FINA4380 Project Report
Group2: Markehance

Name: SID:
Wong Chung Yu   1155126787
Li Wing Ho   1155108494
Poon Hoi Fung   1155142237
Kwok Chi Long Jason   1155126557

1.   Qualitative Ideas

The motivation of our project is to explore the potential of outperforming S&P500 using its own constituents. Therefore, the stocks in S&P500 is the target of our trading strategy.

S&P500 is well-diversified with different industries and therefore, to better predict the return of the stocks, we use factors from different asset classes to cover more aspects of variables in order to minimize the omitted factors. In our project, we used 97 different asset class indices which include interest rates, FX, commodities and so on.

We aim to use the weekly data of different factors to predict the future performance of the 500 stocks and come out with a weekly rebalancing portfolio. Such rebalancing frequency is to balance the trading cost and capture short term price movements.

2.   Methodology

Step 1: Principal Components Analysis (PCA)
Among all the 97 factors, we will use PCA to select the most relevant factors and change the basis of the data by using the first few principal components explaining 80% of variations. By considering the correlation between each factor and principal component, we set the maximum factor correlation in order to prevent duplicated factors. As a result, PCA provides the independent factors with high explanatory power for prediction.

Step 2: Support Vector Regression (SVR)
After selecting the factors, the SVR model is trained by trailing 100-week filtered factors' and stocks' returns for each week. Then, we will input the filtered factors of that week into the trained SVR model, in order to predict every stock's return and select the top 25 greatest expected return stocks as weekly components of the portfolio.

Step 3: Covariance Matrix Modelling
After running SVR, we need to forecast the covariance matrix for smart beta optimization. To account for serial correlation and conditional heteroscedasticity in stock returns, we fit in the univariate GARCH(1,1) models for returns of each stock then forecast their volatility one period ahead.

Next, we use the multivariate Constant Conditional Correlation (GARCH-CCC) model to compute the forecasted covariance matrix. The key idea is that we assume the conditional correlations between error terms of returns are constant throughout the model to ensure that the time-varying covariance matrices are positive definite by splitting it into variances and correlation as well as to enhance computational efficiency for covariance matrix modelling.

Step 4: Smart Beta
As we hypothesise that our selected stocks' returns on average will be higher than S&P500 index, we target to minimise the variance of the portfolio. To avoid portfolio weight heavily tilted to the lowest volatility stocks, we suggest that each stock weight be capped at 10% to diversify some of the idiosyncratic risks. We consider only an unleveraged long-only portfolio to avoid borrowing costs and shorting costs.

Step 5: Rebalancing
The above procedures will be repeated each week with weekly rebalancing as a means to reduce transaction costs and to respond readily to market change.

3.   Other Assumptions

We use a long-only strategy to avoid shorting costs. Also, we assume no slippage cost and 0% risk-free rate. Transaction cost will be 10bps, in and out, included when calculating the results. Regular weekly rebalancing is assumed to be done simultaneously. Our backtest period is 520 weeks (10 years), starting from Dec 2011.

4.   Backtesting Results

(i)   Backtesting focus: Machine Learning Models
The following graphs show the performance of different trading strategies over our backtesting period:

Drawdown Comparison
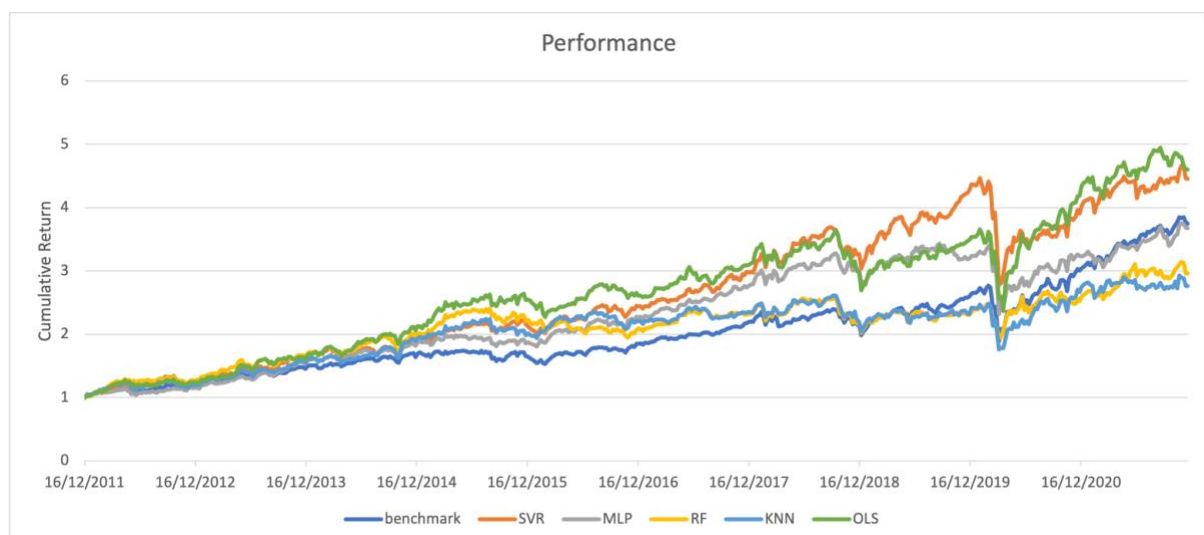
The following table shows the figures related to the benchmark portfolio and ML strategies:

	S&P 500	SVR	OLS	MLP	KNN	RF
Cumulative Return	2.74584	3.45843	3.60092	2.68016	1.76054	1.95826
Annualized Geometric Return	0.14160	0.16171	0.16538	0.13957	0.10719	0.11489
Annualized Volatility	0.14844	0.16334	0.17730	0.15154	0.16598	0.15201
Sharpe Ratio	0.96603	1.00033	0.95106	0.93860	0.69633	0.79096
Maximum Drawdown	-0.31810	-0.37977	-0.35290	-0.32499	-0.32769	-0.24680
Sortino Ratio	1.55916	1.79207	1.43274	1.56157	1.10630	1.42029
Calmar Ratio	0.44513	0.42580	0.46863	0.42947	0.32710	0.46553
VaR	0.02693	0.02995	0.03371	0.02740	0.03605	0.03654
cVaR	0.04994	0.05495	0.05742	0.04988	0.05673	0.05421

From the above results, we can see that OLS and SVR are performing relatively better than the other machine learning models. Comparing the return of the models over the backtesting period, the cumulative return and geometric return of SVR and OLS are higher than the S&P500. In terms of risk, both the SVR and OLS model results in higher volatility, VaR, cVaR and maximum drawdown. Overall, by considering the Sharpe ratio, Sortino ratio and Calmar ratio, the SVR model experiences a better performance than the OLS model and benchmark in general but slightly underperforms when the market suffers an economic downturn. As we assumed the probability of systematic loss happening is low, we believe that the loss of the SVR model can be recovered in the long trading period.

To study the efficacy of our smart beta scheme, we compare our backtest result with that of an equal weight scheme, another way to allocate assets. The following graphs show the performance of the two smart beta schemes:

	S&P 500	SVR-GMV	SVR-EW
Cumulative Return	2.74584	3.45843	4.79813
Annualized Geometric Return	0.14160	0.16171	0.19272
Annualized Volatility	0.14844	0.16334	0.18148
Sharpe Ratio	0.96603	1.00033	1.06157
Maximum Drawdown	-0.31810	-0.37977	-0.35449
Sortino Ratio	1.55916	1.79207	1.68851
Calmar Ratio	0.44513	0.42580	0.54365
VaR	0.02693	0.02995	0.03612
cVaR	0.04994	0.05495	0.06019

Despite our minimum variance portfolio having a lower Sharpe ratio than the equal weight scheme, our smart beta scheme still has its good side such as better Sortino ratio and low volatility and VaR, which means our Smart Beta Scheme do contribute to some of the portfolio benefits. And we did not expect an extreme abnormal drawdown during the week that our GARCH-CCC model cannot account for, such as the COVID outbreak period, thus as we suppose such an extreme event is not likely to occur in the future, we believe that our strategy has its edge for future investment.

5.   Limitation

As the trading strategy is long-only, there will be potential poor performance during an economic downturn. As a result, we may not be able to capture profits by short-selling during a market decline. However, we would still prefer choosing a long-only strategy because short-selling is executed with a high trading cost, or even may not be available.

Strategy performance could be dependent on testing parameters in PCA, Machine Learning and Optimisation. However, we would like to avoid overfitting with specific parameters and stay our focus on comparing different machine models, therefore we set all the models in default settings which may not be the most profitable strategy.

6.   Conclusion

In conclusion, our trading strategy yields a good performance with high returns and relatively moderate volatility, and it is a worth investing portfolio.