Classification Model to Predict the Daily Direction of SPY

Introduction

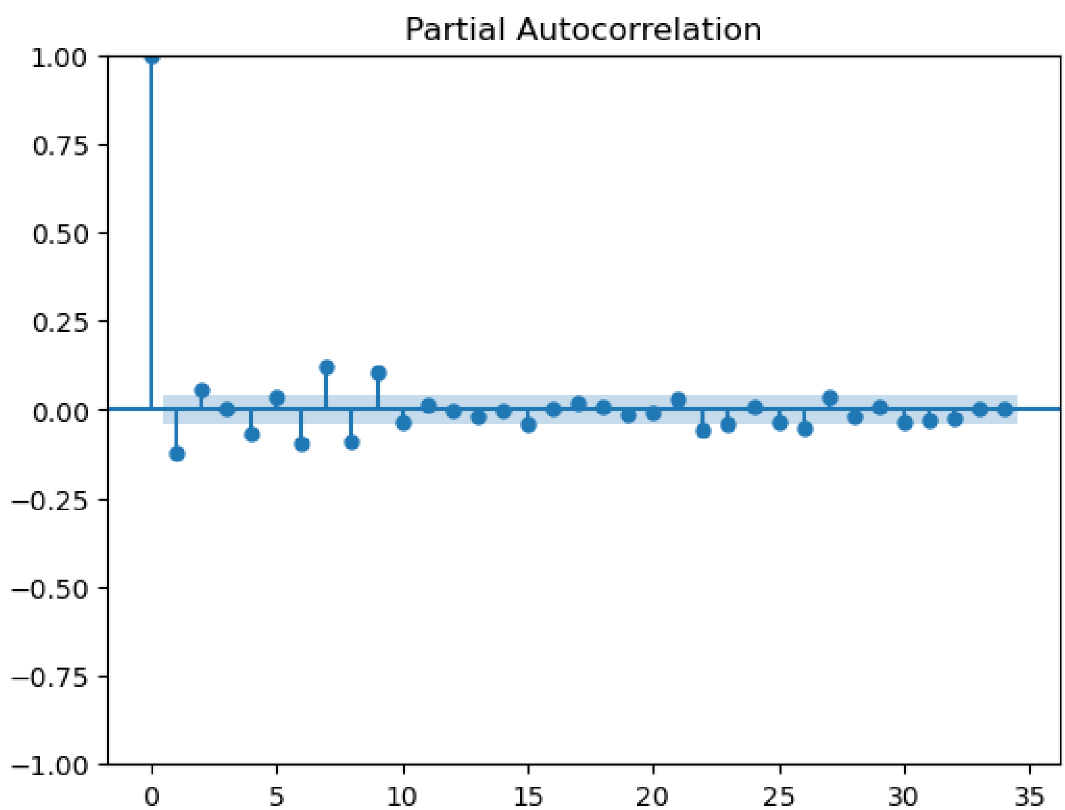
The objective of this project was to create a classification model to determine the direction (sign of return calculated using tomorrow's SPY close price and today's SPY close price) of SPY tomorrow given information today. Stated generally, the model would predict direction at period $t + 1$ given information at t. To set about this task we gained an introduction to this style of problem by reading Zhong and Enke (2019) who fit a neural network to their selected predictors to give classification responses on direction of SPY. Using this as a past precident aided in making decision regarding our model. Where this project differs from Zhong and Enke (2019) is that we include an assessment of the model as part of a trading strategy, albeit a very crude one.

To summarise the findings, we fit a random forest model to 43 predictors which are mostly the same as those in Zhong and Enke (2019). The model was trained on data from 2014-2022 inclusive and tested on data from 2023. This model achieves a testing accuracy of 56.36% and on a basic long only strategy, we obtain an expected monthly return of 3.02% by averaging results over backtests. Furthermore, a sharpe ratio of 0.69 and average maximum drawdown across backtests of 3.38% are obtained. However, none of these numbers can really be trusted as further inspection reveals the model almost never predicts a down day, only up.

Choice of Predictors

We mentioned that the choice of predictors was guided by Zhong and Enke (2019). Firstly, several types of SPY return are used. The daily SPY return at t, a handfull of lagged SPY returns amd various multi-day returns. Zhong and Enke (2019) use up to three lags. Instead, we choose lags according to their partial autocorrelation with period $t + 1$. Figure 1 shows the partial autocorrelation plot (with t not $t + 1$).

Figure 1: Partial Autocorrelation Plot of Daily SPY returns

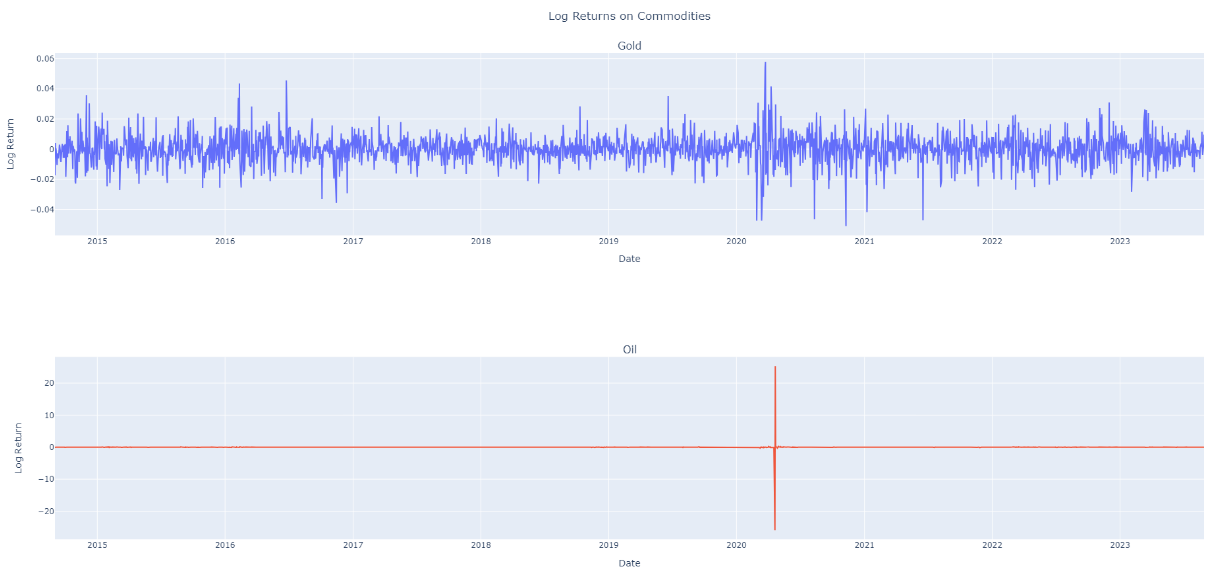


lags 1, 3, 5, 6, 7 and 8 were chosen. As for multi-day returns, these are the same as Zhong and Enke

(2019) at 5, 10, 15 and 20 day returns. Zhong and Enke (2019) include 10, 20, 50 and 200 day exponential moving averages of the SPY close price to guage trend. However, we use the same period simple moving averages with the SPY close price subtracted to achieve stationarity in the series ensuring a stable sample space across all time.

As for treasuries, we opt to get information across the entire yield curve including and above one year maturity, contrary to those chosen by Zhong and Enke (2019). All values are given as daily log changes in yield. For the returns on commodities we only selected gold as the plot of oil returns revealed hardly any variation aside from the flash crash when Covid 19 initially emerged. Figure 2 displays this.

Figure 2: Log Oil Returns



We include the log change in Moody's corporate bond yields (Aaa and Baa), the same as Zhong and Enke (2019). However, with regard to changes in exchange rates between major currencies and the dollar, we decide to use the dollar strength index (DXY) to summarise this information. The returns of the following major indices are included as predictors:

- HSI
- SSE
- FCHI
- FTSE
- DAX
- DJI
- NASDAQ

As are those of eight stocks with the largest market capitalisations in the S&P 500 according to investopedia as of the writing of this report:

- Apple
- Microsoft
- Amazon

- Nvidia
- Alphabet class A shares
- Tesla
- Alphabet class C shares
- Berkshire Hathaway

Model

A random forest was selected to use for predictions of daily SPY direction primarily for it's ability to handle non liner relationships, which mostly likely characterise that between daily SPY direction for $t + 1$ and the predictors. Two parameters were tuned via cross validation on the training data (2014-2022 inclusive) which were maximum depth per tree and loss function criterion. The number of predictors at each split was held fixed at the square root of the total number of predictors. The model was trained on a rolling five year period and tested on each succeeding year for every parameter combination. The best parameters were a maximum depth of one split and gini index criterion for the loss function, achieveing a cross validation accuracy of 54.15%.

Test Results

A model with the best parameters was trained on the entire training period of 2014-2022, this attained an accuracy of 56.36% when tested on data from 2023. The model was then applied to a simple strategy which bought SPY at today's close and sold at tomorrow's close if the model predicted up. This strategy was back-tested on a rollling monthly period (defined as 30 trading days) using the test data, giving a total of 5 periods. Table 1 gives the monthly returns on a \$100 account as well as the maximum drawdowns of each back-test:

Table 1: Back-test Results

Return	Maximum Drawdown
8.36%	2.47%
-3.05%	6.91%
2.09%	2.60%
4.82%	2.52%
2.88%	2.40%

From this an expected return of 3.02% and average maximum drawdown of 3.38% were calculated, which seem quite satisfactory for a month long period. However, a sharpe ratio of 0.69 was obtained, indicating a poor risk reward ratio. All key parameters are arranged in table 2, note that the monthly risk free rate for the Sharpe ratio calculation was taken as the effective federal funds rate on October 10th 2023 (5.33%) divided by 12:

Table 2: Estimated Parameter Values

Expected Return	Variance of Returns	Sharpe Ratio	Average Max Drawdown
3.02%	0.14%	0.69	3.38%

Considerations

The key question to ask given this analysis, is if such a strategy as the one proposed in the back-test can work. In short, no. The design of the back-tested strategy is made to accomodate the model, as daily direction is defined as the sign of SPY returns using close prices. In reality it is not possible to buy and sell at the close. For example, if one were to put a market order in after hours, it would be fulfilled at the next day's open. Therefore, one would not be trading exactly as the model instructs. Furthermore, closer inspection of the model predictions reveals that an overwhelming majority of the time it will predict an up direction. Despite no major class inbalace found in the data. Table 3 shows the counts of each class, where 'sideways' results were changed to 'up' due to how little there were:

Table 3: Estimated Parameter Values

Up	Down	Sideways
1216	1042	6

Given this, the model's performance is mainly deteremined by the number of up days in the data rather than detection of genuine signals. Thus, such a model is not fit for use.

Conclusion

The estimated test accuracy of the random forest model trained with best parameters is above 50%, indicating a possible trading edge. However, such a number cannot be trusted since the main driver of model performance appears to be characteristics of the data it is tested on rather than anything related to how good the training process was. Since practically speaking, the model only predicts 'up' directions. So, the more 'up' days one encounters, the better the model will be. In addition, a trading strategy that only requires one to hold a long position for a day's period everyday does not require machine learning to execute. Therefore, one need not give too much consideration to the results found in this analysis.