

시장 내 자금 흐름 변수를 활용한 변동성 지표 예측

김겨레 김민호

초록

VKOSPI는 KOSPI200를 기초로 하는 옵션의 가격을 통해 산출되는 시장 변동성 지표이다. VKOSPI는 시장 참여자의 기대를 반영하는 정보이기 때문에 이를 활용하여 시장 변동성의 증감을 예측하여 시장 상황에 선제적으로 대응할 수 있다. 본 연구는 기존 변동성 예측 연구에서 주로 활용하던 Heterogeneous Autoregressive(HAR) 모형을 벤치마크 모형으로 두고, 머신러닝 알고리즘의 변동성 예측력을 평가하였다. 본 연구에서는 Long Term Short Memory (LSTM)를 주된 머신러닝 알고리즘으로 활용했다. 해당 알고리즘은 설명변수가 들어갈 수 있는 레이어를 여러 계층으로 나눌 수 있다. 이런 점을 활용하여 성격이 다른 설명변수를 서로 다른 레이어에 넣어 기존의 경제 예측 머신러닝 알고리즘과 차별화된 방법론을 사용하였다. 또한 본 연구에서는 시장의 변동성에 영향을 줄 것으로 보이지만, 기존에는 활용하지 않았던 반대매매, 미수금 등 설명변수를 추가로 도입하여 예측을 시도하였다. Boruta 알고리즘을 통해 중요한 변수를 추출한 결과, 미수금, 차익거래 등 본 연구에서 새롭게 도입한 변수가 높은 순위 나타났다. 예측 결과, 1주, 2주, 4주의 단기 예측에서는 기존의 계량경제학 모형이 머신러닝 알고리즘에 비해 우월한 것으로 나타나지만, 8주에서는 본 연구에서 도입한 LSTM이 우월한 성과를 보여주는 것으로 나타났다. 본 연구의 결과는, 비록 단기 예측에서 본 연구에서 도입한 알고리즘이 기존 모형보다 부족한 예측 성과를 보였지만 중장기 예측에서는 해당 알고리즘이 유용하게 사용될 수 있음을 시사한다.

I. 서론

VKOSPI(KOSPI Volatility)는 주식시장의 한달 뒤 변동성에 대한 옵션시장 참여자의 기대를 보여주는 지표이다. VKOSPI는 미국 CBOE Volatility Index(VIX)와 마찬가지로 옵션가격 가격결정 모형에 따라 산출된다. 옵션가격 결정모형에서 옵션의 이론가격은 간단하게 아래와 같은 수식으로 표현할 수 있다.

$$OptPrice = f(\text{strike price}, \text{Underlying Asset Price}, \text{Left maturity}, \text{Interest rate}, \text{Volatility})$$

옵션가격 결정모형에 따르면 옵션가격(OptPrice)은 거래 당사자 양방이 동의하여 만기에 거래되는 기초 자산의 가격인 행사가격(Strike Price), 기초 자산의 현재 가격(Underlying Asset Price), 남은 만기 (Left Maturity), 이자율 (Interest rate), 변동성

(Volatility)의 함수이다. 변동성을 제외하면 나머지 요소는 지표로 확인이 가능하다. 따라서 옵션가격 결정모형에서 변동성을 제외한 나머지 변수를 대입하면, 변동성을 산출할 수 있다. VKOSPI는 모형 내에서 계산하여 도출하는 지표이기 때문에 내재변동성이라고 한다. VKOSPI는 수익률 변동성 등 다른 변동성 지표와 달리 시장 내 변동성에 대한 기대를 반영하고, 옵션가격과 모형을 통해 산출되는 지표이기 때문에 시장의 변동성을 예측하고 대응하는 데 유용하다. 특히, 주식시장의 변동성이 커질 것으로 예상하면, 시장 참여자들은 옵션을 통해 주식 포지션을 헷지하려는 움직임을 보인다. 이에 따라 옵션시장이 선제적으로 움직인다. 따라서 VKOSPI에 대한 예측을 통해 급격한 시장 변동에 선제적으로 대응할 수 있다.

주식수익률 변동성을 예측하는 연구는 이승희와 한희준(2016), Mingyu Liu(2020), 정희수(2012) 등에 의해 활발히 이루어졌지만, VKOSPI 예측 연구는 부족하다. 또한 변동성 예측에 대한 연구는 주로 이승희와 한희준(2016), 김선웅(2010)와 같이 전통적인 계량경제학 모델을 중심으로 이루어졌다. 하지만 머신러닝 알고리즘을 활용한 변동성 예측 연구도 부족하다. 본 연구에서는 기존 연구에서 사용한 변수를 참고하고 Heterogeneous Autoregressive (HAR) 모형 등 계량경제학 모형을 벤치마크 모형으로 둔 뒤 머신러닝 알고리즘을 활용한 예측과 성능을 비교하는 방식으로 연구를 진행하였다.

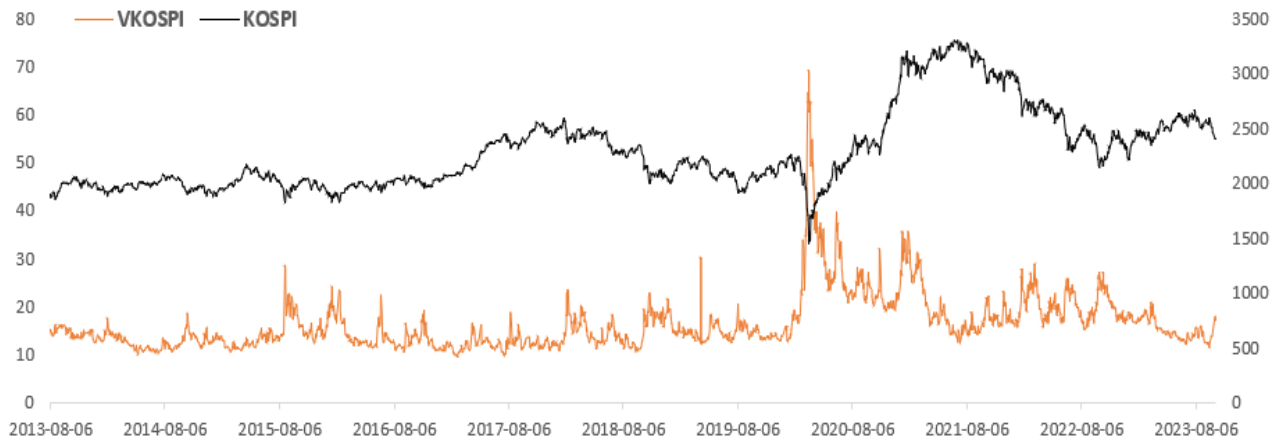
본 연구는 다음과 같이 구성된다. 2장에서는 VKOSPI 예측을 위해 사용된 변수를 주식시장 매매동향, 주식시장 자금흐름 그리고 주식시장 외부변수로 구분하여 설명하며, 이들 변수들의 정상성을 판단하고 정상시계열로 변환하는 과정을 설명한다. 3장에서는 본 연구에서 보이고자 하는 주된 예측방법인 Multi-Layer LSTM에 설명하고, HAR 및 보루타 알고리즘을 통한 변수선택을 적용한 랜덤포레스트(이하 RF-Boruta), XGBoost등의 비교모형을 간략히 소개한다. 4장에서는 보루타 알고리즘을 통한 변수 선택 결과를 소개하며, 모형간 예측 결과를 절대평균오차를 중심으로 비교하고 Giacomini-White 검정 및 MCS 검정 결과를 통해 연구에서 사용한 모형을 평가한다. 5장에서는 연구의 결론을 제시한다.

II. 데이터

II-1. 종속변수

국제 주요 금융지표를 제공하고 있는 웹 사이트인 Investing.com에서는 VKOSPI 데이터를 제공하고 있으며, 2013년 8월 6일부터 자료를 제공하고 있다. 본 연구에서는 해당 웹 사이트를 통해 VKOSPI의 2013년 8월 6일부터 2023년 10월 10일까지의 데이터를 획득한 후, 자연로그를 취해 종속변수로 사용하였다.

< Figure 1 > VKOSPI(Left) and KOSPI(Right) 2013.08.06 – 2023.10.10



< Table 1 > Descriptive Statistics for Logarithm of VKOSPI

Mean	2.762
Median	2.694
Standard Deviation	0.276
Minimum	2.274
Maximum	4.140
Skewness	1.219
Kurtosis	5.078
ADF (p-value)	0.656
PP (p-value)	0.000
KPSS (test statistics)	3.999

II-2. 설명변수

설명변수는 총 52개 변수를 사용하였으며, 주가 변동성 및 수익률을 분석한 기존 연구에 사용된 바 있는 36개 변수와 본 연구에서 신규로 도입한 15개 변수, 그리고 VKOSPI의 Lag을 사용하였다. 기존 연구의 36개 변수에 대해서는 개인/외국인/기관 순매수금액 및 국고채금리 등 금리변수가 다수를 차지한다. Han et al.(2015)는 미국 주가 수익률과 변동성이 VKOSPI에 미치는 영향을 분석하면서 KOSPI 200 Spot index, VIX, S&P500, 환율, 이자율 등을 사용하였으며, Chun et al.(2020)은 주식시장의 변동성을 연구하며 환율 등 변수 외에도 금 가격 및 유가를 활용하였다. 이승희와 한희준(2016)은 한국 주식시장의 변동성 분석 및 예측 연구에서 경기동행지수, 경기선행지수, VIX, 주택매매가격지수, 콜금리, 국고채금리 등을 사용한 바 있으며, 유한수(2006), 류형선

외(2013)에서는 개인, 외국인 및 기관투자자의 순매수금액을 활용한 바 있다. 본 연구에서는 기존에 활용한 변수에 더해, 차익거래 및 비차익거래의 매수/매도/순매수 금액, 신용잔고, 주식형펀드, 신용비율, 고객예탁금, 장내파생상품 예수금, 대고객 RP 매도잔고, 위탁매매 미수금, 미수금 대비 반대매매 금액 및 비중 등 15개 변수를 새롭게 도입하였다. 새롭게 도입한 변수들은 금융투자협회의 종합통계 및 네이버 증권 웹스크래핑을 통해 수집하였다.

52개의 설명변수는 총 3개의 그룹으로 다시 분류되었다. 주식시장 매매동향에 12개 변수, 주식시장 자금흐름에 9개 변수, 그리고 주식시장 외부변수로 31개 변수를 분류하였다. 대표적으로, 주식시장 매매동향에는 KOSPI 지수와 투자주체별 순매수금액이 포함되며 주식시장 자금흐름에는 신용잔고, 미수금 등이 포함된다. 주식시장 외부변수에는 각종 금리 및 미국 주가지수, VIX 및 환율 등이 포함된다. 이에 더해, 모든 설명변수의 종합적인 특성이 최종적으로 종속변수에 반영되는 것으로 보아 설명변수의 Target Factor 4개를 산출하여 설명변수로 활용하였고, VKOSPI의 Lag은 설명변수의 주식시장 매매동향 그룹에 포함하였다.

< Table 2 > 변수 그룹화 및 신규변수 구분

주식시장 매매 동향	주식시장 자금 흐름	주식시장 외부 변수
KOSPI Index, KOSDAQ Index, 개인-외국인-기관 순매수, 차익거래-비차익거래 매수 매도 순매수 추가 : VKOSPI Lag	신용잔고, 주식형펀드, 신용비율, 고객예탁금, 장내파생상품 예수금, 대고객 RP 매도잔고, 위탁매매 미수금, 반대매매 금액, 미수금 대비 반대매매 비중	콜금리 1일, KORIBOR 3M 6M 12M, CD91, 국고채 금리 1Y 3Y 5Y 10Y 20Y 30Y, 통화안정증권 91D 1Y 2Y, 회사채 AA A- BBB, MMF 7D, KOSPI 200 선물, US T-bill 3M 5Y 10Y 30Y, Gold Price, Dollar Index, KRW/USD, IXIC, DJI, S&P 500, VIX, WTI

* 본 연구에서 신규로 도입한 변수 굵은 글씨

본 연구는 주단위 예측을 목표로 하므로, 획득한 데이터 또한 일 단위 데이터를 주 단위 데이터로 변환하여 사용하였다. 개인, 외국인, 기관, 차익거래 및 비차익거래 매수/매도와 같은 변수는 단순 가격 지표가 아닌 거래 규모를 나타내는 변수이므로 월요일부터 금요일까지의 데이터를 누적하여 사용하였으며, 지표 및 가격, 이자율 등의 변수는 해당 주치의 중간값에 해당하는 수요일의 데이터를 사용하였다. 다만, 수요일이 공휴일이어서 결측되었을 경우 해당 주의 첫번째 날을 사용하였다. 결과적으로, 2013년 8월 6일부터 2023년 10월 10일 기간, 총 531주에 해당하는 데이터를 연구에 활용하였다.

II-3. 단위근 검정 및 정상시계열로의 변환

본 연구에서는 시계열 변수를 정상시계열로 변환하여 예측을 진행하였다. 정상시계열로 변환하기 위해서는 변수의 단위근 여부를 판단해야 한다. 따라서 시계열 변수의 단위근 여부를 판단하기 위해 널리 사용되고 있는 Augmented Dickey-Fuller 검정(ADF), Phillips-Perron 검정(PP), KPSS 검정을 실시하여 단위근 여부를 판단하였다. 모든 검정에서 일관된 결과가 나오는 변수는 그에 따른 자료 변환을 실시하였으며, 검정 결과가 서로 충돌할 경우 기존 연구의 변수 사용 및 변수 변환 방식을 추가로 검토하여 변수 변환 여부를 결정하였다.

검정 결과, 단위근을 가지고 있는 변수에 대해서는 FRED-MD의 데이터 변환 방법 및 한희준(2023)을 참고하여 정상시계열로 변환하였다. 지수 및 가격 변수에 대해서는 로그-차분을 실시하였으며, 이자율 변수에 대해서는 로그를 취하지 않고 차분을 실시하여 정상시계열로 변환하였다. 변수의 T-code는 부록에 수록하였다.

III. 연구 방법

III-1. Multi-Layer LSTM

본 연구에서는 서로 다른 성격의 변수를 활용하여 변동성 지표 예측을 시도한다. 기존의 변동성 예측 연구는 변수의 성격과 정보의 반영 경로와 관계 없이 설명변수를 사용하였다. 머신러닝 알고리즘을 활용한 Kim and Han(2022)에서도 보루타 알고리즘을 통해 변수를 선별하여 예측하는 새로운 방법을 사용했지만, 변수의 성격을 고려한 연구는 진행하지 않았다. 본 연구에서는 Long Short Term Memory (LSTM)을 활용하여 다른 성격의 변수를 다른 경로로 머신러닝 알고리즘에 반영하는 새로운 방법론을 도입하여 예측 연구를 진행한다.

LSTM은 Recurrent Neural Network(RNN)을 개선한 머신러닝 알고리즘이다. RNN은 시계열 예측을 할 때, 과거 정보가 희미하게 반영된다. 따라서 중요한 정보임에도 과거의 정보이기에 가중치가 낮게 반영되는 문제가 발생할 수 있다. LSTM은 이런 문제를 개선하여 내재된 함수와 Cell state, Hidden state 등을 통해 과거 정보라도 가중치가 높게 반영될 수 있다. 이런 특성으로 인해 LSTM은 시계열 예측에 주로 활용된다. 변동성 예측에 LSTM을 활용한 사례도 있다. Liu et al.(2020)는 Generalized Autoregressive Conditional Heteroskedasticity (GARCH) 모형과 LSTM을 결합하여 변동성 예측을 시도하였다.

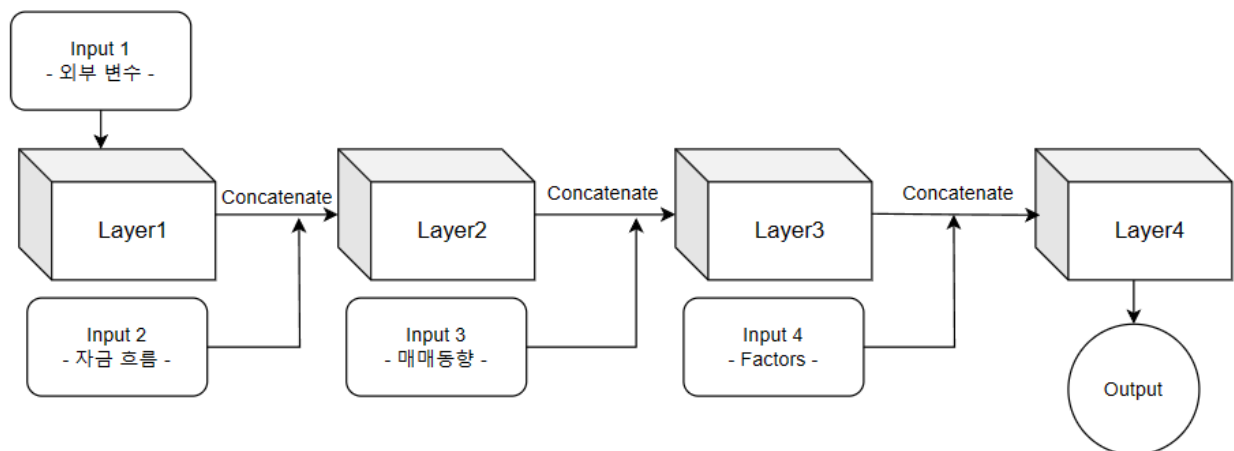
LSTM을 통해 레이어의 수와 구조를 변형하여 변수의 특성을 활용한 알고리즘을 설계할 수 있다. 하지만 변동성 예측에서 변수의 서로 다른 특성을 다른 경로로 반영하여

예측을 시도한 연구는 부족하다. Li et al.(2023)은 종속변수와 양의 상관관계, 음의 상관관계를 가진 설명변수를 구분하여 서로 다른 레이어에 반영하는 Multi input LSTM을 도입하여 주식 가격을 예측하였다. 본 연구에서는 이를 참고하여 서로 다른 특성을 변수를 다른 레이어에 반영하여 합치는 구조로 LSTM을 설계하여 연구를 진행하였다.

앞서 본 연구에서 사용한 변수들을 주식시장 외부변수 31개, 주식시장 자금흐름 변수 9개, 주식시장 매매동향 12개, Target factor 4개로 나누었다. 설명변수의 특성에 기반하여 설명변수가 종속변수에 반영되는 경로에 대한 다음과 같은 가설을 세울 수 있다, 주식시장의 자금흐름이 시장의 배경 여건이라고 할 때, 주식시장 외부변수가 시장의 배경여건에 반영되었을 때, 매매 동향에서 일정 패턴이 나타나고, 이것이 시장의 변동성에 반영된다. 본 연구는 이와 같은 구조를 반영하여 Multi-Layer LSTM 알고리즘을 설계하였다. 해당 알고리즘의 구조는 다음과 같다. 활용한 레이어는 4개이다.

1. 첫번째 레이어에 주식시장 외부변수 31개를 대입하면, 해당 레이어는 주식시장 외부변수에 대한 정보를 반환
2. 첫번째 레이어에서 나온 정보와 두번째 레이어에 주식시장 자금흐름을 대입하여 도출한 결과를 결합
3. 앞서 결합한 정보와 주식시장 매매동향 및 VKOSPI의 Lag를 결합하여 세번째 레이어에 대입
4. 세번째 레이어의 결과와 Target Factor 4개를 결합하여 네 번째 레이어에 대입하고 최종적인 설명변수의 정보 도출
5. VKOSPI 예측
6. epoch 수에 따라 1-5 과정 반복 학습

< Figure 2 > Multi-Layer LSTM Structure



LSTM에서는 epoch 값을 통해 학습하는 수를 정한다. 한번 학습이 이루어지면, 해당 학습의 예측값과 실제값의 차이, 즉 오차가 계산되고, 해당 오차에 따라 알고리즘 내의 파라미터 값이 바뀌면서 지속적으로 학습이 이루어진다. 본 연구에서는 epoch를 100으로 설정하여 알고리즘을 학습시켰으며, unit과 batch size를 각각 16-32-40, 20-25-30으로 바꾸어가며 튜닝을 진행하였다. 결론적으로, 본 연구에서는 epoch 100, batch size 25, unit 40, Adam optimizer를 알고리즘에 반영하여 예측을 진행하였다.

III-2. 예측 과정 및 벤치마크 모형

Kim and Han(2022)은 VIX를 예측하면서 예측 기간을 5일, 10일, 22일, 66일로 설정하였다. 본 연구는 해당 연구와 다르게 주 단위 예측을 진행하므로 예측 기간을 1주, 2주, 4주, 8주로 하여 예측을 진행하였다. 예측 성과는 실제값과 예측값의 오차를 통해 도출하는 Root Mean Squared Error(RMSE) 및 Mean Absolute Error(MAE)를 통해 평가하였다. RMSE와 MAE는 직관적인 표현 방식으로 예측 방법간 예측력을 비교함에 있어 널리 이용되고 있다.

기존 계량경제학에서는 내재변동성 지표를 예측하기 위해 Heterogeneous Autoregressive (HAR) 모형을 활용하였다. 이 모형은 변동성이 급격하게 증가 혹은 감소하는 구간에 의해 예측 모형의 오차가 커지는 것을 방지하기 위해 Lag term이 아닌 Lag term들의 평균값을 활용하여 모형을 추정하고, 예측을 진행한다. 본 연구에서는 HAR을 벤치마크 모형으로 활용하였다. 아래 수식은 HAR(1) 모형을 나타낸 것이며, RV는 Realized Volatility를 의미한다.

$$RV_t = \alpha_0 + \alpha_1 \frac{1}{2} \sum_{k=1}^2 RV(t-k) + u_t$$

또한, Kim and Han(2022)에서 VIX 예측에 보루타 알고리즘을 활용한 변수 선택 이후 랜덤포레스트를 활용한 방식이 기존 HAR보다 우월하였음을 참고하여 보루타 알고리즘을 통한 변수 선택 이후 랜덤포레스트 및 XGBoost를 활용한 예측을 추가적으로 진행하여 결과를 비교하였다.

IV. 결과

IV-1. 보루타 알고리즘을 통한 변수 선택 결과

Kim and Han(2022)는 제시한 보루타 알고리즘, 교차검증 및 랜덤포레스트를 통해 변수를 선택한다. 이를 활용하여 VKOSPI를 예측하는 데 중요한 변수를 선정하였고, 각 예측 기간별로 47개, 43개, 51개, 62개 변수가 선택되었다. 이 중 상위 25개 중요 변수 목록을

살펴보면, 주식시장 매매동향 변수 중에서는 차익거래 및 비차익거래 관련 변수가 중요변수로 선정되었으며, 주식시장 자금흐름 변수에서는 위탁매매 미수금 및 반대매매 비중 변수가 중요변수로 선정되었다. 주식시장 외부변수에서는 VIX, S&P, IXIC(나스닥) 변수가 중요한 것으로 나타났다. 보루타 알고리즘을 통해 본 연구에서 신규로 도입한 변수들이 VKOSPI 예측에 있어 상당히 높은 중요도를 가지고 있는 것으로 밝혀졌다. 자금흐름 변수 중 위탁매매 미수금은 각 예측기간에서 모두 10위 이내에 위치하였으며, 매매동향 변수 중 차익거래 및 비차익거래 변수 또한 각 예측 기간에서 15위 이내의 중요도를 가진 것으로 나타났다. 보루타 알고리즘을 통해 선택된 모든 변수는 부록에 수록하였다.

< Table 3 > Variable Rankings 1 to 15 determined by Boruta Algorithm

Rank	1 Weeks Ahead	2 Weeks Ahead	4 Weeks Ahead	8 Weeks Ahead
1	VKOSPI(1)	VKOSPI(1)	VKOSPI(1)	차익거래(매수)(1)
2	VIX(1)	VIX(1)	위탁매매_미수금(1)	위탁매매_미수금(1)
3	VKOSPI(2)	VKOSPI(2)	위탁매매_미수금(2)	위탁매매_미수금(3)
4	VKOSPI(3)	VKOSPI(3)	VIX(1)	위탁매매_미수금(4)
5	VIX(2)	VIX(2)	위탁매매_미수금(4)	위탁매매_미수금(2)
6	VKOSPI(4)	위탁매매_미수금(3)	위탁매매_미수금(3)	VIX(1)
7	VIX(3)	위탁매매_미수금(4)	VKOSPI(2)	VKOSPI(1)
8	위탁매매_미수금(4)	위탁매매_미수금(2)	VIX(4)	차익거래(매도)(1)
9	VIX(4)	VIX(3)	VKOSPI(4)	VIX(2)
10	위탁매매_미수금(3)	VKOSPI(4)	VKOSPI(3)	차익거래(매수)(4)
11	차익거래(매도)(2)	차익거래(매도)(1)	차익거래(매수)(1)	VIX(3)
12	차익거래(매도)(1)	차익거래(매도)(2)	차익거래(매도)(1)	VIX(4)
13	차익거래(매도)(3)	비차익거래(매도)(1)	VIX(3)	VKOSPI(2)
14	차익거래(매수)(2)	위탁매매_미수금(1)	차익거래(매수)(4)	차익거래(매수)(2)
15	비차익거래(매도)(2)	차익거래(매수)(3)	VIX(2)	차익거래(매수)(3)

IV-2. 예측력 비교

먼저, 예측 방법간 RMSE 및 MAE를 비교하였다. MAE를 기준으로, 1주 다단계 예측에서는 HAR 모형이 가장 낮은 MAE를 기록하였다. 2주 다단계 예측 또한 HAR 모형이 가장 낮은 MAE를 기록하였는데, 이는 변동성 예측에 있어 HAR의 예측성능이 우수한 결과를 보여왔던 기존 연구들을 고려하면 놀라운 결과는 아니다. 4주 다단계 예측에서는

Boruta-XGBoost가 가장 우수한 결과를 보였으며, 8주 다단계 예측에서는 Boruta-RF가 가장 우수한 결과를 보였으며 이 결과는 Kim and Han(2022)에서 5 Days Ahead Forecast, 10 Days Ahead Forecast, 22 Days Ahead Forecast에서 Boruta-RF가 우수한 예측성능을 보였던 바와 유사한 결과이다. Multi-Layer LSTM의 경우, 1주, 2주 및 4주 다단계 예측에서 다른 예측모형들과 비교했을 때 눈에 띄는 우수한 예측력을 보이지는 못하고 있으나, 8주 다단계 예측에서는 Boruta-RF 다음으로 우수한 예측성능을 보였다.

< Table 4 > Model Confidence Set for Each Forecast Horizon

Model Algorithm	1 Week Ahead		2 Weeks Ahead		4 Weeks Ahead		8 Weeks Ahead	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Multi-Layer LSTM	0.121	0.092	0.153	0.117	0.176	0.139	0.162	0.124
LSTM	0.139	0.111	0.159	0.122	0.195	0.156	0.167	0.132
Boruta - XGBoost	0.102	0.079	0.131	0.101	0.140	0.109	0.160	0.127
Boruta - RF	0.100	0.079	0.124	0.097	0.142	0.115	0.153	0.121
HAR	0.098	0.073	0.130	0.097	0.161	0.125	0.171	0.136
HARX	0.127	0.101	0.490	0.148	0.204	0.168	0.232	0.187

다음으로, 위 RMSE 및 MAE 비교 결과가 통계적으로 예측 방법간 유의미한 예측력 차이를 나타내는지 검정하기 위해 Giacomini-White 검정 및 MCS 검정을 실시하였다. 먼저, 각 다단계 예측 수준에서 Multi-Layer LSTM과 각 예측 수준의 최소 MAE를 기록한 예측방법간 Giacomini-White 검정을 실시하였다. 1주 다단계 예측에서는 HAR모형이 가장 우수한 MAE결과를 보였으며, 2주 다단계 예측 또한 HAR모형이 가장 우수하였다. 4주 다단계 예측은 보루타 알고리즘을 통한 변수선택 후의 XGBoost가 가장 우수하였으며, 8주 다단계 예측은 RF-Boruta가 가장 우수한 결과를 보였으므로 위 4가지 경우의 검정을 실시하였다. Giacomini-White 검정은 두 모형 간 유의미한 예측력 차이가 있는지 검정하는 방법이며, 상대 모형인 모형1은 각 다단계 예측별로 가장 우수한 MAE를 가진 모형으로, 그리고 모형2는 Multi-Layer LSTM으로 설정하였다. 만약 Giacomini-White 통계량이 충분히 큰 음(-)의 값을 가질 경우, 상대 모형인 모형1의 예측력이 모형2보다 유의하게 높다는 결과로 해석될 수 있다. 검정 결과, 1주, 2주 및 4주 다단계 예측에서는 위 RMSE, MAE 비교 결과에서 예상하였듯 Multi-Layer LSTM의 예측력이 유의하게 낮다는 결과도 출되었다. 하지만, 8주 다단계 예측에서는 최소 MAE를 가지는 RF-Boruta 대비 예측력이 낮으나 그 결과는 유의하지 않은 것으로 나타났으며, 8주 다단계 예측에서는 Multi-Layer LSTM이 RF-Boruta와 함께 가장 우수한 예측력을 가졌다는 결론을 도출할 수 있다.

< Table 5 > Giacomini-White Test Result for Each Forecast Horizon

Forecast Horizon	Minimum MAE	Statistics	P-Value
1 Week ahead	HAR	-3.961	0.000
2 Weeks ahead	HAR	-3.320	0.001
4 Weeks ahead	XGB-Boruta	-2.100	0.035
8 Weeks ahead	RF-Boruta	-0.362	0.717

* Multi-Layer LSTM과 각 예측기간별 최소 MAE 모형 비교

마지막으로, Model Confidence Set (MSC) 검정을 실시하였다. MCS 검정은 예측에 사용된 모든 모형을 대상으로 검정을 실시하며, 각 기간별 다단계 예측별로 예측력이 우수하며, 동시에 모형간 예측력 차이가 유의미하게 발생하지 않는 모형들이 Superior Model로 결정되어 Confidence Set으로 분류된다. 여기서는 한희준(2023)에서 설정한 바와 같이 유의수준을 50%로 설정하였다. Multi-Layer LSTM은 1주, 2주 및 4주 다단계 예측에서는 Confidence Set에 포함되지 못하였으나, 8주 다단계 예측에서는 Confidence Set에 포함되어 우수한 예측력을 가진다는 결과를 발견하였다. 보루타 알고리즘을 활용한 Boruta-XGBoost 및 Boruta-RF은 모든 예측기간에서 Confidence Set에 포함되었다.

< Table 6 > Model Confidence Set for Each Forecast Horizon

Model	1 Week Ahead	2 Weeks Ahead	4 Weeks Ahead	8 Weeks Ahead
Multi-Layer LSTM				○
LSTM				○
Boruta-XGBoost	○	○	○	○
Boruta-RF	○	○	○	○
HAR	○	○		○
HARX				

IV-3. LSTM과 Multi-Layer-LSTM의 비교

본 연구에서 중점적으로 다루고 있는 Multi-Layer LSTM은 기본적으로 LSTM의 구조를 조정하여 활용하고 있으므로, 일반적인 LSTM 예측과의 비교결과 또한 중요하다. 이는, 만약 Multi-Layer LSTM이 일반 LSTM 대비 예측 성능이 유의미하게 좋지 않다면 Layer와 변수를 차별화하여 예측을 설계할 필요가 없기 때문이다. 먼저, RMSE와 MAE를 비교하였다. 모든 기간의 다단계 예측에서 Multi-Layer LSTM은 일반적인 LSTM 예측보다 우수한 RMSE와 MAE를 기록하였다. 다음으로, 두 예측방법간 Giacomini-White 검정을 실시하였다. 1주 및 2주 다단계 예측에서는 두 모형간 유의미한 예측력 차이를 발견할 수 없었으나, 4주 및 8주 다단계 예측에서는 LSTM 대비 Multi-Layer LSTM이 유의미하게 예측력이 우수하다는 결과를 도출하였다. 더욱이, 각 기간의 P-value는 4주 예측 0.018, 8주 예측 0.004로 예측 기간이 늘어날수록 귀무가설(두 모형간 예측력 차이가 없음)을 기각할 수 있는 유의수준 또한 낮아짐을 발견하였다.

< Table 7 > Giacomini-White Test Result for LSTM and Multi-Layer LSTM

Forecast Horizion	Test Statistics	P-Value
1 Week ahead	1.254	0.210
2 Weeks ahead	0.066	0.947
4 Weeks ahead	2.361	0.018**
8 Weeks ahead	2.857	0.004***

V. 결론

본 연구에서는 새로운 변수와 새로운 머신러닝 알고리즘을 도입해 VKOSPI를 예측하였다. 본 연구의 주요 결과는 다음과 같다.

첫 번째, VKOSPI를 예측하는 데 유효한 신규 변수들을 식별하였다. 미수금과 관련된 주식시장 자금흐름을 나타내는 변수들이 VKOSPI를 예측함에 있어 효과적으로 활용될 수 있음을 발견하였다. Kim and Han(2022)가 제시한 보루타 알고리즘, 교차검증 및 랜덤포레스트를 통한 최적 변수 선택 결과, 위탁매매 미수금 및 미수금 대비 반대매매 비중 등의 변수가 VKOSPI를 예측함에 있어 높은 중요도를 가짐을 발견하였다. 또한, 주식시장 매매동향과 관련된 차익거래, 비차익거래 매수/매도 변수들이 VKOSPI를 예측함에 있어 높은 중요도를 가진 변수로 활용될 수 있음을 발견하였다.

두 번째, 8주 다단계 예측에서 Multi-Layer LSTM이 우수한 예측성능을 가짐을 발견하였다. HAR 및 랜덤포레스트 등 다른 예측방법과 비교했을 때, 우수한 절대평균오차를

기록하였으며 최소 절대평균오차를 기록한 보루타 알고리즘을 통한 변수 선택을 적용한 랜덤 포레스트와 Giacomini-White 검정을 실시한 결과 유의미한 예측력 차이는 발견되지 않았다. 마지막으로, VKOSPI 예측에 있어 Multi-Layer LSTM이 LSTM보다 우수한 예측성능을 가짐을 발견하였다. 모든 기간의 다단계 예측에서 LSTM 대비 낮은 절대평균오차를 기록하였으며, Giacomini-White 검정 결과 4, 8주 다단계 예측에서 유의미한 예측력 차이를 나타냄을 발견하였다. 기존 경제 시계열 예측에서는 설명력이 부족한 것으로 평가받았던 LSTM이 알고리즘 구조를 통해 설명력을 확보하면서 동시에 예측력도 높일 수 있음을 시사한다.

본 연구에서는 주식시장 외부 변수, 주식시장 자금 흐름, 주식시장 매매동향으로 변수의 그룹의 구분하여 LSTM 알고리즘에 대입하는 방식으로 예측을 진행했다. 하지만 LSTM의 다층 레이어를 활용하면 다른 구조도 설계할 수 있기 때문에 예측력이 개선된 다른 알고리즘이 있을 가능성을 배제할 수 없다. 또한, 본 연구는 8주 이상의 예측 기간에 대해 예측을 진행하지 않았다. 따라서 8주를 넘어선 중장기 예측력은 확인할 수 없었다. 따라서 향후 경제 시계열 예측 연구에서 LSTM 알고리즘을 개선하고 예측 기간과 Hyperparameter를 다양하게 조절하면서 추가 연구를 진행한다면, 본 연구에서 나타난 것보다 개선된 성과를 얻을 수 있을 것으로 기대한다.

[참고문헌 Reference]

- Chun, D., Cho, H., & Ryu, D. (2020). Economic indicators and stock market volatility in an emerging economy. *Economic Systems*, 44(2), 100788.
- Han, H., Kutan, A. M., & Ryu, D. (2015). Effects of the US stock market return and volatility on the VKOSPI. *Economics*, 9(1), 20150035.
- Kim, B. Y., & Han, H. (2022). Multi-Step-Ahead Forecasting of the CBOE Volatility Index in a Data-Rich Environment: Application of Random Forest with Boruta Algorithm. *Korean Economic Review*, 38(3), 541-569.
- Kim, S. W. (2010). A study on developing a VKOSPI forecasting model via GARCH class models for intelligent volatility trading systems. *Journal of Intelligence and Information Systems*, 16(2), 19-32.
- Liu, M., Ye, J., & Yu, L. (2022). Volatility Prediction via Hybrid LSTM Models with GARCH Type Parameters. *Proceedings of Business and Economic Studies*, 5(6), 37-46.
- Li, H., Shen, Y., & Zhu, Y. (2018, November). Stock price prediction using attention-based multi-input LSTM. In *Asian conference on machine learning* (pp. 454-469). PMLR.
- 류형선, 양성국, & 김봉현. (2013). 한국 주식시장의 투자주체별 투자행태가 주가지수 수익률에 미치는 영향. *글로벌경영학회지*, 10(3), 355-370.
- 유한수. (2006). 투자주체별 거래와 주가변동성. *기업경영연구*, 13(2), 67-77.
- 이승희, & 한희준. (2016). 경제/금융 변수를 이용한 한국 주식시장의 변동성분석 및 예측. *경제학연구*, 64(2), 67-95.

[부록 Appendix]

< Table 1 > T-code

Variables	T-Code
VKOSPI (Target)	4
KOSPI	5
KOSDAQ	5
개인	1
외국인	1
기관계	1
차익거래(매수)	1
차익거래(매도)	1
차익거래(순매수)	1
비차익거래(매수)	1
비차익거래(매도)	1
비차익거래(순매수)	1
Dollar Index	5
USDKRW	5
신용잔고	5
주식형	5
신용비율	2
고객예탁금	5
장내파생상품 예수금	5
대고객 환매조건부채권 매도잔고	5
위탁매매 미수금	1
반대매매 금액	5
미수금 대비 반대매매 비중	1
DJI	5
IXIC	5
S&P	5
VIX	4
WTI	5
콜금리 1일	2

KORIBOR 3m	2
KORIBOR 6m	2
KORIBOR 12m	2
CD 91	2
국고채 1y	2
국고채 3y	2
국고채 5y	2
국고채 10y	2
국고채 20y	2
국고채 30y	2
통화안정증권 91d	2
통화안정증권 1y	2
통화안정증권 2y	2
회사채 AA+	2
회사채 BBB	2
회사채 AA-	2
MMF 7d	2
KOSPI200 선물	5
usrate 3m	2
usrate 5y	2
usrate 1y	2
usrate 30y	2
gold	5

< Table 2 > Variable Ranking Determined by Boruta Algorithm

Rank	1 Weeks Ahead	2 Weeks Ahead	4 Weeks Ahead	8 Weeks Ahead
1	VKOSPI(1)	VKOSPI(1)	VKOSPI(1)	차익거래(매수)(1)
2	VIX(1)	VIX(1)	위탁매매 미수금(1)	위탁매매 미수금(1)
3	VKOSPI(2)	VKOSPI(2)	위탁매매 미수금(2)	위탁매매 미수금(3)
4	VKOSPI(3)	VKOSPI(3)	VIX(1)	위탁매매 미수금(4)
5	VIX(2)	VIX(2)	위탁매매 미수금(4)	위탁매매 미수금(2)
6	VKOSPI(4)	위탁매매 미수금(3)	위탁매매 미수금(3)	VIX(1)
7	VIX(3)	위탁매매 미수금(4)	VKOSPI(2)	VKOSPI(1)
8	위탁매매 미수금(4)	위탁매매 미수금(2)	VIX(4)	차익거래(매도)(1)
9	VIX(4)	VIX(3)	VKOSPI(4)	VIX(2)
10	위탁매매 미수금(3)	VKOSPI(4)	VKOSPI(3)	차익거래(매수)(4)
11	차익거래(매도)(2)	차익거래(매도)(1)	차익거래(매수)(1)	VIX(3)
12	차익거래(매도)(1)	차익거래(매도)(2)	차익거래(매도)(1)	VIX(4)
13	차익거래(매도)(3)	비차익거래(매도)(1)	VIX(3)	VKOSPI(2)
14	차익거래(매수)(2)	위탁매매_미수금(1)	차익거래(매수)(4)	차익거래(매수)(2)
15	비차익거래(매도)(2)	차익거래(매수)(3)	VIX(2)	차익거래(매수)(3)
16	차익거래(매도)(4)	차익거래(매수)(2)	차익거래(매도)(2)	VKOSPI(3)
17	차익거래(매수)(4)	차익거래(매수)(1)	차익거래(매도)(4)	VKOSPI(4)
18	차익거래(매수)(3)	차익거래(매도)(3)	차익거래(매수)(3)	차익거래(매도)(2)
19	위탁매매_미수금(2)	VIX(4)	비차익거래(매도)(1)	비차익거래(매도)(4)
20	S&P(1)	반대매매 비중(1)	차익거래(매수)(2)	비차익거래(매도)(3)
21	차익거래(매수)(1)	차익거래(매도)(4)	비차익거래(매수)(1)	비차익거래(매도)(2)
22	위탁매매_미수금(1)	비차익거래(매수)(1)	차익거래(매도)(3)	비차익거래(매도)(1)
23	반대매매_비중(1)	차익거래(매수)(4)	반대매매 비중(4)	반대매매 비중(1)
24	IXIC(1)	비차익거래(매도)(2)	반대매매 비중(1)	차익거래(매도)(4)
25	비차익거래(매수)(2)	비차익거래(매수)(2)	반대매매 비중(3)	차익거래(매도)(3)

26	비차익거래(매도)(1)	비차익거래(매도)(3)	반대매매 비중(2)	반대매매 비중(3)
27	반대매매_비중(2)	비차익거래(매수)(3)	비차익거래(매도)(2)	반대매매 비중(4)
28	비차익거래(순매수)(1)	반대매매_비중(2)	MMF 7d(1)	반대매매 비중(2)
29	외국인(1)	반대매매_비중(3)	MMF 7d(4)	비차익거래(매수)(4)
30	개인(1)	MMF 7d(3)	비차익거래(매수)(2)	비차익거래(순매수)(3)
31	비차익거래(매도)(3)	비차익거래(순매수)(1)	MMF 7d(2)	비차익거래(매수)(1)
32	비차익거래(매도)(4)	비차익거래(매도)(4)	비차익거래(매도)(4)	비차익거래(매수)(2)
33	KOSPI(1)	반대매매 비중(4)	신용잔고(1)	비차익거래(매수)(3)
34	비차익거래(매수)(1)	IXIC(1)	개인(2)	MMF 7d(2)
35	비차익거래(매수)(4)	비차익거래(순매수)(4)	비차익거래(매도)(3)	MMF 7d(1)
36	비차익거래(매수)(3)	MMF 7d(4)	비차익거래(순매수)(4)	MMF 7d(4)
37	KOSPI200 선물(1)	개인(4)	MMF 7d(3)	개인(1)
38	신용잔고(1)	비차익거래(매수)(4)	비차익거래(매수)(4)	KOSPI(3)
39	DJI(1)	USRATE 3M(4)	비차익거래(순매수)(2)	개인(3)
40	반대매매 비중(3)	S&P 500(1)	비차익거래(순매수)(1)	KOSPI(1)
41	반대매매 비중(4)	신용잔고(2)	비차익거래(순매수)(3)	MMF 7d(3)
42	MMF 7d(4)	비차익거래(순매수)(3)	USRATE 3M(2)	비차익거래(순매수)(2)
43	KOSDAQ(1)	신용잔고(3)	비차익거래(매수)(3)	비차익거래(순매수)(4)
44	신용잔고(3)		외국인(2)	외국인(3)
45	비차익거래(순매수)(4)		개인(1)	비차익거래(순매수)(1)
46	비차익거래(순매수)(2)		콜금리_1일(1)	개인(2)
47	신용잔고(4)		KORIBOR 3m(1)	개인(4)
48			USRATE 3M(3)	KORIBOR 3m(1)
49			신용잔고(4)	KOSPI(2)
50			KOSPI200 선물(1)	KOSPI200 선물(1)
51			신용잔고(2)	주식형(2)
52				외국인(2)

53	외국인(1)
54	신용잔고(1)
55	신용잔고(4)
56	KORIBOR_6m(1)
57	기관계(1)
58	기관계(2)
59	콜금리_1일(1)
60	주식형(3)
61	통안증권_91d(1)
62	KOSPI(4)

< Table 3 > RMSE and MAE Result

Model Algorithm	1 week ahead		2 weeks ahead		4 weeks ahead		8 weeks ahead	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
LSTM	0.139	0.111	0.159	0.122	0.195	0.156	0.167	0.132
Multi-Layer LSTM	0.121	0.092	0.153	0.117	0.176	0.139	0.162	0.124
HAR	0.098	0.073	0.130	0.097	0.161	0.125	0.171	0.136
HARX	0.127	0.101	0.490	0.148	0.204	0.168	0.232	0.187
Random Walk	0.099	0.076	0.137	0.105	0.177	0.141	0.202	0.162
LASSO	0.100	0.075	0.143	0.111	0.166	0.131	0.185	0.141
Adaptive LASSO	0.100	0.073	0.150	0.115	0.165	0.128	0.173	0.134
ElasticNet	0.103	0.079	0.144	0.112	0.167	0.132	0.187	0.142
Ada.ElasticNet	0.102	0.077	0.149	0.114	0.168	0.133	0.176	0.135
Target Factor	0.100	0.078	0.141	0.112	0.179	0.143	0.197	0.158
Random Forest	0.102	0.082	0.128	0.103	0.149	0.121	0.163	0.131
XGBoost	0.108	0.083	0.131	0.103	0.150	0.120	0.169	0.134
Boruta-RF	0.100	0.079	0.124	0.097	0.142	0.115	0.153	0.121
Boruta-XGB	0.102	0.079	0.131	0.101	0.140	0.109	0.160	0.127

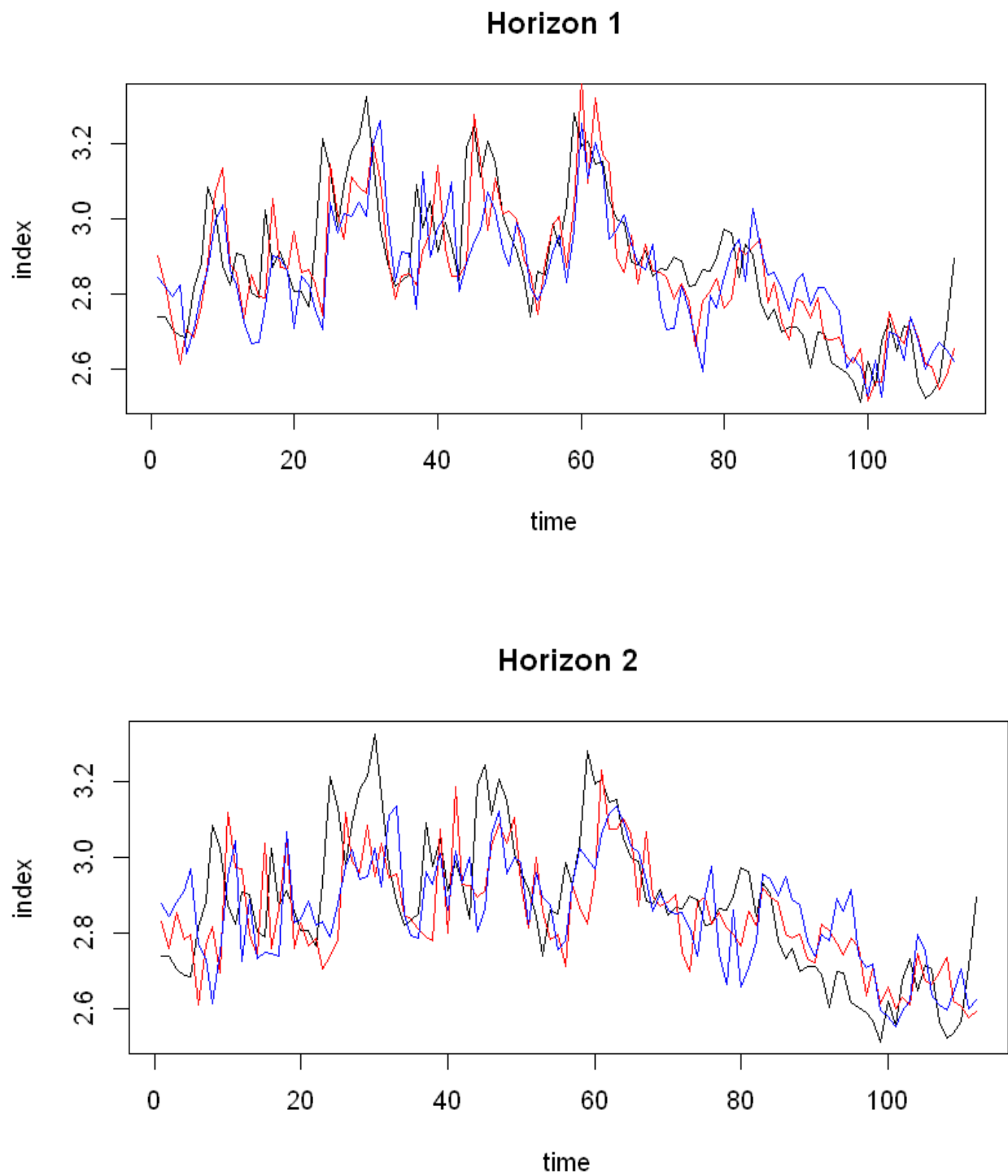
< Table 4 > Model Confidence Set Result

Model Algorithm	1 Week Ahead	2 Weeks Ahead	4 Weeks Ahead	8 Weeks Ahead
Multi-Layer LSTM				○
LSTM				○
Boruta – XGBoost	○	○	○	○
Boruta – RF	○	○	○	○
HAR	○	○		○
HARX				
Random Walk	○			
LASSO	○			○
Adaptive LASSO	○			○
E.Net	○			○
Adaptive E.Net	○			○
Target Factor	○			○
Random Forest	○	○		○
XGBoost	○	○		○

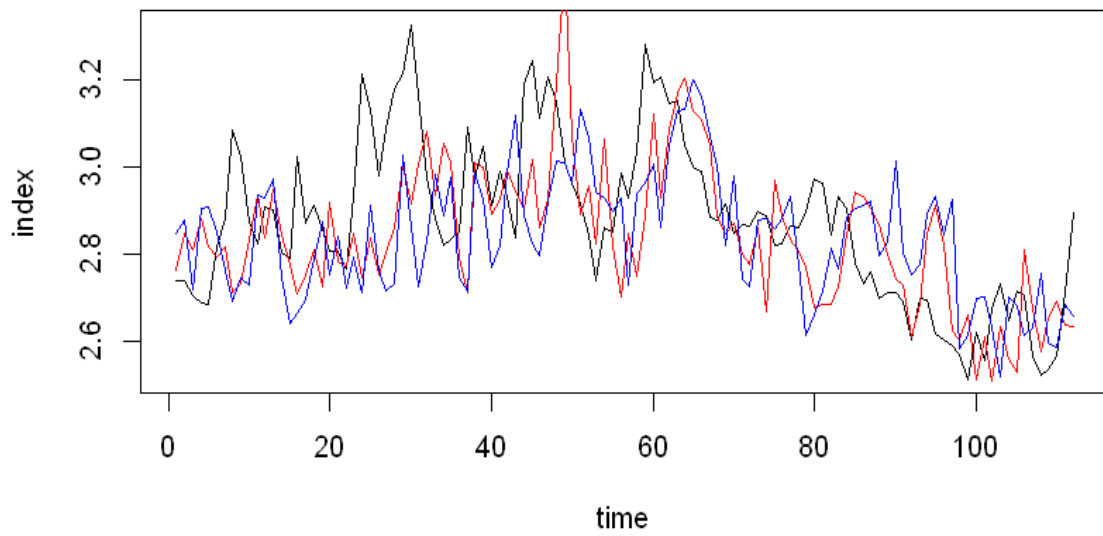
< Table 5 > LSTM Tuning

Model Algorithm	1 week ahead		2 weeks ahead		4 weeks ahead		8 weeks ahead		Parameter	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	Batch	Unit
LSTM	0.146	0.114	0.181	0.140	0.191	0.157	0.199	0.156	30	32
M-LSTM	0.125	0.095	0.164	0.123	0.178	0.146	0.179	0.142	30	32
LSTM	0.140	0.107	0.160	0.123	0.188	0.150	0.183	0.152	25	32
M-LSTM	0.128	0.092	0.152	0.121	0.175	0.141	0.163	0.123	25	32
LSTM	0.141	0.110	0.160	0.128	0.178	0.140	0.181	0.143	20	32
M-LSTM	0.122	0.093	0.155	0.123	0.168	0.132	0.162	0.125	20	32
LSTM	0.139	0.111	0.159	0.122	0.195	0.156	0.167	0.132	25	40
M-LSTM	0.121	0.092	0.153	0.117	0.176	0.139	0.162	0.124	25	40
LSTM	0.136	0.106	0.176	0.145	0.181	0.143	0.198	0.158	30	16
M-LSTM	0.130	0.099	0.145	0.113	0.170	0.134	0.181	0.143	30	16
LSTM	0.130	0.100	0.160	0.122	0.190	0.152	0.177	0.143	20	40
M-LSTM	0.119	0.092	0.153	0.121	0.169	0.129	0.166	0.125	20	40

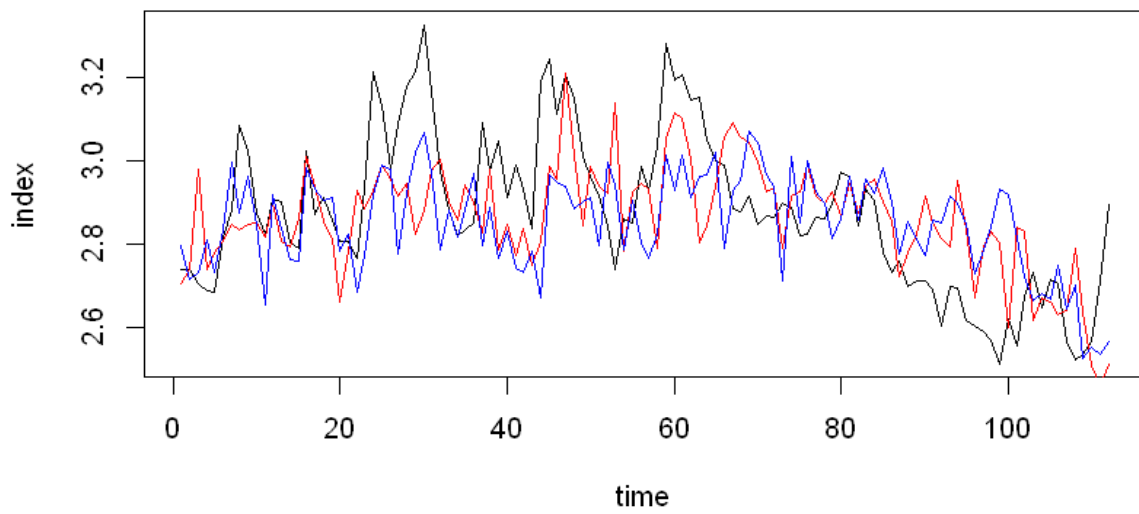
< Figure 1 > Comparison between LSTM and Multi-Layer LSTM



Horizon 3



Horizon 4



본 연구에서 사용한 코드와 데이터는 모두 Github에 업로드하였다.

https://github.com/popper6508/202302_MLFTF_final