
LSTM-Based Volatility Forecasting: Effects and Limitations of Feature Expansion

Team Members

남수경	20259135
손민재	20259217
지윤승	20259431



Introduction

연구 목적

- S&P 500 지수의 실현 변동성(realized volatility) 을 높은 정확도로 예측하는 모델 개발
- 가격 정보 기반 LSTM 모델에 GARCH 변동성·VIX 내재 변동성을 결합해 정보 추가가 예측력을 얼마나 개선하는지 검증

예측 대상: 실현 변동성(Realized Volatility)

- S&P 500 일별 로그수익률을 사용
- 22일(약 1개월) 롤링 표준편차로 계산
- 시장 변동성의 실제 움직임을 반영하는 핵심 지표

프로젝트 개요

- LSTM 기반 변동성 예측 모델 3종 비교
 - LSTM:** 가격 정보 + 전일 실현 변동성
 - LSTM-GARCH:** 가격 정보 + 전일 실현 변동성 + GARCH(2,2) 단기 예측 변동성
 - Hybrid (LSTM + GARCH + VIX):** 가격 정보 + 전일 실현 변동성 + GARCH(2,2) 단기 예측 변동성 + VIX(30일 내재 변동성) 포함

Motivation

변동성 예측의 중요성

- 변동성은 금융시장의 위험 수준을 직접적으로 반영하는 핵심 지표
- 옵션 가격결정, 리스크 관리, 자산배분, 스트레스 테스트 등 대부분의 금융 모델이 변동성에 민감하게 의존
- 특히 S&P 500 변동성은 글로벌 시장 전반의 위험 신호로 활용됨

기존 접근(GARCH, VIX)의 한계

- GARCH 계열 모델
과거 변동성 패턴을 잘 설명하지만 비선형적·장기적 패턴 반영이 제한적
극단적 이벤트(팬데믹, 금융위기) 대응력이 낮음
- VIX
시장 기대를 반영하지만 옵션 시장 수급·왜곡에 영향을 받음
단독으로는 실현 변동성을 안정적으로 예측하기 어려움

LSTM 기반 접근의 필요성

- LSTM은 장·단기 의존성을 모두 학습할 수 있어서 Volatility Clustering, Regime Shift, Nonlinearity를 자연스럽게 포착
- 가격 정보, 과거 변동성, GARCH 예측치, VIX 등 이질적 정보를 시계열 구조에서 통합적으로 학습 가능
- 전통 모델에서는 설명하기 어려운 실제 시장의 복잡한 변동성 움직임을 데이터 기반으로 직접 학습할 수 있음

Data & Experimental Setting

Data

항목	내용
데이터 소스	Yahoo Finance (SPX, VIX)
기간	2000-01-01 ~ 2024-12-31
관측 빈도	일별
사용 변수	SPX, VIX , ret(SPX 로그수익률), realized_vol(실현 변동성), lag_realized_vol(전일 실현 변동성), garch_vol_pred(GARCH(2,2) 예측 변동성)

Experimental Setting

항목	설정 내용
예측 대상	S&P 500 Realized Volatility (22일 롤링 표준편차)
Lookback Window	최근 22일 시계열 입력
Feature 구성	LSTM: ret, lag_realized_vol / LSTM-GARCH: ret, lag_realized_vol, garch_vol_pred/ Hybrid: ret, lag_realized_vol, garch_vol__pred, vix
Garch Train 기간	2000-01-01 ~ 2006-02-27
Train 기간	2006-02-28 ~ 2019-05-06(70%)
Validation 기간	2019-05-07 ~ 2022-03-01(15%)
Test 기간	2022-03-02 ~ 2024-12-31(15%)

Feature Construction

공통 입력 변수

- ret: S&P 500 일별 로그수익률
- lag_realized_vol: 22일 실현변동성의 지연값으로, 가격 변동성과 변동성 클러스터링 패턴을 반영하는 핵심 입력 변수임

모델별 Feature 세트 구성

- LSTM: ret, lag_realized_vol
- LSTM-GARCH: ret, lag_realized_vol, garch_vol_pred
- Hybrid (LSTM + GARCH + VIX): ret, lag_realized_vol, garch_vol, vix (옵션 시장이 반영한 30일 기대 변동성)

Feature 구성 의도

- 공통적으로 가격 정보와 변동성의 자기상관 구조를 사용
- GARCH는 단기 조건부 변동성을 제공
- VIX는 시장 참여자의 기대 변동성 정보를 제공
- 결합할수록 더 풍부한 정보 기반 예측 가능

Hyperparameter Search & Model Selection

LSTM 모델의 구조에서 **units, dropout, learning rate** 세 가지에 대해 Tuning을 진행

버전	Units	Dropout	Learning rate	Best Val loss
1	64	0.1	0.001	0.000591
2	128	0.1	0.001	0.062041
3	128	0.2	0.001	0.000625
4	128	0.1	0.0005	0.000629

결과 : 버전 1 (units=64, dropout=0.1, lr=0.001) 채택

- 모든 조합에 대해 validation loss가 큰 차이를 보이지는 않았으나, units=64, dropout=0.1, lr=0.001 조합이 가장 낮은 val_loss를 기록
- 64-unit 모델은 validation loss 대비 연산 복잡도가 낮아 실전 활용성을 고려할 때 효율적이라는 장점이 있음
- 128-unit 조합 중 일부(버전 2)는 오히려 val_loss가 크게 증가하여, 은닉 유닛 수를 단순히 늘리는 것이 항상 성능 개선으로 이어지지 않음을 확인
 - 해당 버전은 validation loss가 개선되지 않아 EarlyStopping에 의해 약 10 epoch에서 조기 종료됨
 - 모델 규모가 커질수록 초기 수렴이 느려지는 경향이 있어, 충분한 학습 기회를 갖지 못해 상대적으로 큰 val_loss가 기록된 것으로 보임
- Layer의 경우, 단층보다 layer가 2일때 장기 패턴 포착에 유리하면서도, 깊은 네트워크에 비해 과적합 위험이 상대적으로 낮아, 튜닝 대상에서 제외

Model Architecture

공통 LSTM 구조 (2-Layer LSTM)

세 모델은 모두 동일한 LSTM 구조를 사용하여 Feature 차이로 인한 성능 변화를 공정하게 비교하도록 설계함

- LSTM Layer 1
 - 64 units
 - activation = tanh
 - return_sequences = True
 - Dropout 0.1 적용
- LSTM Layer 2
 - 64 units
 - activation = tanh
 - return_sequences = False
 - Dropout 0.1 적용
- Output Layer
 - Dense(1)
 - activation = ReLU (변동성은 음수가 될 수 없기 때문에 ReLU 사용)

시계열 입력 구조

- Lookback Window는 지난 22일간의 시계열 데이터를 하나의 입력 샘플로 사용
- 입력 형태(22 × Feature 수)

LSTM: 22×2
LSTM-GARCH: 22×3
Hybrid: 22×4

학습 설정

- Optimizer: Adam (lr=0.001)
- Loss: MSE
- EarlyStopping

기준: val_loss

10 epoch 동안 추가 개선이 없으면 학습 중단 및 best weights 복원

Model Architecture

Model 1: LSTM

- 입력 차원: 22×2
- 사용 Feature: ret (SPX 로그수익률), lag_realized_vol (전일 실현 변동성)
- 특징

가격 정보와 변동성의 자기상관만으로 예측

딥러닝만으로 어느 정도까지 변동성을 설명할 수 있는지 보여주는 기본 기준

Model 2: LSTM-GARCH

- 입력 차원: 22×3
- 사용 Feature: ret, lag_realized_vol, garch_vol_pred
- 특징

전통적 GARCH 모델이 추정한 단기 조건부 변동성 정보를 LSTM에 주입

전통적 통계 모델(GARCH)과 딥러닝(LSTM)을 결합한 하이브리드 구조로 예측 안정성 향상 시도

Model 3: Hybrid (LSTM + GARCH + VIX)

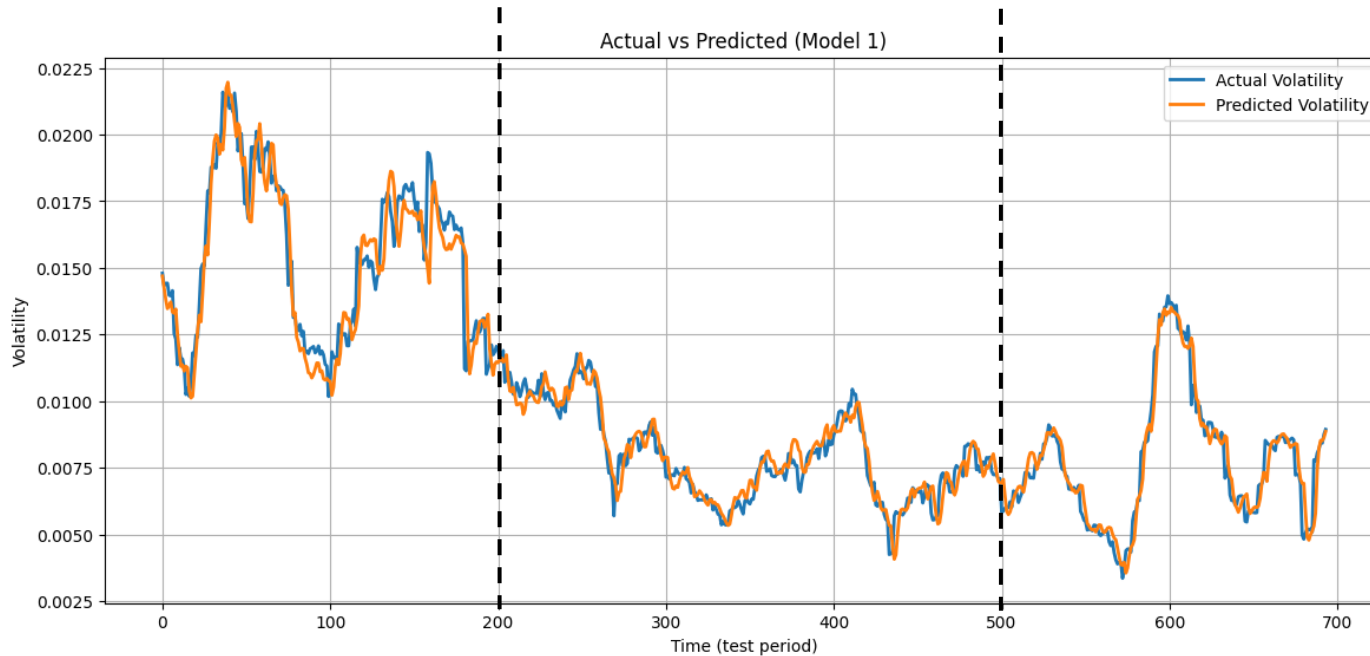
- 입력 차원: 22×4
- 사용 Feature: ret, lag_realized_vol, garch_vol_pred, vix
- 특징

과거 realized vol + GARCH 예측 + 옵션 시장의 기대(VIX) 를 모두 결합한 하이브리드 구조

데이터 기반(실현 변동성) + 모형 기반(GARCH) + 시장 기대(VIX)를 LSTM이 한 번에 학습하도록 설계한 가장 리치한 정보 조합

Results – LSTM(Model 1)

모델	MAE	RMSE
LSTM	0.000546	0.000791



성능 지표 분석

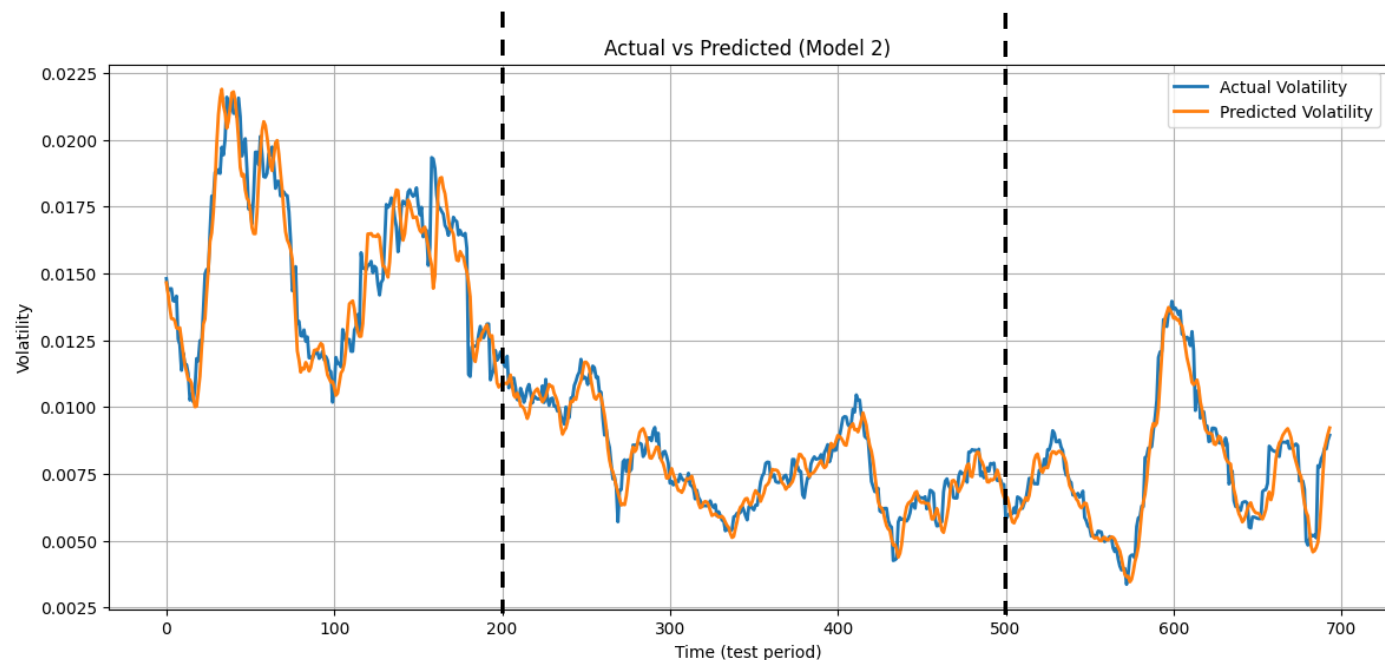
- MAE와 RMSE가 가장 낮은 모델
- 예측 정확성과 안정성 측면에서 가장 우수한 성능을 보임

예측 그래프 분석

- 초반(0~200): 급등 구간을 정확히 따라가며 과도한 과대반응이나 과소반응 없이 매끄럽게 대응함
- 중반(200~500): 두 선의 차이가 거의 보이지 않고 완만한 하락·횡보 흐름을 정확히 재현
- 후반(500~700): 변동성 재상승 구간에서 피크의 위치와 높이를 정확히 반영, 세부 파형까지 거의 동일함
- 전체 구간에서 실제 변동성과 예측값이 거의 완전히 겹치며 급등·급락 시점에서도 민감하고 안정적인 추적이 이루어짐

Results – LSTM-GARCH (Model 2)

모델	MAE	RMSE
LSTM_GARCH	0.000605	0.000839



성능 지표 분석

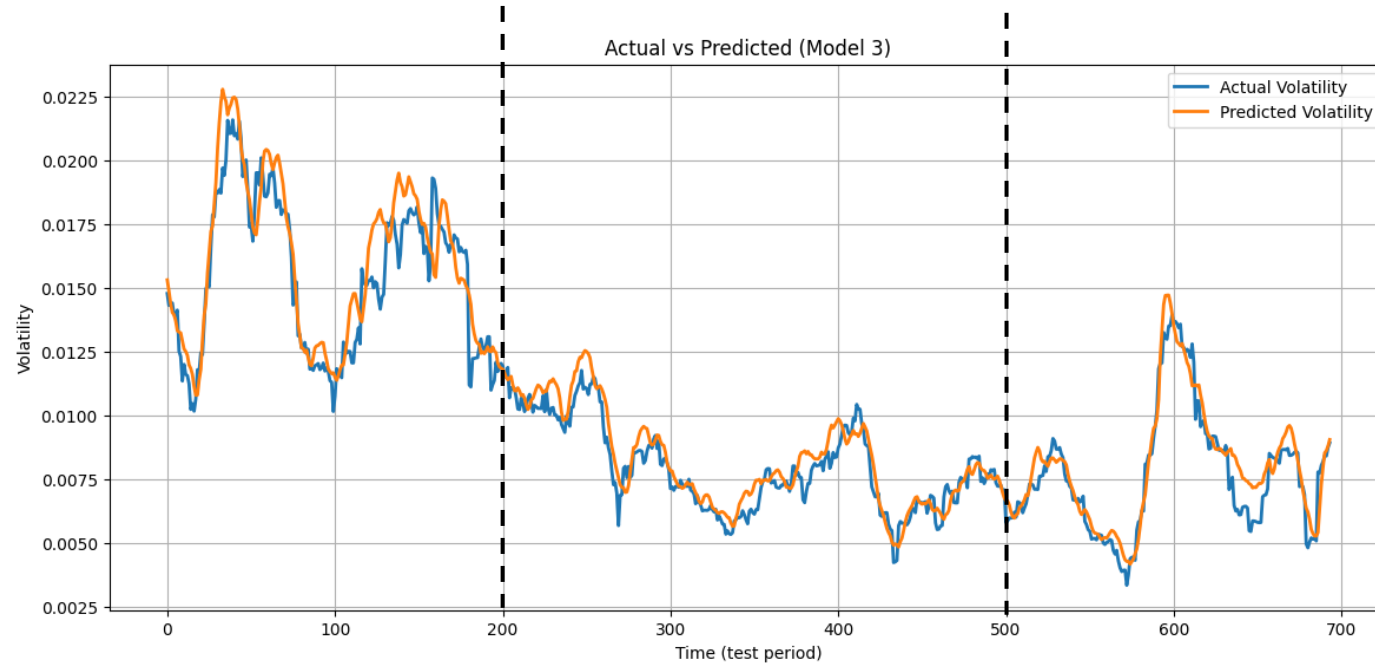
- MAE와 RMSE가 낮은 값으로 나타나면서 전반적인 예측 오차의 안정적 축소가 이루어짐
- 두 지표의 균형적 수준 유지로 일관된 예측 품질 확보가 확인됨

예측 그래프 분석

- 초반(0~200): 급격한 변동을 정확히 따라가며 부드러운 추적력 확보
- 중반(200~500): 완만한 흐름을 거의 동일하게 재현하며 오차가 거의 없음
- 후반(500~700): 변동성 재상승 국면의 피크 위치와 형태를 정확히 포착
- 전체적으로 실제 변동성과 유사한 흐름을 보이지만, 급등·급락 구간에서 다소 약한 반응이 나타나는 등 LSTM에 비해 추적력이 조금 떨어지는 모습이 확인됨

Results – Hybrid(Model 3)

모델	MAE	RMSE
Hybrid	0.000699	0.000949



성능 지표 분석

- MAE와 RMSE가 세 모델 중 가장 높게 나타남
- LSTM, LSTM-GARCH 대비 전반적인 예측 정확도가 가장 낮은 수준임

예측 그래프 분석

- 초반(0~200): 급등 구간의 큰 흐름은 따라가지만, 피크 근처에서 실제보다 크게 흔들리거나 작게 반응하는 모습이 자주 나타남
- 중반(200~500): 하락·횡보 구간을 전체적으로 잘 재현하지만, 세밀한 굴곡에서는 예측선의 변동폭이 실제보다 약간 더 큰 편
- 후반(500~700): 변동성 재상승 흐름은 잘 맞추지만, 피크의 높낮이나 작은 파형에서 차이가 비교적 자주 나타남
- 실제 변동성과 큰 흐름은 비슷하지만, 세부 파형에서 과대·과소 반응이 자주 나타나며 전체적인 추적력이 두 다른 모델보다 떨어짐

Conclusion & Implications

단순 LSTM 모델의 우수한 예측력 확인

- 가격 정보(ret)와 전일 실현 변동성(lag RV)만을 사용한 LSTM이 MAE, RMSE 모두에서 가장 낮은 오차를 기록하며 최고 성능을 보임
- 예측 그래프에서도 전체 구간에 걸쳐 실제 변동성과 가장 유사한 흐름을 나타냄
- 복잡한 Feature 없이도 변동성의 자기상관 구조와 시계열 패턴을 충분히 학습한 것으로 판단됨

GARCH Feature 추가 효과의 한계

- LSTM-GARCH 모델은 전반적 흐름은 잘 재현하나 오차가 LSTM 단독 대비 소폭 증가
- GARCH(2,2) 예측 변동성이 기존 입력과 높은 중복성을 가지며 모델 복잡성만 증가시키고 실질적인 예측력 개선은 제한적임

Hybrid(VIX 포함) 모델의 성능 저하

- VIX 추가 시 예측선이 매끄럽지만 세부 구간에서 실제 변동성과의 차이가 더 자주 발생
- 옵션 프리미엄·수급 요인이 섞여 있는 VIX가 실현 변동성 예측에는 오히려 노이즈로 작용하는 경향 확인
- 결과적으로 세 모델 중 MAE, RMSE가 가장 높음

결론

- 가장 단순한 LSTM이 가장 뛰어난 예측성능을 보임
- GARCH와 VIX의 정보는 단독 LSTM 기반 예측력을 실질적으로 향상시키지 못함
- 변동성 예측에서는 핵심 Feature(가격 수익률 + 최근 RV)가 이미 충분한 정보량을 제공하며 모델 단순성이 오히려 일반화 성능을 강화하는 결과로 이어짐

시사점

- 핵심 Feature(가격, 실현 변동성)만으로도 충분한 예측력 확보 가능함
- 불필요한 Feature(GARCH, VIX) 추가는 모델 복잡성만 증가시키고 예측력 개선은 제한적임
- 변동성 예측에서는 Feature 확장보다 모델의 단순성·안정성이 더 중요한 요소로 확인됨
- 향후 모델 확장은 모든 정보를 넣기보다는 상황별 선택적 Feature 활용이 효과적일 가능성이 높음

Effect of Scaling & Activation Choices : Additional Analysis

기존 Model과 Additional Analysis의 차이점

- 추가 분석에서는 target variable의 scaling을 진행하지 않음
 - Scaling하지 않는 경우 모델이 실제 값 단위에서 학습하므로 오차 감도가 커짐
- 금융 시계열(특히 변동성 예측)에 잘 맞는 손실함수로 Huber loss를 사용함
- 추가 분석에서는 output layer의 activation function으로 softplus를 사용함
- 추가 분석에서는 raw scale에서의 예측 능력 검증 + activation 변화의 영향을 테스트

Huber loss

- MAE의 강건함(이상치에 덜 민감)과 MSE의 매끄러운 경사(gradient 안정성)를 적절히 결합한 손실함수
- 작은 오차는 MSE처럼 미세하게 조정하며, 큰 오차는 MAE처럼 robust하게 처리

Softplus

- 항상 양수 출력을 보장하면서도 미분이 가능한 활성화 함수
- ReLU는 0에서 gradient가 사라져 작은 변동성을 예측할 때 불리
- Logistic-like shape 때문에 tail에서 더 크게 반응

Additional Analysis의 특징

- Target scaling을 하지 않음으로 인해 작은 값을 예측해야 함. 이는 Loss gradient의 축소를 의미
 - Softplus가 이를 상쇄하는 역할을 함. 그러나 학습이 저하될 가능성이 있음
- 실제 금융 데이터 적용 시 스케일링을 하지 않으면 해석 가능성이 좋아짐
 - 실제 변동성 값 단위로 비교 및 예측 가능
- 성능 비교 해석
 - Scaling + ReLU 조합 : 값이 균일하게 되어 예측이 수렴하기 쉬움
 - Non-scaling + Softplus : prediction variance가 커지고 spike를 더 잘 잡지만 noise에 민감

Results of Original Analysis

모델	MAE	RMSE
LSTM	0.000546	0.000791
LSTM_GARCH	0.000605	0.000839
Hybrid	0.000699	0.000949

Results of Additional Analysis

모델	MAE	RMSE
LSTM	0.000824	0.001115
LSTM_GARCH	0.001066	0.001426
Hybrid	0.000808	0.001058

➔ 작은 값 예측 난이도 증가로 인해
MAE와 RMSE가 소폭 상승