Report

Step 1

1(a)

We download daily prices for ["SPY", "TLT", "SHY", "GLD", "DBO"] between 2018-01-01 to 2022-12-30 for testing and 2010-01-01 to 2017-12-31 for training and validation.

We will be using "Adjusted Close" for the prices to ensure consistency.

1(b)

Here we evaluate the stationary of the log return. Before that we look at the summary of the various values, both in terms of prices as well as daily returns , followed by looking at the autocorrelation of the time series with its various lags. We present below the observations.

A. Summary:-

Summary Statistics for Prices:

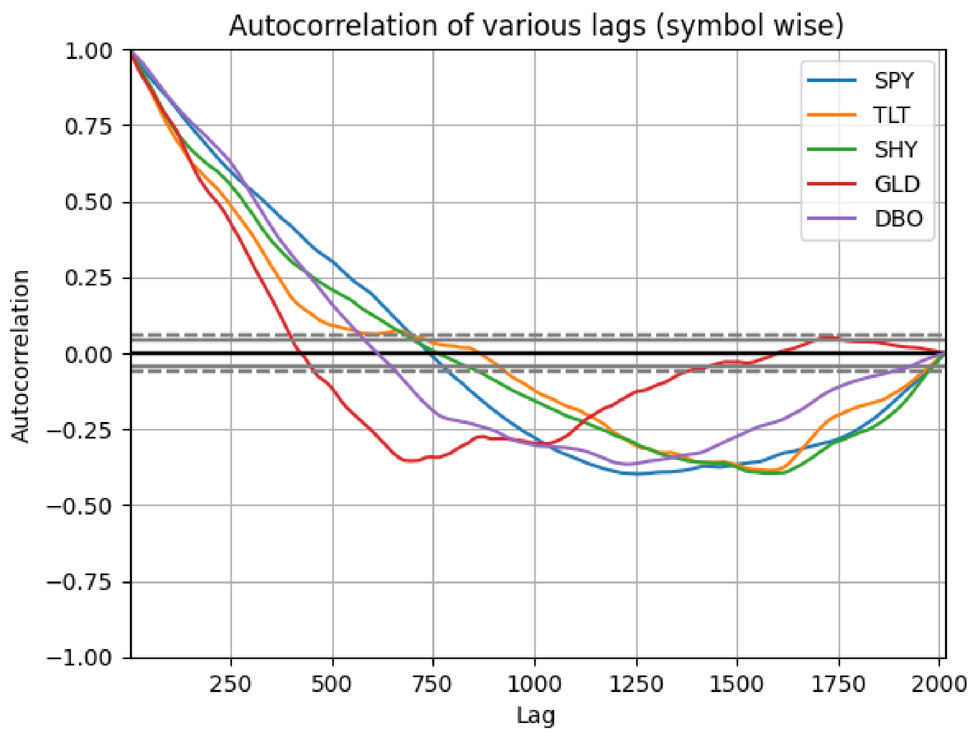
	DBO	GLD	SHY	SPY	TLT
count	2013.000000	2013.000000	2013.000000	2013.000000	2013.000000
mean	19.866224	130.406284	76.461696	148.316466	93.053119
std	8.268948	19.304866	1.132778	43.707799	15.174869
min	6.538252	100.500000	73.340599	79.874084	62.060909
25%	9.336240	116.940002	75.952644	106.352585	83.922028
50%	24.191534	123.389999	76.408325	151.392227	94.305405
75%	26.604918	142.639999	77.411011	181.209763	104.970505
max	32.941246	184.589996	78.350510	244.399261	123.215515

Summary Statistics for Daily Returns:

	DBO	GLD	SHY	SPY	TLT
count	2012.000000	2012.000000	2012.000000	2012.000000	2012.000000
mean	-0.000506	0.000059	0.000030	0.000506	0.000290
std	0.017450	0.010400	0.000545	0.009230	0.009081
min	-0.096989	-0.091905	-0.002766	-0.067340	-0.051766
25%	-0.009505	-0.005155	-0.000238	-0.003258	-0.005205
50%	0.000360	0.000394	0.000000	0.000635	0.000830
75%	0.008703	0.005540	0.000354	0.005059	0.005831
max	0.085298	0.047874	0.002386	0.045450	0.038889

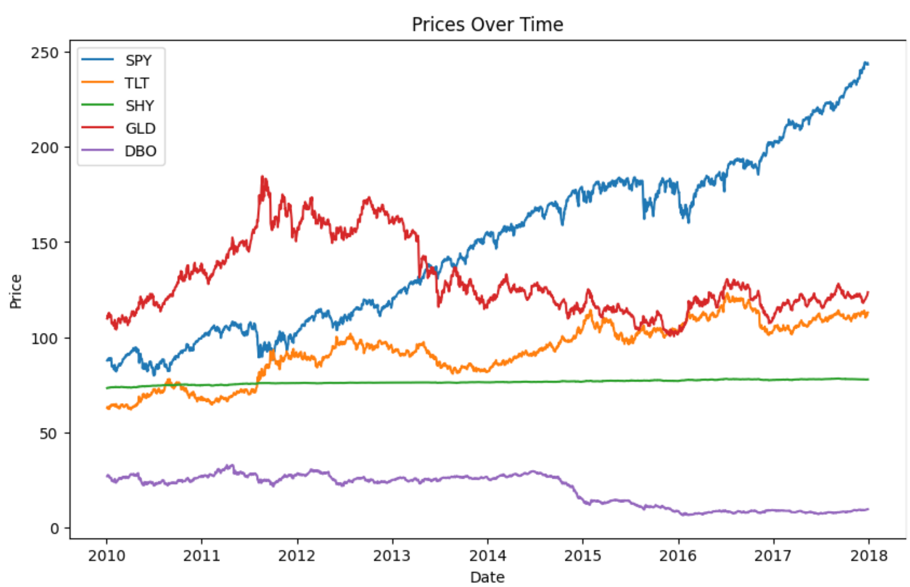
Note here that the count of daily returns is 1 less than the count of prices which is expected, next we see a very clear representation of the various ranges of the various time series in question.

B. Autocorrelation with lags.

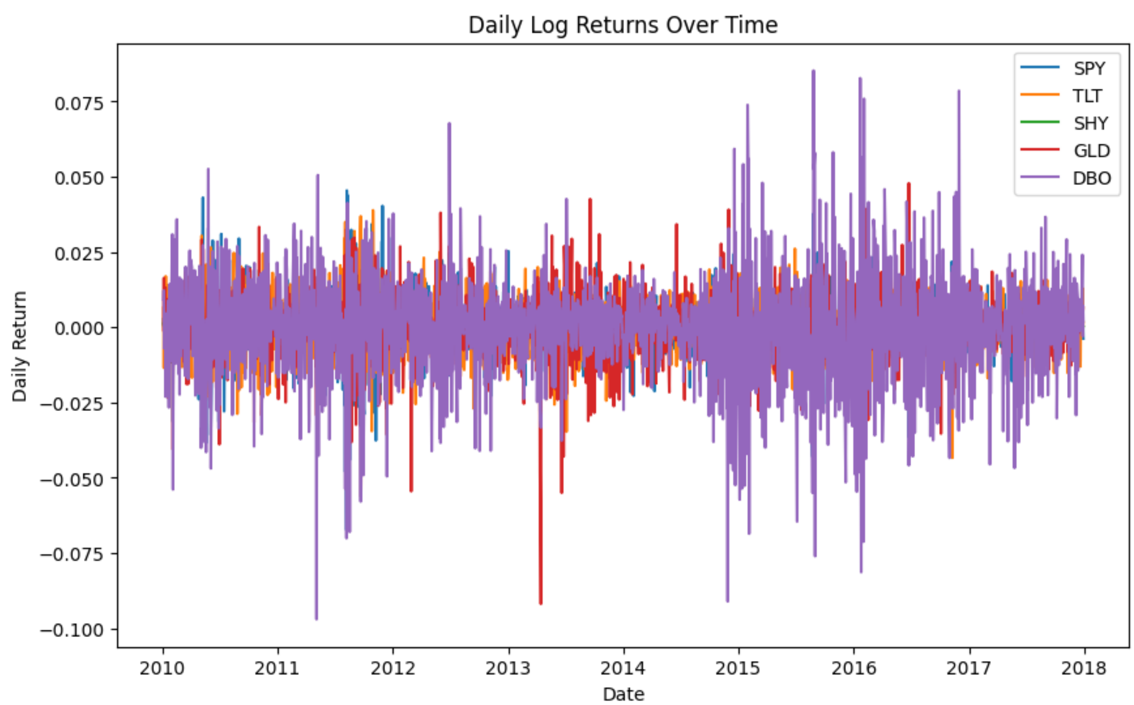


In the graph we can clearly see that the smaller lags are very much correlated , whereas as the lags increase the correlation decreases. It is also worthwhile to note that as the lags are increasing the number of sample points to get the autocorrelation is also decreasing , hence making those values less reliable.

C. Visualization of prices and returns over time.

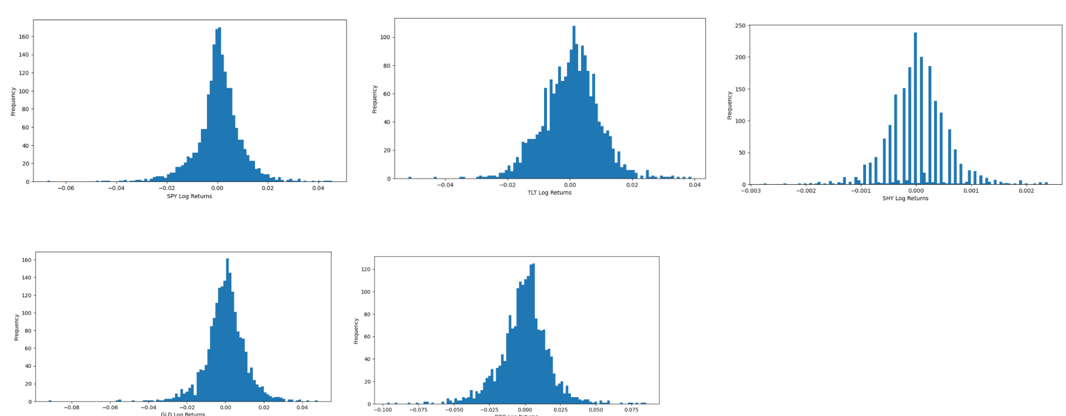


Here we see the price fluctuations of the five symbols in question. Here we can correlate the nature of these ETFs with the category they belong to , for example SHY or cash like treasury bonds tend to stay at the same price throughout the period which is expected since it is cashlike. DBO has weakened over the time maybe due to the coming of more green energy emerging. SPY has given good returns in line with the equity market. Similar numerous observations can be listed but we would leave the rest to the reader.



The log returns shows it is a mean reverting process with periods of high volatility followed by periods of low volatility.

D. Histogram of log returns

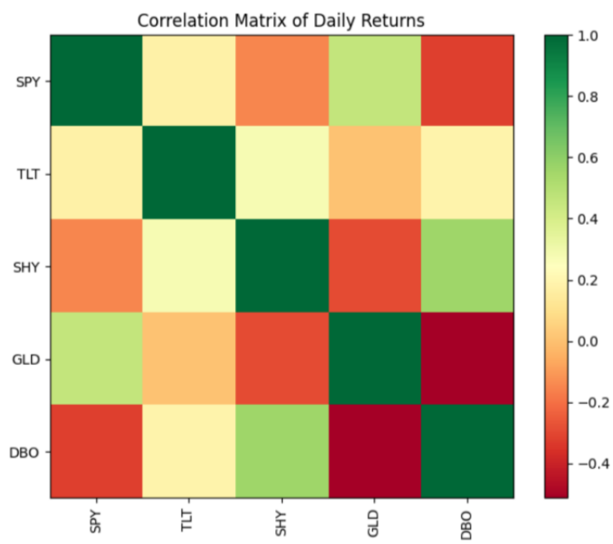


We have presented miniscule graphs because we just wanted to only highlight that the log returns follow a gaussian-like distribution from the graphs.

E. ADF test for stationarity.

We have received the p value of all the 5 series close to 0. Which means that the probability that null hypothesis will not be rejected is near to 0 . Hence the Null hypothesis will be rejected. Note that the Null hypothesis here is that Non Stationarity exists in the series. Hence we can safely say that all the 5 series are Stationary.

F. Correlation Matrix of Daily Returns



We see that correlation shouldn't be much of an issue in our study, Since no value is hauntingly close to 1 or -1.

Step 2

a) We use a LSTM architecture with 10 input channels and 64 nodes in the first layer , followed by a layer with 32 nodes and then a dense layer with 1 node. We use MSE loss and adams optimizer. The 10 input channels are nothing but lags =1 ,2,3…10 and the output we are expecting is the 25th lag of the time series.

b) The In sample losses are reported as

```
In-sample Loss for SPY: 9.824313019635156e-05
In-sample Loss for TLT: 9.69843240454793e-05
In-sample Loss for SHY: 2.972676895751647e-07
In-sample Loss for GLD: 0.00012503222387749702
In-sample Loss for DBO: 0.0003115987638011575
```

The In sample losses are pretty good with the best performance achieved in SHY and worst performance in DBO. The time series of SHY is simpler than the rest of the series hence it is expected that our model would be able to predict it well. For DBO it is not very clear why it is performing the worst.

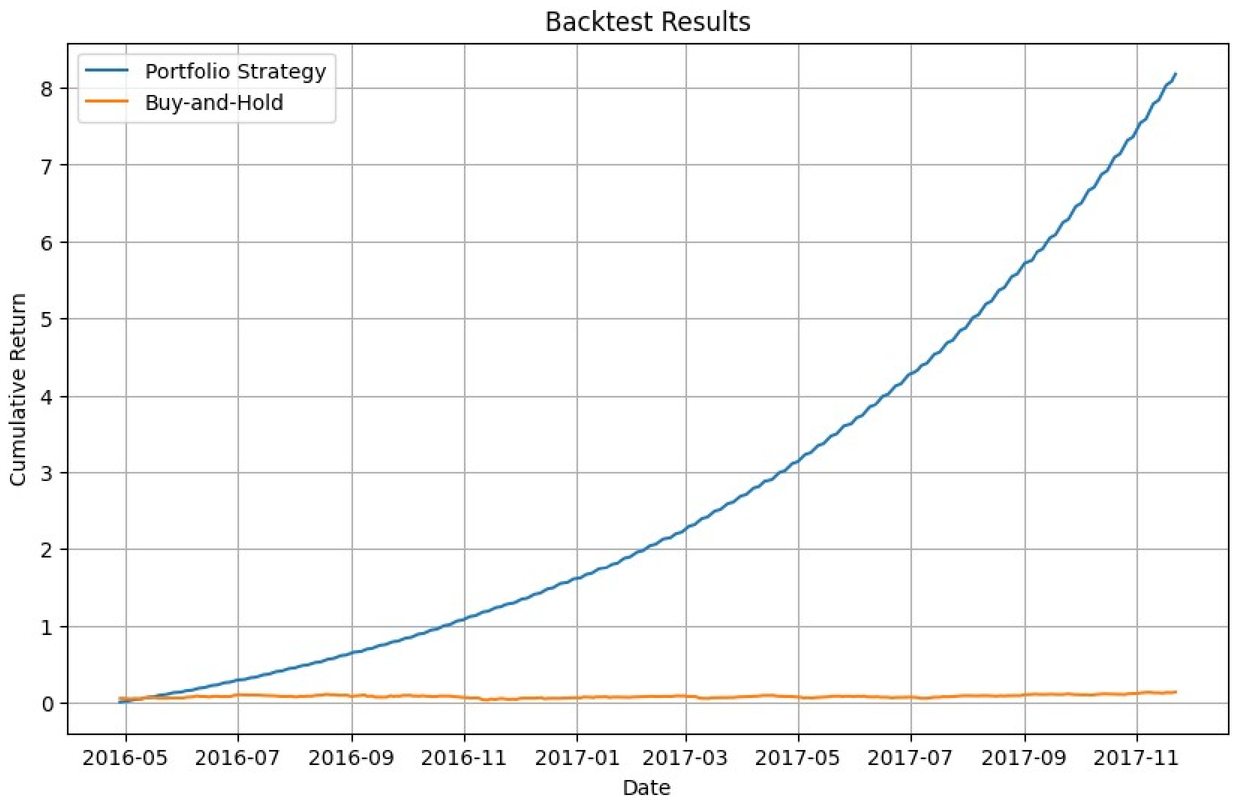
c) The out sample losses are reported as

```
Out-of-sample Loss for SPY: 2.928815047198441e-05
Out-of-sample Loss for TLT: 5.51321281818673e-05
Out-of-sample Loss for SHY: 3.217404298538895e-07
Out-of-sample Loss for GLD: 6.238381320144981e-05
Out-of-sample Loss for DBO: 0.0002796747721731663
```

The out sample losses are quite good and comparative to the in sample losses which suggests that our LSTM has generalized well. Here again the best is SHY and the worst performer is DBO. Another thing to note here is that out of sample loss of DBO is less than that of in sample. It might be because we have more sample points in the in sample as compared to the out sample.

d) Here we choose a strategy in which we go long for the best 2 performers among the 5 and go short on the worst 2 performers. We use this strategy to get our portfolio returns.

e) We back tested our strategy for out of sample performance and we got the following result.



All we can comment on is that the portfolio strategy appears to have generated positive cumulative returns over the entire backtesting period, increasing from an initial value to 8.180987. This suggests that the strategy, based on the trading signals, has been successful in generating profitable returns.

The buy-and-hold approach, using an equally weighted portfolio, also shows positive cumulative returns over the backtesting period, but with a lower final value of 0.152451. This indicates that the portfolio strategy has outperformed the buy-and-hold approach in terms of generating higher returns.

Step 3

Introduction

We use LSTM to forecast actual stock price 25 trading days into the future. Then we use this forecast to determine portfolio allocations.
First, we train LSTM model using data between 2018-2021. Then we start trading in 2022. During trading, we do not re-train the model.
Our allocation strategy is proportional to the forecast gain in the next 25 trading days while negative forecast will receive zero allocation.
We will try to fully deploy all cash whenever possible. So, this is a long-only strategy.
Our rebalance strategy is every 25 days. This means every 25 trading days, we look at our forecast and determine allocations.
Note that our backtest is simplied quite a bit by ignoring slippage, fee and utilize the same closing price for that day to determine current portfolio value and calculate number of shares based on new allocation.
In the future, we will want to incorporate negative allocation; that is short selling asset classes. We will then try a long-short strategy.

Implementation

We have created a complete class with the Backtest module inside of which we use a predictor module. In the predictor module we have created a model with 64 nodes in the first LSTM layer with input having multiple channels (as many as the symbols). We add another LSTM layer with 64 nodes followed by a dense layer with as many nodes as there are symbols. We get a training loss of 0.003725 and a test loss of 0.1203, which is actually good. Our strategy explained in the introduction is also implemented in the Backtest class itself. We made a return of 13.72% with the strategy using our predictions with a capital of 1 million USD, over a period of 1 year.

Comparison with Step2 and Buy and Hold Strategy
As compared to step 2 we got quite a good result because of the the combined LSTM where the LSTM nodes which learned information from returns of one security was able to utilize in the other. Hence the joint model is really helpful as opposed to the individual models. Had we used the buy and hold strategy , the returns would have been 10.8% , 0.7% , -3.8% , -20.6% and -34.78% respectively for DBO,GLD,SHY,SPY,TLT. Clearly we have beaten them individually as well as collectively on an equally weighted portfolio.

STEP 4

Comparison Analysis of the models

In multi-classes model, we can see model evaluation result in STEP 3 as follow:

train_loss 0.003718577791005373 test_loss 0.023464126512408257

This shows that trained loss was 0.003 while loss from test data was 0.02, which was significantly higher. Thus, the model is noticeably less performant with out-of-sample data. - possibly a sign of Overfitting? We might try to reduce the size of our Model and try to reduce the out of sample loss.

Using this model, however, still generates positive return in our trading strategy, at around 4-11%. Actual returns will probably be lower when slippage and fee are taken into account. The returns fluctuate as the model accuracy changes in each training.

However, if we consider the buy-and-hold strategy for these 5 assets, our model and trading strategy has done quite well since 3 of 5 assets had negative returns, GLD was virtually flat, while DBO gained 10%. Our portfolio gained upto 11% total.

We have not yet tried shorting asset class, which should give us an even higher return.

DBO 0.108640 GLD 0.007752 SHY -0.038404 SPY -0.206365 TLT -0.347868

Our Portfolio is up 11%

References

1. Shraddha Shekhar, "What is LSTM for Text Classification?" ,website , Analytics Vidya
   June 14 2021
2. Jason Brownlee , "Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras" ,website , Deep Learning for Time Series
   July 21 2016
3. Karthik Ram, "Algorithmic Trading - Backtesting a strategy in python", website , DevGenius
   January 3 2022
4. NUTHDANAI WANGPRATHAM, "Master the art of backtesting with Python: A step-by-step guide",website , DataDrivenInvestor
   December 26 2022
5. Shreyans Jain , "A Rookie Guide to Getting Started with Backtesting in Python!", website
   November 24 2021