Appendix

Extending the KOSPI pipeline to predict daily excess

Introduction

This project extends the main project of S&P 500 pipeline to estimate and forecast the daily excess returns for the *Korea Composite Stock Price Index* (KOSPI index/KS11), which is a South Korea's equity benchmark. Here we wanted to show how our core model can be applied to a different asset class by creating features that are unique to KOSPI.

We selected KOSPI because it has trade-driven volatility (price movements) as opposed to the S&P 500 (stock market index) which is driven by technology stocks. Our plan is to evaluate the performance of our model through: sharpe ratio, cumulative returns, maximum draw-down and volatility ratio. Additionally, we will provide some qualitative insight about how KOSPI behaves in terms of market trends. We also hope to showcase how we have found another way to develop features and demonstrate the strength of a risk management approach. Finally, we aim to make our model reproducible and based upon sound logic.

Methodology

As project requires, the pipeline for KOSPI will be very similar to the one for S&P 500; however, it will be tailored to the specific KOSPI 's characteristics.

1. *Data Collection and Cleaning*: historical daily data was obtained from Yahoo Finance using the 'yfinance' API; data covers the time frame of Decebmer 4th 2019 to December 4th 2024 (equal to about 1200 trading days). This time frame is useful, because it includes several markets condition examples, such as unexpected behaviour due to COVID-19, and Trade Wars. After collecting, data was cleaned, all NaNs removed, to ensure that there were no missing values in the data set.
2. *Feature Engineering* 4 features:
   - *ret_5d* --> cumulative returns (total changes in prices) over the last five trading days (to show how short is life-cycle of index's stability)
   - *vol_5d* --> rolling volatility (standard deviation of returns) over those last 5 trading days (amount of risks investors take)
   - *vol_ratio* --> liquidity or in other words volume divided by its 5 day average
   - *momentum*: *ret_5d / (vol_5d + epsilon)* --> risk adjusted momentum of KOSPI, shows impact of trade volatility (one that used in exports etc.)

* visible improvements on prediction by raw prices. Returns that exceed a benchmark for us considered with a 0% risk-free rate and targeted to 3 days in the future to forecast.

3. *Model Architecture and Training:* Sequential LSTM was chosen as it is linked with time-series predictions. The layers logic goes as presented:

    1) *Input* - - > shape(1,4) – single timestamp with 4 features
    2) *LSTM* - - > 64 units – for temporal dependencies in overall time features, has no overhitting gradients
    3) *Dropout* - - > 0.2 rate – from overfitting (random deactivation)
    4) *Dense* - - > 16 units + ReLU – makes non-linear + mins dimensions
    5) *Output* - - > linear – pred excess return.
    *hyperparameters (64 units, 20 epochs, batch size 32) are good fit; Adam optimiser helps with MSE losses. Train/Test split 80, 10,10
4. *Weighting and Backtesting:* sigmoid function (0-2 range) for scale dynamic; returns = weights × actual excess returns. (points model more on correct predictions)
5. *Volatility Constraint*: if > than 1.2× benchmark(120) - > weights scaled down (risk control)

Results

Evaluation held on 20 split. (table with results do not fit in a 2 page criteria)

Plots showed that:

- losses in train/val go smooth with no overfitting;
- cumulative returns are a bit outperforming
- weights are balanced around 1.0
- risks align on drawdown and volatility
- our prediction is fine, but with weak correlation

Conclusion and Discussion

We found that our model predicts KOSPI returns fairly well, with a bit of overfitting, while generally showing adaptability. We can see that KOSPI's behavior is cyclical and depends on trading and external economic conditions. The graphs show that the model is good at predicting downturns but has difficulty during periods of consistent growth. The low Sharpe ratio indicates potential for improvement in the settings; in future work, macroeconomic data could be included to improve the ratio.