTM	2	3	200	0	5	0.372599
LSTM	4	3	200	0	10	0.390548
LSTM	4	3	200	0	15	0.390581

<표9> GRU 모형 결과 예시 - 주택매매가격지수(KB)

모형	Hidden Layer	Sequence	Epoch	Dropout	Batch Size	RMSE
GRU	4	2	500	0	35	0.285579
GRU	4	2	300	0	15	0.321141
GRU	4	2	200	0	10	0.341151
GRU	3	2	500	0	30	0.343658
GRU	3	2	500	0	45	0.373886
GRU	2	2	400	0	20	0.374942
GRU	2	2	900	0	10	0.375183

1) LSTM 모형

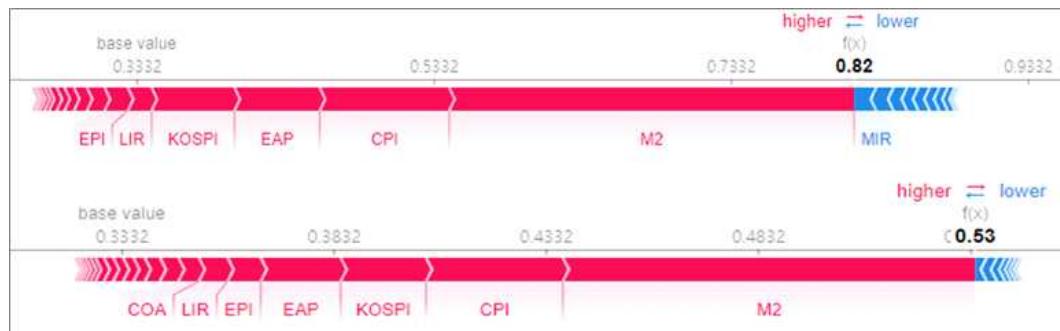
2006년 1월부터 2021년 11월까지를 191개월을 Train data, 2021년 12월부터 2022년 5월까지 6개월을 Test data로 분리하였다. 가장 최적의 모형은 Hidden Layer와 Hidden Units은 4(128, 256, 128, 64), sequence 2, batch_size 15, epoch 400, dropout 0 로 설정하였을 때 주택매매가격지수(KB) 실제값과 예측값의 비교 RMSE 값이 0.27952로 측정되었다. LSTM 모형에서 Train data 및 Test data 의 성능 지표는 아래 <표8>와 같다.

<표10> LSTM 최적 모형 결과 - 주택매매가격지수(KB)

구분	MAE	MSE	MAPE	RMSE
Train data	0.27779	0.12798	0.43031	0.35775
Test data	0.21489	0.07813	0.21453	0.27952

LSTM 모형 예측을 해석하기 위하여 shap 중 DeepExplainer 활용하여 변수 중요도를 확인하였다. $f(x)$ 를 기준으로 왼쪽의(higher) 변수들은 예측값을 높게 하는 변수의 영향도를 나타내며, 오른쪽의(lower) 변수들은 예측값을 낮게 하는 변수의 영향도를 나타낸다. sequence가 2이기 때문에 2번의 변수 중요도를 확인하였다. 30개의 거시경제 변수 중 M2, 소비자물가지수(CPI), 경제활동인구(EAP), KOSPI, 연간소득 대비 대출금액 비율(서울)(LIR), 수출물가지수(EPI)가 중요한 것으로 나타났고, 주택담보대출 금리(MIR), 수입물가지수(IPI)는 중요하지 않은 것으로 나타났다.

<그림4> LSTM 변수 중요도 결과 - 주택매매가격지수(KB)



2) GRU 모형

2006년 1월부터 2021년 11월까지를 191개월을 Train data, 2021년 12월부터 2022년 5월까지 6개월을 Test data로 분리하였다. 가장 최적의 모형은 Hidden Layer와 Hidden Units은 4(128, 256, 128, 64), sequence 2, batch_size 35, epoch 500, dropout 0 로 설정하였을 때 주택매매가격지수(KB) 실제값과 예측값의 비교 RMSE 값이 0.28557로 측정되었다. GRU 모형에서 Train data 및 Test data 의 성능 지표는 아래 표 <표9>와 같다.

<표11> GRU 최적 모형 결과 - 주택매매가격지수(KB)

구분	MAE	MSE	MAPE	RMSE
Train data	0.48575	0.31232	0.77937	0.55885
Test data	0.24792	0.08155	0.24728	0.28557

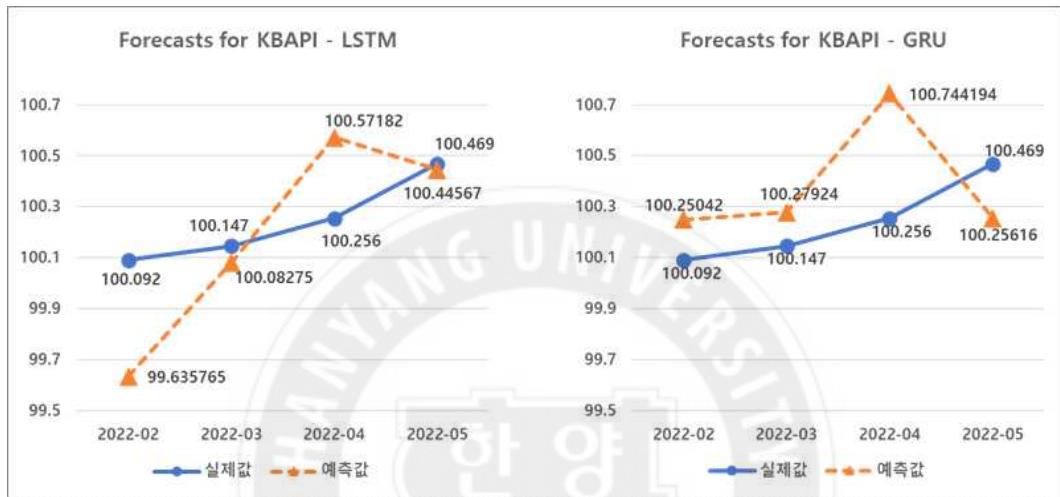
GRU 모형 예측을 해석하기 위하여 shap 중 DeepExplainer 활용하여 변수 중요도를 확인하였다. $f(x)$ 를 기준으로 왼쪽의(higher) 변수들은 예측값을 높게 하는 변수의 영향도를 나타내며, 오른쪽의(lower) 변수들은 예측값을 낮게 하는 변수의 영향도를 나타낸다. sequence가 2이기 때문에 2번의 변수 중요도를 확인하였다. 30개의 거시경제 변수 중 M2, 소비자물가지수(CPI), 경제활동 인구(EAP), KOSPI, 수출물가지수(EPI), 연간소득 대비 주택가격 비율(서울)(PIR)이 중요한 것으로 나타났고, 생산자물가지수(PPI), 수입물가지수(IPI), 주택담보대출 금리(MIR)는 중요하지 않은 것으로 나타났다.

<그림5> GRU 변수 중요도 결과 - 주택매매가격지수(KB)



LSTM에서 RMSE 값이 최소가 되는 sequence는 2이기 때문에, 4개월의 실제값과 예측값을 비교하였다. GRU 모형에서 RMSE 값이 최소가 되는 sequence는 2이기 때문에 4개월의 실제값과 예측값을 비교하였다.

<그림6> LSTM 및 GRU 예측 결과 - 주택매매가격지수(KB)



주택매매가격지수(KB) 예측력을 RMSE로 비교한 결과 시계열 모형 VECM 보다 딥러닝 모형 LSTM과 GRU 모형이 예측력이 높은 것으로 나타났고, GRU 모형보다 LSTM 모형이 예측력이 높은 것으로 나타났다.

<표 12> 주택매매가격지수(KB) 예측력 비교

구분	모형	RMSE
주택매매가격지수(KB)	VECM	3.59395
	LSTM	0.27952
	GRU	0.28557

제2절 매매실거래가격지수의 예측력 비교

1. 시계열 분석 모형

1) 단위근 검정 (Unit root test)

분석 자료 매매실거래가격지수와 30개의 거시경제변수를 시계열 분석 모형을 설계하기에 앞서 시계열의 안정성을 확인 하였다. ADF 단위근 검정을 진행하였고, 유의수준 0.05 기준으로 검정 결과를 해석하였다. 앞서 주택매매가격지수(KB)와 동일한 거시경제 변수를 사용하기 때문에 자세한 설명은 생략하고, 단위근이 존재하는 원시계열을 1차 및 2차 차분하여 안정적인 시계열로 변환 후 사용하였다.

<표13> 단위근 검정 - 매매실거래가격지수

지표명	원시계열		차분	
	검정통계량	p-value	검정통계량	p-value
매매실거래가격지수(서울)	-0.63662	0.9747	-4.5955	0.01
행정구역별 아파트 매매거래현황(서울)	-2.9491	0.1788	-7.8006	0.01
아파트 매매가격 대비 전세가격비율(서울)	-1.5173	0.7782	-8.0039	0.01
지가변동률(서울)	-4.7661	0.01	-7.761	0.01
연간소득 대비 주택가격 비율(서울)	-2.5284	0.3549	-7.1656	0.01
연간소득 대비 대출금액 비율(서울)	-2.7381	0.2671	-5.9189	0.01
소비자물가지수	-1.7189	0.6938	-3.9673	0.01202
생산자물가지수	-1.5836	0.7504	-3.9434	0.01321
수출물가지수	-1.436	0.8123	-4.2434	0.01
수입물가지수	-1.4924	0.7887	-3.7491	0.02281
선행지수순환변동치	-4.3094	0.01	-4.9066	0.01
동행지수순환변동치	-3.4908	0.04498	-6.0889	0.01
국내건설수주액	-2.257	0.4685	-9.5847	0.01
건축착공현황	-2.0475	0.5562	-9.4419	0.01
건축허가현황	-4.2252	0.01	-9.0582	0.01
경제활동인구	-4.6367	0.01	-7.9229	0.01

실업률	-5.5027	0.01	-7.7497	0.01
고용률	-5.4229	0.01	-7.0498	0.01
국내총생산	-4.0585	0.01	-4.7452	0.01
총저축률	-2.3037	0.449	-6.873	0.01
국내총투자율	-3.7769	0.02142	-7.1632	0.01
경상수지	-5.0643	0.01	-7.7649	0.01
금융계정	-4.4585	0.01	-7.3247	0.01
환율(원/미국달러)	-3.4468	0.0491	-5.3175	0.01
M2	-0.21121	0.99	-7.8301	0.01
국고채(3년)	-0.68955	0.9699	-5.4446	0.01
회사채(3년)	-1.5271	0.7741	-5.316	0.01
주택담보대출 금리	-1.5564	0.7618	-5.5095	0.01
KOSPI	-3.9071	0.01498	-5.1034	0.01
경제심리지수	-3.5259	0.0417	-5.8698	0.01
기업경기실사지수	-3.3192	0.06967	-5.185	0.01

2) 그랜저 인과관계 검정 (Granger causality test)

매매실거래가격지수와 행정구역별 아파트매매거래현황(서울)에 대해서 그랜저 인과관계 분석한 결과, 행정구역별 아파트매매거래현황(서울)은 매매실거래가격지수에 시차 6에서 인과관계가 있는 것으로 나타났고, 매매실거래가격지수는 행정구역별 아파트매매거래현황(서울)에 인과관계가 없는 것으로 나타났다. 매매실거래가격지수와 아파트매매가격 대비 전세가격비율(서울)에 대해서 그랜저 인과관계 분석한 결과, 아파트 매매가격 대비 전세가격비율(서울)은 매매실거래가격지수에 시차 4에서 인과관계가 있는 것으로 나타났고, 매매실거래가격지수는 아파트 매매가격대비 전세가격비율(서울)에 인과 관계가 없는 것으로 나타났다. 매매실거래가격지수와 지가변동률(서울)에 대해서 그랜저 인과관계 분석한 결과, 2개 변수 간 시차 5에서 양방향으로 인과관계가 있었다.

매매실거래가격지수와 연간소득 대비 주택가격 비율(서울)에 대해서 그랜저 인과관계 분석한 결과, 연간소득 대비 주택가격 비율(서울)은 매매실거래가격지수에 시차 3에서 인과관계가 있는 것으로 나타났고, 매매실거래가격지수는 연간소득 대비 주택가격비율(서울)에 인과관계가 없는 것으로 나타났다. 매매실거래가격지수와 소비자물가지수에 대해서 그랜저 인과관계 분석한 결과, 소

비자물가지수는 매매실거래가격지수에 인과관계가 없는 것으로 나타났고, 매매실거래가격지수는 소비자물가지수에 시차 7에 인과관계가 있는 것으로 나타났다. 매매실거래가격지수와 생산자물가지수에 대해서 그랜저 인과관계 분석한 결과, 생산자물가지수는 매매실거래가격지수에 인과관계가 없는 것으로 나타났고, 매매실거래가격지수는 생산자물가지수에 시차 7에 인과관계가 있는 것으로 나타났다. 매매실거래가격지수와 수출물가지수에 대해서 그랜저 인과관계 분석한 결과, 수출물가지수는 매매실거래가격지수에 시차 4에 인과관계가 있는 것으로 나타났고, 매매실거래가격지수는 수출물가지수에 인과관계가 없는 것으로 나타났다. 매매실거래가격지수와 수입물가지수에 대해서 그랜저 인과관계 분석한 결과, 2개 변수 간 시차 4에서 양방향으로 인과관계가 있었다.

매매실거래가격지수와 선행지수순환변동치에 대해서 그랜저 인과관계 분석한 결과, 선행지수순환변동치는 매매실거래가격지수에 시차 10에 인과관계가 있는 것으로 나타났고, 매매실거래가격지수는 선행지수순환변동치에 인과관계가 없는 것으로 나타났다. 매매실거래가격지수와 국내건설수주액에 대해서 그랜저 인과관계 분석한 결과, 국내건설수주액은 매매실거래가격지수에 영향이 없는 것으로 나타났고, 매매실거래가격지수는 국내건설수주액에 시차 11에 인과관계가 있는 것으로 나타났다. 매매실거래가격지수와 고용률에 대해서 그랜저 인과관계 분석한 결과, 고용률은 매매실거래가격지수에 영향이 없는 것으로 나타났고, 매매실거래가격지수는 고용률에 시차 12에 인과관계가 있는 것으로 나타났다.

매매실거래가격지수와 국내총투자율에 대해서 그랜저 인과관계 분석한 결과, 2개 변수 간 시차 3에서 양방향으로 인과관계가 있었다. 매매실거래가격지수와 환율(원/미국달러)에 대해서 그랜저 인과관계 분석한 결과, 2개 변수 간 시차 5에서 양방향으로 인과관계가 있었다. 매매실거래가격지수와 M2에 대해서 그랜저 인과관계 분석한 결과, M2는 매매실거래가격지수에 인과관계가 없는 것으로 나타났고, 매매실거래가격지수는 M2에 시차 5에 인과관계가 있는 것으로 나타났다. 매매실거래가격지수와 국고채(3년)에 대해서 그랜저 인과관계 분석한 결과, 2개 변수 간 시차 4에서 양방향으로 인과관계가 있었다. 매매실거래가격지수와 회사채(3년)에 대해서 그랜저 인과관계 분석한 결과, 회사채(3년)은 매매실거래가격지수에 인과관계가 없는 것으로 나타났고, 매매실거래

가격지수는 회사채(3년)에 시차 7에 인과관계가 있는 것으로 나타났다. 매매실거래가격지수와 주택담보대출 금리에 대해서 그랜저 인과관계 분석한 결과, 주택담보대출 금리는 매매실거래가격지수에 시차 5에 인과관계가 있는 것으로 나타났고, 매매실거래가격지수는 주택담보대출 금리에 인과관계가 없는 것으로 나타났다.

매매실거래가격지수와 KOSPI에 대해서 그랜저 인과관계 분석한 결과, KOSPI는 매매실거래가격지수에 인과관계가 없는 것으로 나타났고, 매매실거래가격지수는 KOSPI에 시차 2에 인과관계가 있는 것으로 나타났다.

동행지수순환변동치, 건축착공현황, 건축허가현황, 경제활동인구, 실업률, 국내총생산, 총저축률, 경상수지, 금융계정, 경제심리지수, 기업경기실사지수는 매매실거래가격지수와 양방향 모두 인과관계가 없었다.

그랜저 인과관계 검정 결과를 통해 매매실거래가격지수와 선행관계가 있는 변수는 행정구역별 아파트매매 거래현황(서울), 아파트 매매가격 대비 전세가격비율(서울), 지가변동률(서울), 연간소득 대비 주택가격 비율(서울), 연간소득 대비 대출금액 비율(서울), 수출물가지수, 수입물가지수, 선행지수순환변동치, 국내총투자율, 환율(원/미국달러), 국고채(3년), 주택담보대출 금리 총 12개 변수이며, 후행관계가 있는 변수는 지가변동률(서울), 소비자물가지수, 생산자물가지수, 수입물가지수, 국내건설수주액, 고용률, 국내총투자율, 환율(원/미국달러), M2, 국고채(3년), 회사채(3년), KOSPI 총 12개 변수로 나타났다. 다음으로 공적분 검정을 진행하는데, 입력변수 개수 제한이 있어 선행관계가 있는 변수 중에서 가장 인과관계가 높은 변수 행정구역별 아파트매매 거래현황(서울), 주택담보대출 금리, 국고채(3년), 수입물가지수, 연간소득 대비 주택가격비율(서울) 총 5개 변수를 사용하였다.

<표14> 그랜저 인과관계 검정 결과 - 매매실거래가격지수

인과관계	시차	F값	Pr(>F)	인과관계
APC_DIF→API_DIF API_DIF→APC_DIF	6	7.1001	8.596e-07 ***	선행
	6	1.9026	0.08278 .	
APR_DIF2→API_DIF API_DIF→APR_DIF2	4	3.3652	0.011 *	선행
	4	0.9303	0.4475	
FLP_DIF→API_DIF API_DIF→FLP_DIF	5	2.9729	0.01326 *	선행
	5	4.3479	0.0009259 ***	후행

PIR_DIF→API_DIF API_DIF→PIR_DIF	3	4.7672	0.003168 **	선행
	3	0.3565	0.7845	
LIR_DIF→API_DIF API_DIF→LIR_DIF	3	2.6715	0.04883 *	선행
	3	1.234	0.2987	
CPI_DIF→API_DIF API_DIF→CPI_DIF	7	1.4108	0.2036	
	7	2.4789	0.0189 *	후행
PPI_DIF→API_DIF API_DIF→PPI_DIF	7	1.4471	0.1894	
	7	4.2176	0.0002539 ***	후행
EPI_DIF→API_DIF API_DIF→EPI_DIF	4	3.9602	0.004164 **	선행
	4	1.4547	0.2179	
IPI_DIF→API_DIF API_DIF→IPI_DIF	4	5.3071	0.0004574 ***	선행
	4	3.4115	0.01019 *	후행
CLI_DIF→API_DIF API_DIF→CLI_DIF	10	2.789	0.003259 **	선행
	10	1.0208	0.4281	
CCI_DIF→API_DIF API_DIF→CCI_DIF	7	1.1598	0.3284	
	7	1.0989	0.3659	
COA_DIF→API_DIF API_DIF→COA_DIF	11	1.3706	0.1913	
	11	2.1001	0.02298 *	후행
CW_DIF→API_DIF API_DIF→CW_DIF	5	0.1733	0.9723	
	5	0.8525	0.5143	
RBP_DIF→API_DIF API_DIF→RBP_DIF	12	1.5656	0.1067	
	12	0.773	0.6775	
EAP_DIF→API_DIF API_DIF→EAP_DIF	12	0.672	0.7764	
	12	1.3237	0.2098	
UR_DIF→API_DIF API_DIF→UR_DIF	12	1.4378	0.1539	
	12	0.9643	0.4853	
ER_DIF→API_DIF API_DIF→ER_DIF	12	1.4442	0.1512	
	12	1.8418	0.0457 *	후행
GDP_DIF→API_DIF API_DIF→GDP_DIF	12	1.0613	0.3964	
	12	1.4164	0.1633	
SR_DIF→API_DIF API_DIF→SR_DIF	3	0.255	0.8577	
	3	0.4635	0.7081	
GDIR_DIF→API_DIF API_DIF→GDIR_DIF	3	3.3508	0.02019 *	선행
	3	5.1897	0.001824 **	후행
CAB_DIF→API_DIF API_DIF→CAB_DIF	11	1.5479	0.1192	
	11	1.4806	0.1432	
FA_DIF→API_DIF API_DIF→FA_DIF	12	1.0797	0.3807	

	12	1.6098	0.0936 .	
ECR_DIF→API_DIF API_DIF→ECR_DIF	5	2.3733	0.04091 * .	선행
	5	3.7101	0.02627 * .	후행
M2_DIF2→API_DIF API_DIF→M2_DIF2	5	0.6795	0.6395 .	
	5	3.3455	0.006499 ** .	후행
GB_DIF→API_DIF API_DIF→GB_DIF	4	6.5813	5.706e-05 *** .	선행
	4	4.9616	0.0008062 *** .	후행
CB_DIF→API_DIF API_DIF→CB_DIF	7	3.0064	0.08454 .	
	7	3.523	0.001458 ** .	후행
MIR_DIF→API_DIF API_DIF→MIR_DIF	5	6.3183	1.984e-05 *** .	선행
	5	2.1505	0.06154 .	
KOSPI_DIF→API_DIF API_DIF→KOSPI_DIF	2	0.1698	0.8439 .	
	2	4.063	0.01872 * .	후행
ESI_DIF→API_DIF API_DIF→ESI_DIF	3	0.9431	0.4209 .	
	3	0.8132	0.488 .	
BSI_DIF→API_DIF API_DIF→BSI_DIF	3	2.0075	0.1144 .	
	3	1.8657	0.1369 .	

3) 공적분 검정 (Cointegration test)

분석하고자 하는 변수간의 공적분 관계가 있으면 VECM 모형으로 예측 모형을 설계해야 하므로 공적분 검정을 진행하였다. 우선 앞서 그랜저 인과관계 검정을 통해 확인한 선행관계 변수 행정구역별 아파트매매거래현황(서울), 주택담보대출 금리, 국고채(3년), 수입물가지수, 연간소득 대비 주택가격비율(서울) 5개 변수를 가지고 VAR 모델에서 최대 시차 12로 설정하여 확인하였고, AIC가 최소가 되는 시차는 3 이었다.

다음으로 공적분 검정하기 위하여 요한센 검정 방법을 사용하였다. 위에서 확인한 시차 3을 입력하였고, type은 trace으로 확인하였다. 공적분 없다($r=0$) test 검정통계량 175.58 값이 유의수준 5% 102.14 값보다 크므로 공적분 관계가 있다고 할 수 있다. 적어도 공적분 5개 이하 ($r \leq 5$)에서 test 검정통계량 5.99 값이 유의수준 5% 9.24 값보다 작으므로 $r=5$ 라고 할 수 있다.

<표15> 공적분 검정 결과 - 매매실거래가격지수

	test	10pct	5pct	1pct
$r \leq 5$	5.99	7.52	9.24	12.97
$r \leq 4$	22.48	17.85	19.96	24.60
$r \leq 3$	43.52	32.00	34.91	41.07

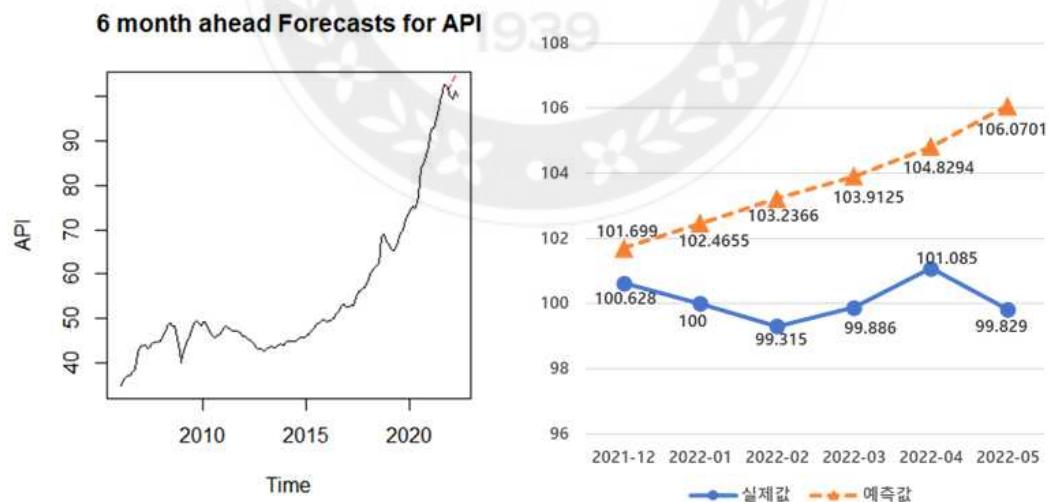
r<=2	70.46	49.65	53.12	60.16
r<=1	111.11	71.86	76.07	84.45
r=0	175.58	97.18	102.14	111.01

4) VECM 모형

공적분 검정을 통해 매매실거래가격지수와 행정구역별 아파트매매거래현황(서울), 주택담보대출 금리, 국고채(3년), 수입물가지수, 연간소득 대비 주택가격비율(서울) 5개 변수 간 장기적인 균형관계가 존재하는 것을 확인하였다. 다변량 데이터 분석에서 공적분 관계가 있기 때문에 VECM 모형을 이용하였고, 원시계열 그대로 입력하여 6개월간의 시계열 데이터를 예측하고 실제 값과의 차이를 RMSE 지표로 모형 성능평가 진행하였다.

2006년 1월부터 2021년 11월까지를 191개월 데이터를 입력하였고, VAR 모델에서 최대 시차 12로 설정하여 확인하였고, AIC가 최소가 되는 시차는 4로 입력하여 VECM 모형으로 6개월을 예측하였다. 매매실거래가격지수 실제값과 예측값 비교 RMSE 값은 3.91130로 측정되었다.

<그림7> VECM 예측 결과 - 매매실거래가격지수



2. 딥러닝 분석 모형

LSTM과 GRU 모형은 VECM 모형과 동일하게 원시계열 그대로 사용했으며, 다만 모형에서 입력변수에 제한이 없기 때문에 매매실거래가격지수를 예측하기 위하여 30개 거시경제변수를 모두 사용하였다. LSTM과 GRU 모형 최적화를 위해 하이퍼파라미터를 변경하면서 최적의 모형을 찾으려고 하였다. 또한 모형의 성능평가는 RMSE 지표로 측정하였다. 또한 매번 분석 할 때마다 모형의 성능 측정값이 변하지 않도록 Random seed(50)를 설정하여 분석을 진행하였다. 하이퍼파라미터는 Hidden Layer와 Hidden Units은 2(128, 64), 3(128, 256, 64), 4(128, 256, 128, 64)으로 하였고, sequence는 1 ~ 5, batch_size는 5 ~ 50, epoch는 100 ~ 1000, dropout은 0으로 설정하여 반복테스트를 진행하였다.

<표16> LSTM 모형 결과 예시 - 매매실거래가격지수

모형	Hidden Layer	Sequence	Epoch	Dropout	Batch Size	RMSE
LSTM	3	2	600	0	20	0.504805
LSTM	4	2	600	0	25	0.56907
LSTM	3	2	600	0	30	0.58556
LSTM	4	2	900	0	5	0.591248
LSTM	4	1	1000	0	5	0.600703
LSTM	4	2	600	0	10	0.618789
LSTM	4	2	500	0	10	0.64226

<표17> GRU 모형 결과 예시 - 매매실거래가격지수

모형	Hidden Layer	Sequence	Epoch	Dropout	Batch Size	RMSE
GRU	4	1	900	0	20	0.248764
GRU	3	1	900	0	10	0.453835
GRU	2	2	800	0	35	0.476337
GRU	2	2	900	0	35	0.490458
GRU	2	2	500	0	25	0.539134
GRU	2	2	500	0	15	0.560387
GRU	2	2	900	0	40	0.582048

1) LSTM 모형

2006년 1월부터 2021년 11월까지를 191개월을 Train data, 2021년 12월부터 2022년 5월까지 6개월을 Test data로 분리하였다. 가장 최적의 모형은 Hidden Layer와 Hidden Units은 3(128, 256, 64), sequence 2, batch_size 20, epoch 600, dropout 0 로 설정하였을 때 매매실거래가격지수 실제값과 예측값의 비교 RMSE 값이 0.50480로 측정되었다. LSTM 모형에서 Train data 및 Test data 의 성능 지표는 아래 <표13> 와 같다.

<표18> LSTM 최적 모형 결과 - 매매실거래가격지수

구분	MAE	MSE	MAPE	RMSE
Train data	0.63579	0.73357	1.05601	0.85649
Test data	0.40236	0.25482	0.40074	0.50480

LSTM 모형 예측을 해석하기 위하여 shap 중 DeepExplainer 활용하여 변수 중요도를 확인하였다. $f(x)$ 를 기준으로 왼쪽의(higher) 변수들은 예측값을 높게 하는 변수의 영향도를 나타내며, 오른쪽의(lower) 변수들은 예측값을 낮게 하는 변수의 영향도를 나타낸다. sequence가 2이기 때문에 2번의 변수 중요도를 확인하였다. 30개의 거시경제 변수 중 M2, KOSPI, 소비자물가지수(CPI), 경제활동인구(EAP)가 중요한 것으로 나타났고, 수입물가지수(IPI), 경제심리지수(ESI), 연간소득 대비 주택가격 비율(서울)(PIR), 행정구역별 아파트 매매거래 건수(서울)(APC)은 중요하지 않은 것으로 나타났다.

<그림8> LSTM 변수 중요도 결과 - 매매실거래가격지수



2) GRU 모형

2006년 1월부터 2021년 11월까지를 191개월을 Train data, 2021년 12월부터 2022년 5월까지 6개월을 Test data로 분리하였다. 가장 최적의 모형은 Hidden Layer와 Hidden Units은 4(128, 256, 128, 64), sequence 1, batch_size 20, epoch 900, dropout 0으로 설정하였을 때 매매실거래가격지수실제값과 예측값의 비교 RMSE 값이 0.24876로 측정되었다. GRU 모형에서 Train data 및 Test data 의 성능 지표는 아래 <표14>와 같다.

<표14> GRU 최적 모형 결과 - 매매실거래가격지수

구분	MAE	MSE	MAPE	RMSE
Train data	0.62287	0.70791	1.04171	0.84137
Test data	0.18454	0.06188	0.18438	0.24876

GRU 모형 예측을 해석하기 위하여 shap 중 DeepExplainer 활용하여 변수 중요도를 확인하였다. $f(x)$ 를 기준으로 왼쪽의(higher) 변수들은 예측값을 높게 하는 변수의 영향도를 나타내며, 오른쪽의(lower) 변수들은 예측값을 낮게 하는 변수의 영향도를 나타낸다. sequence가 1이기 때문에 1번의 변수 중요도를 확인하였다. 30개의 거시경제 변수 중 M2, KOSPI, 소비자물가지수(CPI), 경제활동인구(EAP)가 중요한 것으로 나타났고, 수입물가지수(IPI), 연간소득대비 주택가격 비율(서울)(PIR), 환율(원/달러)(ECR) 은 중요하지 않은 것으로 나타났다.

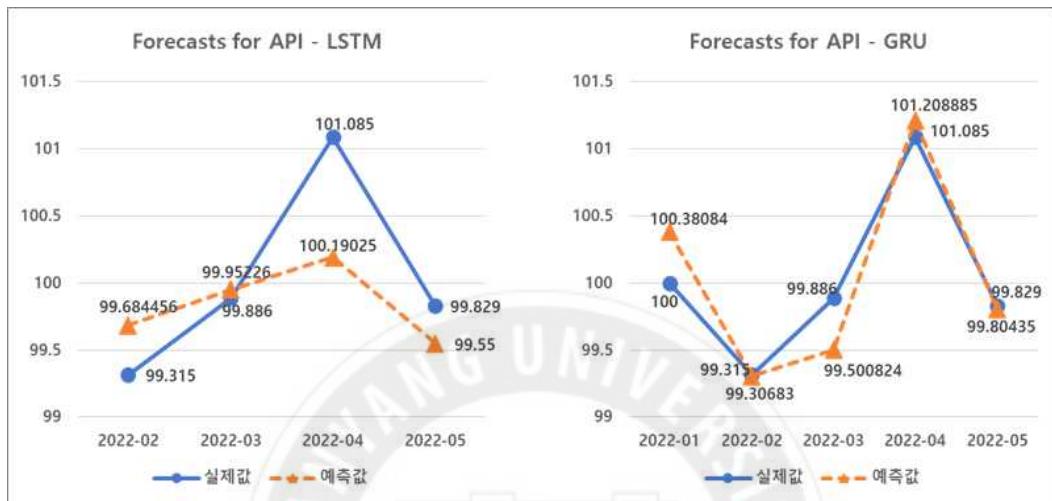
<그림9> GRU 변수 중요도 결과 - 매매실거래가격지수



LSTM 및 GRU 모형은 모형의 sequence에 따라 예측의 값 범위가 상이하며, LSTM에서 RMSE 값이 최소가 되는 sequence는 2이기 때문에, 4개월의 실제값과 예측값을 비교하였다. GRU 모형에서 RMSE 값이 최소가

되는 sequence는 1이기 때문에 5개월의 실제값과 예측값을 비교하였다.

<그림10> LSTM 및 GRU 결과 - 매매실거래가격지수



매매실거래가격지수 예측력을 RMSE로 비교한 결과 시계열 모형 VECM 보다 딥러닝 모형 LSTM과 GRU 모형이 예측력이 높은 것으로 나타났고, LSTM 모형보다 GRU 모형이 예측력이 높은 것으로 나타났다.

<표20> 매매실거래가격지수 예측력 비교

구분	모형	RMSE
매매실거래가격지수	VECM	3.91130
	LSTM	0.50480
	GRU	0.24876

제5장 결론 및 시사점

제1절 요약 및 결론

본 연구는 서울시 아파트 매매 가격을 나타내는 2가지 대표 지수인 주택매매가격지수(KB)와 매매실거래가격지수에 대해서 시계열 분석 모형과 딥러닝 분석 모형으로 예측하고 어떠한 모형이 예측력이 더 좋은지 비교 연구하는 것이 목적이이다.

시계열 분석 모형 설계에 앞서 그랜저 인과관계 검정을 통해 각 지수와 거시경제 변수와의 인과관계를 확인하였다. 서울시 아파트 주택매매가격지수(KB)의 경우 행정구역별 아파트매매거래현황(서울), 실업률, 국고채(3년), 회사채(3년), 주택담보대출 금리, 동행지수순환변동치, 국내건설수주액, KOSPI, 경제심리지수, 기업경기실사지수 30개 거시경제 변수 중 10개 변수가 선행, 후행 관계가 있는 것으로 나타났다. 선행관계에서는 행정구역별 아파트매매거래 건수(서울), 주택담보대출 금리, 국고채(3년), 회사채(3년), 실업률 순으로 인과관계가 높은 것으로 나타났고, 후행관계에서는 행정구역별 아파트매매거래 건수(서울), 경제심리지수, KOSPI, 기업경기실사지수, 동행지수순환변동치, 국내건설수주액, 회사채(3년), 주택담보 대출 금리 순으로 인과관계가 높은 것으로 나타났다.

다음으로 서울시 아파트 매매실거래가격지수의 경우 행정구역별 아파트매매거래현황(서울), 아파트 매매가격 대비 전세가격비율(서울), 지가변동률(서울), 연간소득 대비 주택가격 비율(서울), 연간소득 대비 대출금액 비율(서울), 수출물가지수, 수입물가지수, 선행지수순환변동치, 국내총투자율, 환율(원/미국달러), 국고채(3년), 주택담보 대출 금리, 소비자물가지수, 생산자물가지수, 국내건설수주액, 고용률, M2, 회사채(3년), KOSPI 30개 거시경제 변수 중 19개 변수와 선행, 후행 관계가 있는 것으로 나타났다. 선행관계에서는 행정구역별 아파트매매거래 건수(서울), 수입물가지수, 국고채(3년), 주택담보대출 금리, 연간소득 대비 주택가격 비율(서울), 수출물가지수, 선행지수수환변동치, 아파트매매가격대비 전세가격비율(서울), 지가변동률(서울), 연간소득 대비 대출금액 비

율(서울), 국내총투자율, 환율(원/미국달러) 순으로 인과관계가 높은 것으로 나타났고, 후행관계에서는 지가변동률(서울), 생산자물가지수, 국고채(3년), 국내총투자율, M2, 회사채(3년), 소비자물가지수, 수입물가지수, 국내건설수주액, 고용률, 환율(원/미국달러), KOSPI 순으로 인과관계가 높은 것으로 나타났다.

서울시 아파트 주택매매가격지수(KB)를 예측하기 위해 시계열 분석 모형의 경우는 그랜저 인과관계 검정을 통해 확인된 선행관계 변수 행정구역별 아파트매매거래 건수(서울), 실업률, 국고채(3년), 회사채(3년), 주택담보대출 금리 총 5개의 변수를 입력변수로 하여 분석하였고, 딥러닝 분석 모형의 경우는 입력 변수 개수에 제한이 없어 30개 거시경제 변수를 모두 입력하여 분석하였다. 2006년 1월부터 2021년 11월까지 191개월의 과거 데이터를 분석하여 2021년 12월부터 2022년 5월까지 6개월을 예측하여 실제값과 예측값을 비교해본 결과 VECM 모형은 3.59395이고, LSTM 모형은 0.27952, GRU 모형은 0.28557으로 나타났다. 딥러닝 분석 모형이 시계열 분석 모형보다 예측력이 더 높은 것으로 나타났고, 딥러닝 분석 모형에서는 LSTM 모형의 예측력이 GRU보다 더 높은 것으로 나타났다.

서울시 아파트 매매실거래가격지수를 예측하기 위해 시계열 분석 모형의 경우는 그랜저 인과관계 검정을 통해 확인된 선행관계 변수 행정구역별 아파트매매거래 건수(서울), 수입물가지수, 주택담보대출 금리, 연간소득 대비 주택가격 비율, 국고채(3년) 총 5개 변수를 입력변수로 하여 분석하였고, 딥러닝 모형의 경우는 30개 거시경제 변수를 모두 입력하여 분석하였다. 앞서 주택매매가격지수(KB)와 동일하게 학습 데이터와 테스트 데이터는 동일하게 분리하였고, 각 모형의 실제값과 예측값을 비교해본 결과 VECM 모형은 3.91130, LSTM 모형은 0.50480, GRU 모형은 0.24876으로 나타났다. 딥러닝 분석 모형이 시계열 분석 모형보다 예측력이 더 높은 것으로 나타났고, 딥러닝 분석 모형에서는 GRU 모형의 예측력이 LSTM 보다 높은 것으로 나타났다.

서울시 아파트 주택매매가격지수(KB)와 매매실거래가격지수의 실제값과 예측값을 그래프로 시각해본 결과 시계열 분석 모형 VECM 모형은 과거 데이터의 추세에 따라 계속해서 상승하는 것으로 예측을 하였고, 딥러닝 분석 모형은 실제값과 비슷한 형태로 예측값이 측정된다는 부분에서 의미 있는 결과를 확인할 수 있었다.

제2절 시사점 및 향후 연구 방향

본 연구는 서울시 아파트 매매 가격을 나타내는 국민은행의 주택매매가격지수(KB)와 한국부동산원의 매매실거래가격지수 2가지를 각각 독립변수로 두고 선행연구에서 확인한 30개의 거시경제변수와 다변량 시계열 데이터 분석을 진행하였고, 분석 결과를 통해 시계열 분석 모형 보다 딥러닝 분석 모형의 예측력이 높다는 것을 확인하였다.

딥러닝 분석 모형의 결과 해석을 하기 위해서 shap의 DeepExplainer을 이용하여 변수의 영향도를 확인하였다. 우선 주택매매가격지수(KB)를 예측한 LSTM과 GRU 모형 모두 30개의 거시경제 변수 중에 M2, 소비자물가지수, 경제활동인구, KOSPI, 수출물가지수 순으로 중요한 것으로 나타났고, 주택담보대출 금리와 수입물가지수는 중요하지 않은 것으로 나타났다. 다음으로 매매실거래가격지수를 예측한 LSTM과 GRU 모형 모두 30개 거시경제 변수 중에서 M2, KOSPI, 경제활동인구 순으로 중요한 것으로 나타났고, 수입물가지수, 연간소득 대비 주택가격 비율(서울)은 중요하지 않은 것으로 나타났다.

향후 연구 방향은 딥러닝 분석 모형에서 주택매매가격지수(KB)와 매매실거래가격지수 2가지에 대해서 Feature Importance 분석 결과에 따라 입력변수를 변경해 보면서 모형의 성능을 비교 분석하는 것이 필요할 것으로 보인다. 추가적으로 초매개변수 설정 시에 연구자에 의해 설계되는 것이 아니라 Grid Search, Random Search 방법 등을 이용하여 최적의 모형의 초매개변수를 찾는 연구가 필요하다.

참 고 문 헌

- 이주미·박성훈·조상호·김주형(2021), “머신러닝을 이용한 부동산 지수 예측 모델 비교”, 「대한건축학회논문집」, 37(1):191-199
- 전해정(2020), “시계열분석모형과 머신러닝을 이용한 주택가격 예측력 연구”, 「한국주거환경학회논문집」, 18. 1 (2020): 49-65.
- 윤만식·신성윤(2020), “계량시계열과 인공신경망모형을 이용한 아파트매매가격 지수 예측 - 서울과 6대 광역도시를 중심으로 -”, 「지역사회발전학회 논문집」, 45권2호: 9-19.
- 배성완·유정석(2018), “머신 러닝 방법과 시계열 분석 모형을 이용한 부동산 가격지수 예측”, 「주택연구」, 26. 1 (2018): 107-133.
- 최문기·이성화(2021), “거시경제변수 및 부동산 대책이 강남4구 아파트 시장에 미치는 영향”, 「한국지적학회지」, 37. 3 (2021): 105-114.
- 조한국·김영곤(2020), “아파트 가격 및 거래량과 거시경제변수 간의 동태적 관계연구”, 「부동산학보」, 82. (2020): 45-58.
- 조한국·서충원(2019), “VAR에 의한 임대료 및 공실률과 거시경제변수간의 동태적 연구”, 「부동산학보」, 79. (2019): 35-49.
- 김경민(2018), “서울아파트시장과 거시경제변수 요인들간 동학적 상관관계 분석”, 「부동산학보」, 73. (2018): 115-129.
- 박성균·이현식(2012), “주거용과 상업용 부동산의 가격 결정요인 비교”, 「부동산연구」, 2012 22(2):171
- S. Hochreiter, J. Schmidhuber, J.(1997), “Long Short-Term Memory” Neural Computation, vol. 9, no. 8, pp. 1735-1780
- Cho, Kyunghyu(2014), “Learning Phrase Representations using RNN Encoder - Decoder for Statistical Machine Translation”, Proceedings of EMNLP: 1724 - 1734.
- 유원준·안상준, 「딥 러닝을 이용한 자연어 처리 입문」 Wikidocs
한국은행 ECOS <https://ecos.bok.or.kr/>
통계청 KOSIS <https://kosis.kr/>
한국부동산원 R-ONE <https://www.reb.or.kr/r-one/main.do>

KB부동산 데이터허브 <https://data.kbland.kr/>

주택금융공사 주택금융통계시스템 <https://houstat.hf.go.kr/>

KDI 경제정보센터 <https://www.kdi.re.kr/>

국토교통부 <https://www.molit.go.kr/portal.do>



부 록

[부록1] 주택매매가격지수(KB) - LSTM 결과

Hidden Layer	Sequence	Epoch	Dropout	Batch Size	RMSE
2	3	600	0	5	2.099953454
2	3	600	0	10	0.605089357
2	2	600	0	10	1.069347717
2	1	600	0	10	1.520702202
2	4	600	0	10	2.144307947
2	5	600	0	10	5.39541449
2	3	600	0	15	0.562694429
2	3	600	0	20	0.949501545
2	3	600	0	25	2.194203256
2	3	600	0	30	1.049746252
2	3	600	0	35	1.541341514
2	3	600	0	40	1.774223473
2	3	600	0	45	2.413016902
2	3	700	0	5	5.13115708
2	3	700	0	10	0.347190775
2	1	700	0	10	1.67487805
2	2	700	0	10	2.184970465
2	4	700	0	10	2.41686589
2	5	700	0	10	3.415426392
2	3	700	0	15	0.667573875
2	3	700	0	20	0.562713071
2	3	700	0	25	1.260724163
2	3	700	0	30	0.676142092
2	3	700	0	35	1.459742933
2	3	700	0	40	3.572098472
2	3	700	0	45	2.213409112
2	1	800	0	5	1.150802908
2	1	800	0	10	1.55693576
2	3	800	0	10	1.877998665
2	2	800	0	10	2.373231997
2	4	800	0	10	3.03358994
2	5	800	0	10	5.758108276

2	1	800	0	15	2.18810694
2	1	800	0	20	1.903244815
2	1	800	0	25	1.584088714
2	1	800	0	30	2.390097283
2	1	800	0	35	2.279866302
2	1	800	0	40	1.989431689
2	1	800	0	45	2.472694248
3	2	600	0	5	2.810048692
3	2	600	0	10	0.426291412
3	4	600	0	10	1.592399178
3	3	600	0	10	1.995055622
3	1	600	0	10	2.850993258
3	5	600	0	10	3.948995728
3	2	600	0	15	1.825283521
3	2	600	0	20	2.995522748
3	2	600	0	25	2.622871094
3	2	600	0	30	3.448491775
3	2	600	0	35	1.676735406
3	2	600	0	40	2.692748931
3	2	600	0	45	3.59945197
3	2	700	0	5	2.15649084
3	2	700	0	10	0.500964235
3	1	700	0	10	1.900607752
3	3	700	0	10	2.361612717
3	4	700	0	10	3.045348635
3	5	700	0	10	3.360212463
3	2	700	0	15	1.662114457
3	2	700	0	20	2.643115386
3	2	700	0	25	2.981282429
3	2	700	0	30	3.192864317
3	2	700	0	35	4.260400638
3	2	700	0	40	3.860897637
3	2	700	0	45	4.096081723
3	2	800	0	5	1.716111346
3	2	800	0	10	0.564388356
3	1	800	0	10	0.759305173
3	3	800	0	10	2.235858099
3	4	800	0	10	3.274899763
3	5	800	0	10	3.717466492

3	2	800	0	15	5.407066231
3	2	800	0	20	4.234450633
3	2	800	0	25	1.635829774
3	2	800	0	30	3.01331847
3	2	800	0	35	2.867735975
3	2	800	0	40	3.774690734
3	2	800	0	45	2.991146287
4	1	600	0	5	2.46960817
4	1	600	0	10	0.428712274
4	2	600	0	10	1.093277709
4	3	600	0	10	3.019272065
4	4	600	0	10	3.062491263
4	5	600	0	10	4.485777039
4	1	600	0	15	1.905672994
4	1	600	0	20	1.281498493
4	1	600	0	25	1.213557215
4	1	600	0	30	2.224268633
4	1	600	0	35	2.518257073
4	1	600	0	40	1.880497335
4	1	600	0	45	1.293872027
4	2	700	0	5	2.22065329
4	2	700	0	10	1.320313084
4	1	700	0	10	1.71187309
4	4	700	0	10	2.765580779
4	3	700	0	10	3.086609893
4	5	700	0	10	3.832319397
4	2	700	0	15	1.435476592
4	2	700	0	20	0.885338282
4	2	700	0	25	1.055579873
4	2	700	0	30	1.871490559
4	2	700	0	35	2.85577709
4	2	700	0	40	3.066463798
4	2	700	0	45	3.395566321
4	2	800	0	5	1.574778173
4	2	800	0	10	0.498317928
4	1	800	0	10	0.585564273
4	3	800	0	10	2.2288383
4	4	800	0	10	3.329711112
4	5	800	0	10	4.332723755

4	2	800	0	15	0.656886381
4	2	800	0	20	1.970405013
4	2	800	0	25	1.873187638
4	2	800	0	30	0.514940307
4	2	800	0	35	1.412938401
4	2	800	0	40	2.492842011
4	2	800	0	45	1.133811723

[부록2] 주택매매가격지수(KB) - GRU 결과

Hidden Layer	Sequence	Epoch	Dropout	Batch Size	RMSE
2	1	600	0	5	0.966975
2	1	600	0	10	0.78327
2	2	600	0	10	1.725172
2	4	600	0	10	3.040002
2	3	600	0	10	3.229194
2	5	600	0	10	5.163435
2	1	600	0	15	0.748935
2	1	600	0	20	0.832639
2	1	600	0	25	1.246478
2	1	600	0	30	0.796199
2	1	600	0	35	1.20446
2	1	600	0	40	1.098569
2	1	600	0	45	1.384657
2	2	700	0	5	0.940749
2	2	700	0	10	0.539232
2	1	700	0	10	1.593772
2	3	700	0	10	3.020512
2	4	700	0	10	5.757561
2	5	700	0	10	5.778219
2	2	700	0	15	0.576159
2	2	700	0	20	0.949246
2	2	700	0	25	0.504568
2	2	700	0	30	0.540449
2	2	700	0	35	0.654571
2	2	700	0	40	1.005985
2	2	700	0	45	0.38854
2	1	800	0	5	0.789917

2	1	800	0	10	0.839033
2	2	800	0	10	1.477371
2	3	800	0	10	2.979277
2	4	800	0	10	3.974365
2	5	800	0	10	5.061072
2	1	800	0	15	1.544411
2	1	800	0	20	0.604186
2	1	800	0	25	1.116579
2	1	800	0	30	1.301612
2	1	800	0	35	1.542548
2	1	800	0	40	1.774464
2	1	800	0	45	1.307225
3	2	600	0	5	1.641853
3	2	600	0	10	0.803773
3	5	600	0	10	2.291136
3	3	600	0	10	2.485947
3	4	600	0	10	2.779097
3	1	600	0	10	2.884664
3	2	600	0	15	0.427315
3	2	600	0	20	3.266434
3	2	600	0	25	0.799121
3	2	600	0	30	1.251586
3	2	600	0	35	1.703255
3	2	600	0	40	0.839811
3	2	600	0	45	0.981425
3	2	700	0	5	0.680199
3	2	700	0	10	0.398778
3	1	700	0	10	1.406132
3	3	700	0	10	2.555417
3	4	700	0	10	4.155999
3	5	700	0	10	4.523482
3	2	700	0	15	0.669696
3	2	700	0	20	0.714089
3	2	700	0	25	0.538868
3	2	700	0	30	0.690255
3	2	700	0	35	0.677177
3	2	700	0	40	0.416303
3	2	700	0	45	0.552832
3	5	800	0	5	1.612078

3	2	800	0	10	0.944153
3	1	800	0	10	1.354042
3	3	800	0	10	2.992816
3	4	800	0	10	3.688609
3	5	800	0	10	4.006842
3	2	800	0	15	1.45558
3	2	800	0	20	1.858207
3	2	800	0	25	1.672829
3	2	800	0	30	0.583159
3	2	800	0	35	0.882171
3	2	800	0	40	1.267085
3	2	800	0	45	0.539561
4	1	600	0	5	2.483502
4	1	600	0	10	0.994771
4	2	600	0	10	2.357053
4	3	600	0	10	2.74064
4	4	600	0	10	3.555879
4	5	600	0	10	5.842611
4	1	600	0	15	0.971414
4	1	600	0	20	0.617831
4	1	600	0	25	1.352547
4	1	600	0	30	0.776919
4	1	600	0	35	0.88007
4	1	600	0	40	0.892409
4	1	600	0	45	0.91181
4	1	700	0	5	1.401836
4	1	700	0	10	0.955622
4	2	700	0	10	1.458179
4	3	700	0	10	2.567735
4	4	700	0	10	4.221617
4	5	700	0	10	5.317244
4	1	700	0	15	1.255961
4	1	700	0	20	0.752554
4	1	700	0	25	2.233779
4	1	700	0	30	0.787666
4	1	700	0	35	0.674504
4	1	700	0	40	0.774913
4	1	700	0	45	1.551595
4	1	800	0	5	3.262634

4	1	800	0	10	0.783176
4	2	800	0	10	1.329622
4	3	800	0	10	2.867257
4	4	800	0	10	3.710867
4	5	800	0	10	4.775015
4	1	800	0	15	1.085881
4	1	800	0	20	1.275733
4	1	800	0	25	0.956285
4	1	800	0	30	1.469572
4	1	800	0	35	0.625856
4	1	800	0	40	1.102833
4	1	800	0	45	0.675934

[부록3] 매매실거래가격지수 - LSTM 결과

Hidden Layer	Sequence	Epoch	Dropout	Batch Size	RMSE
2	1	600	0	5	9.449654
2	1	600	0	10	1.354432
2	2	600	0	10	1.535316
2	3	600	0	10	3.939044
2	4	600	0	10	4.855981
2	5	600	0	10	7.286031
2	1	600	0	15	5.013782
2	1	600	0	20	2.751462
2	1	600	0	25	1.943058
2	1	600	0	30	3.37207
2	1	600	0	35	3.212579
2	1	600	0	40	2.879253
2	1	600	0	45	3.768267
2	2	700	0	5	0.898803
2	2	700	0	10	0.778394
2	4	700	0	10	4.151582
2	3	700	0	10	4.185
2	1	700	0	10	4.64077
2	5	700	0	10	7.040609
2	2	700	0	15	2.786355
2	2	700	0	20	2.104001
2	2	700	0	25	3.1052

2	2	700	0	30	3.000807
2	2	700	0	35	3.495674
2	2	700	0	40	5.111414
2	2	700	0	45	3.02935
2	2	800	0	5	0.823603
2	2	800	0	10	1.74653
2	3	800	0	10	3.192887
2	1	800	0	10	4.946786
2	4	800	0	10	6.113754
2	5	800	0	10	6.806386
2	2	800	0	15	3.188205
2	2	800	0	20	2.908192
2	2	800	0	25	2.715657
2	2	800	0	30	3.998252
2	2	800	0	35	3.421076
2	2	800	0	40	4.406087
2	2	800	0	45	4.671416
3	2	600	0	5	1.075511
3	2	600	0	10	1.07287
3	1	600	0	10	3.502541
3	3	600	0	10	4.401801
3	4	600	0	10	4.532157
3	5	600	0	10	7.1238
3	2	600	0	15	0.747306
3	2	600	0	20	0.504805
3	2	600	0	25	1.315336
3	2	600	0	30	0.58556
3	2	600	0	35	1.305398
3	2	600	0	40	2.789303
3	2	600	0	45	4.877827
3	2	700	0	5	1.154281
3	2	700	0	10	0.925402
3	1	700	0	10	2.601295
3	3	700	0	10	2.972569
3	4	700	0	10	4.480929
3	5	700	0	10	5.258092
3	2	700	0	15	1.489851
3	2	700	0	20	1.267149
3	2	700	0	25	3.757634

3	2	700	0	30	3.959918
3	2	700	0	35	3.074986
3	2	700	0	40	5.007578
3	2	700	0	45	3.106504
3	2	800	0	5	1.060129
3	2	800	0	10	1.178387
3	3	800	0	10	3.070427
3	1	800	0	10	3.679525
3	4	800	0	10	4.931722
3	5	800	0	10	6.815504
3	2	800	0	15	1.618644
3	2	800	0	20	1.877202
3	2	800	0	25	2.126326
3	2	800	0	30	1.604648
3	2	800	0	35	2.371642
3	2	800	0	40	4.050932
3	2	800	0	45	5.733763
4	2	600	0	5	1.148882
4	2	600	0	10	0.618789
4	3	600	0	10	0.654958
4	4	600	0	10	1.096492
4	1	600	0	10	3.053111
4	5	600	0	10	5.967542
4	2	600	0	15	1.184651
4	2	600	0	20	1.328625
4	2	600	0	25	0.56907
4	2	600	0	30	0.892562
4	2	600	0	35	2.733817
4	2	600	0	40	3.97358
4	2	600	0	45	5.267711
4	2	700	0	5	0.815759
4	2	700	0	10	0.929163
4	3	700	0	10	1.03848
4	1	700	0	10	4.810006
4	4	700	0	10	5.141162
4	5	700	0	10	5.220983
4	2	700	0	15	1.236285
4	2	700	0	20	1.844644
4	2	700	0	25	2.787263

4	2	700	0	30	2.906424
4	2	700	0	35	3.099674
4	2	700	0	40	4.955558
4	2	700	0	45	2.998333
4	3	800	0	5	3.423464
4	3	800	0	10	0.877649
4	2	800	0	10	1.301388
4	1	800	0	10	2.797185
4	5	800	0	10	4.571233
4	4	800	0	10	5.274375
4	3	800	0	15	1.109902
4	3	800	0	20	2.34673
4	3	800	0	25	0.878225
4	3	800	0	30	1.571032
4	3	800	0	35	0.7234
4	3	800	0	40	2.387228
4	3	800	0	45	3.058973

[부록4] 매매실거래가격지수 - GRU 결과

Hidden Layer	Sequence	Epoch	Dropout	Batch Size	RMSE
2	2	600	0	5	3.802305
2	2	600	0	10	1.119905
2	1	600	0	10	3.428442
2	5	600	0	10	3.924894
2	3	600	0	10	5.211519
2	4	600	0	10	5.879236
2	2	600	0	15	1.044114
2	2	600	0	20	4.169692
2	2	600	0	25	0.763339
2	2	600	0	30	0.738504
2	2	600	0	35	0.715697
2	2	600	0	40	1.613505
2	2	600	0	45	1.736756
2	2	700	0	5	1.109332
2	2	700	0	10	0.779784
2	3	700	0	10	3.07702

2	1	700	0	10	5.145768
2	5	700	0	10	6.850545
2	4	700	0	10	7.432313
2	2	700	0	15	1.017569
2	2	700	0	20	0.923089
2	2	700	0	25	1.604872
2	2	700	0	30	1.330209
2	2	700	0	35	0.673679
2	2	700	0	40	1.137847
2	2	700	0	45	1.136746
2	2	800	0	5	1.102351
2	2	800	0	10	1.709942
2	3	800	0	10	2.979723
2	1	800	0	10	4.968595
2	5	800	0	10	6.771032
2	4	800	0	10	7.542444
2	2	800	0	15	1.862194
2	2	800	0	20	2.216148
2	2	800	0	25	1.704151
2	2	800	0	30	1.146237
2	2	800	0	35	0.476337
2	2	800	0	40	0.844937
2	2	800	0	45	2.858608
3	2	600	0	5	0.830708
3	2	600	0	10	0.932244
3	1	600	0	10	2.45189
3	5	600	0	10	5.143514
3	4	600	0	10	5.222104
3	3	600	0	10	5.348073
3	2	600	0	15	0.707547
3	2	600	0	20	1.407697
3	2	600	0	25	3.226049
3	2	600	0	30	1.687788
3	2	600	0	35	1.825984
3	2	600	0	40	0.757585
3	2	600	0	45	2.895945
3	2	700	0	5	0.796261
3	2	700	0	10	1.336426
3	1	700	0	10	2.49561

3	3	700	0	10	3.819803
3	5	700	0	10	4.474622
3	4	700	0	10	5.033262
3	2	700	0	15	0.775029
3	2	700	0	20	0.989959
3	2	700	0	25	2.3207
3	2	700	0	30	1.079544
3	2	700	0	35	0.852667
3	2	700	0	40	1.74742
3	2	700	0	45	0.764232
3	2	800	0	5	0.894744
3	2	800	0	10	1.158018
3	1	800	0	10	2.573992
3	5	800	0	10	4.345563
3	3	800	0	10	5.194518
3	4	800	0	10	5.441875
3	2	800	0	15	1.210013
3	2	800	0	20	1.020219
3	2	800	0	25	1.066187
3	2	800	0	30	1.117108
3	2	800	0	35	0.63455
3	2	800	0	40	2.980056
3	2	800	0	45	1.512589
4	1	600	0	5	2.354861
4	1	600	0	10	0.617087
4	2	600	0	10	0.79859
4	4	600	0	10	5.055363
4	5	600	0	10	5.082189
4	3	600	0	10	6.493193
4	1	600	0	15	1.678312
4	1	600	0	20	1.630096
4	1	600	0	25	1.946059
4	1	600	0	30	0.842565
4	1	600	0	35	1.992106
4	1	600	0	40	2.038637
4	1	600	0	45	1.851808
4	2	700	0	5	0.759595
4	2	700	0	10	0.70747
4	3	700	0	10	3.320844

4	5	700	0	10	3.641645
4	4	700	0	10	4.314845
4	1	700	0	10	5.089179
4	2	700	0	15	0.772995
4	2	700	0	20	0.842242
4	2	700	0	25	0.846728
4	2	700	0	30	0.870018
4	2	700	0	35	0.752311
4	2	700	0	40	1.043462
4	2	700	0	45	0.945198
4	2	800	0	5	0.772657
4	2	800	0	10	1.017832
4	1	800	0	10	2.655423
4	3	800	0	10	2.885238
4	5	800	0	10	3.530614
4	4	800	0	10	4.708236
4	2	800	0	15	0.923155
4	2	800	0	20	1.144052
4	2	800	0	25	0.957595
4	2	800	0	30	1.488976
4	2	800	0	35	1.744555
4	2	800	0	40	0.633422
4	2	800	0	45	1.506707

ABSTRACT

The purpose of this study is to analyze the housing sales price index (KB) of KB Kookmin Bank & actual transaction price index of apartment houses by the Korea Real Estate Agency, 30 macroeconomic variables and multivariate data for apartments in Seoul. And also, the prediction capability of the model is compared and analyzed by designing a time series analysis model and a deep learning analysis model. The housing sales price index (KB) used in the analysis can be checked the price fluctuation rate of all housing types regardless of whether they are transacted or not, and the price change rate of the house actually traded can be checked by the actual transaction price index.

Since actual transaction price index tends to fluctuate more than the index of housing sales price(KB), in order to grasp the overall market, two indices needed to be examined and the prediction capability is compared and analyzed together

In order to design a time series analysis model, causal relationship between variables was identified through granger causality test among the 30 macroeconomic variables, housing sales price index and actual transaction price index. And also, variables with precedence were selected and the prediction model was designed with the VECM(Vector Error Correction Model) model by the cointegration test. In order to design a deep learning analysis model, 30 macroeconomic variables were used as input variables, and the prediction model was designed with the LSTM(Long Short-Term Memory) model and the GRU(Gated Recurrent Unit) model. In order to design the optimal deep learning model, the performance of the model was measured by repeatedly changing the hyper

parameter.

Performance evaluation of the prediction capability of the time series analysis model and the deep learning analysis model was compared with the RMSE(Root Mean Square of Errors) evaluation index.

A time series analysis model and a deep learning analysis model were designed to study 191 months from January 2006 to November 2021 and predict 6 months from December 2021 to May 2022 and both the housing sales price index (KB) and the actual transaction price index showed that the deep learning analysis model was found to have higher prediction capability than the time series analysis model. However, in the case of the time series analysis model, because it can be analyzed statistically, such as unit root test, the granger causality test, and the cointegration test, It seems that using the time series analysis model and the deep learning analysis model together can objectively check the relationship.

연구 윤리 서약서

본인은 한양대학교 대학원생으로서 이 학위논문 작성 과정에서 다음과 같이 연구 윤리의 기본 원칙을 준수하였음을 서약합니다.

첫째, 지도교수의 지도를 받아 정직하고 엄정한 연구를 수행하여 학위논문을 작성한다.

둘째, 논문 작성시 위조, 변조, 표절 등 학문적 진실성을 훼손하는 어떤 연구 부정행위도 하지 않는다.

셋째, 논문 작성시 논문유사도 검증시스템 "카피킬러" 등을 거쳐야 한다.

2022년12월13일

학위명 : 석사

학과 : 도시·부동산개발학과

지도교수 : 성현곤

성명 : 조민지



한 양 대 학 교 부 동 산 융 합 대 학 원 장 귀 하

Declaration of Ethical Conduct in Research

I, as a graduate student of Hanyang University, hereby declare that I have abided by the following Code of Research Ethics while writing this dissertation thesis, during my degree program.

"First, I have strived to be honest in my conduct, to produce valid and reliable research conforming with the guidance of my thesis supervisor, and I affirm that my thesis contains honest, fair and reasonable conclusions based on my own careful research under the guidance of my thesis supervisor.

Second, I have not committed any acts that may discredit or damage the credibility of my research. These include, but are not limited to : falsification, distortion of research findings or plagiarism.

Third, I need to go through with Copykiller Program(Internet-based Plagiarism-prevention service) before submitting a thesis."

DECEMBER 13, 2022

Degree : Master

Department : DEPARTMENT OF URBAN & REAL ESTATE DEVELOPMENT

Thesis Supervisor : Hyungun Sung

Name : CHO MIN JI

