****

**BT4016 Risk Analytics for Financial Services**

**Semester 1 AY 20/21**

**Portfolio Risk Management Project**

**Final Report**

***Group 10***

**Prepared by:**

Chew En Ren Janson (A0183927R)

Ho Ying Rong (A0188843M)

Koh Kai Jie Bryan (A0183858L)

Lim Jermaine (A0187866E)

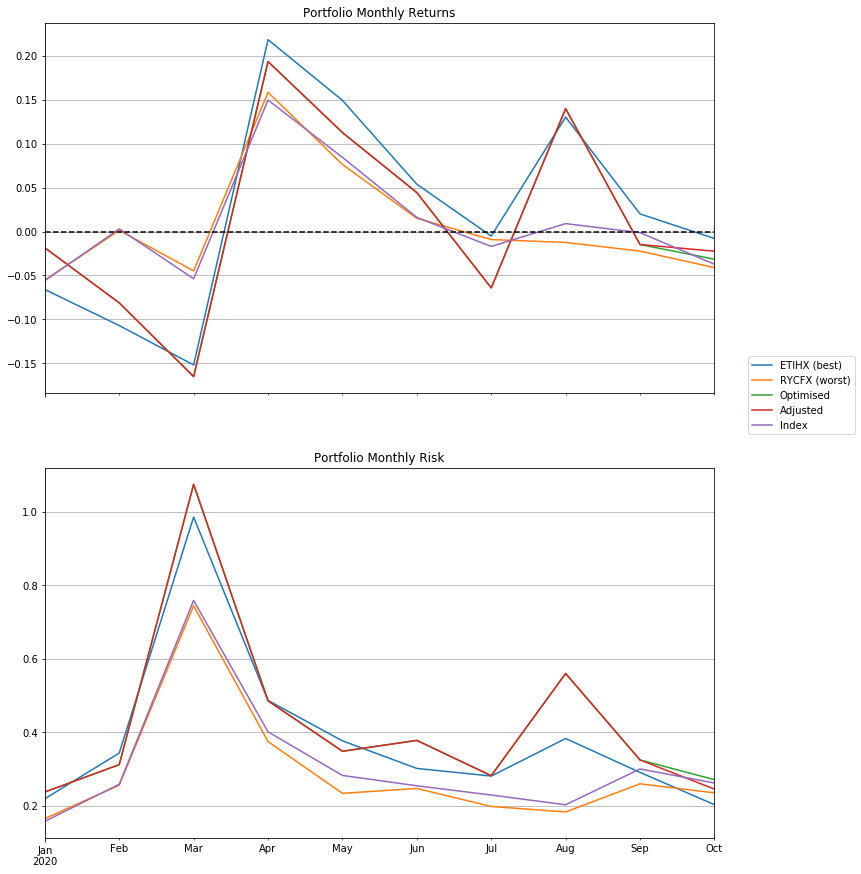
Timothy Lim Shyen En (A0167597H)

This report documents the methods our group applied to attempt to improve our portfolio’s performance. The report is structured in the following manner. First, we discuss the performance of the various portfolios from January to October 2020. Next, we explore using options as a cross-hedging strategy to limit the downside risk of our portfolio. Finally, we present a machine learning method to forecast future alpha values of assets and its applications to portfolio management.

**Portfolio Analysis**

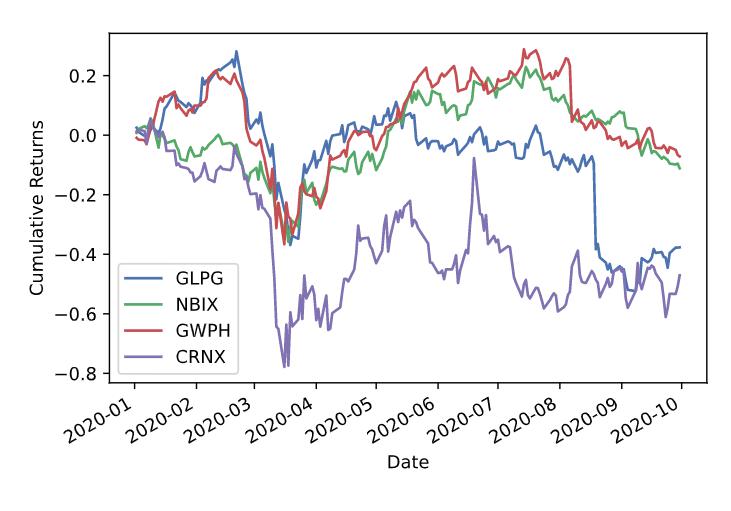
*Evaluation of risk & return*

To analyse the relationships between risk and return for our best and worst portfolios, we take a look at their performance in 2020.



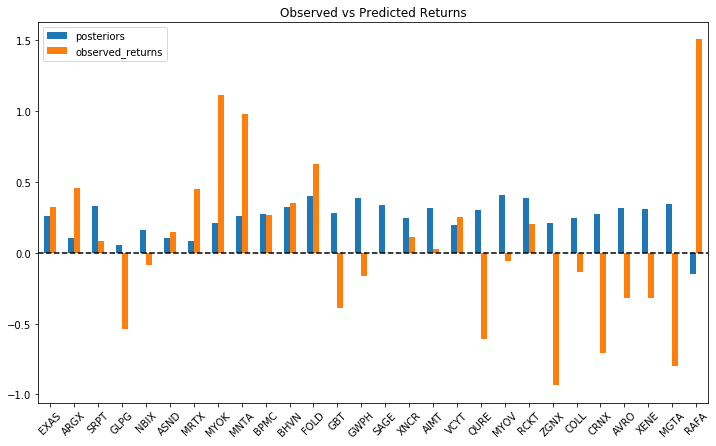
The above plot describes the monthly returns and risk of the various portfolios in 2020. The best portfolio arguably gave the best return, and our worst portfolio gave the lowest return. We observe that the best and optimised portfolio returns fluctuated wildly throughout, while the worst portfolio’s returns were consistent but generally poor. For the month of October, we can see that the adjusted portfolio returned marginally smaller losses than the optimal portfolio. Quantitatively, the optimised portfolio returned a loss of -3.15% in October, while the adjusted portfolio returned a loss of -2.24%.

The monthly risk was highest for the optimised portfolio but still similar to the best portfolio, with the worst portfolio consistently having the lowest risk among the 3 portfolios. In the month of October, we see that our adjustments reduced the amount of risk slightly, but it still had considerably higher risk than both the best and worst portfolio.

Both the returns and risk were consistent with our expectations on how the two mutual funds would perform. As expected, since RYCFX’s portfolio consisted of larger companies, they were able to have significantly smaller fluctuations in both risk and return during this pandemic compared to our best and optimised portfolio. The increased volatility for the optimised and adjusted portfolio from August to September could be due to the smaller biotech companies announcing more promising trials of the COVID-19 vaccine, which would attract speculative investors to gamble on their pending success.

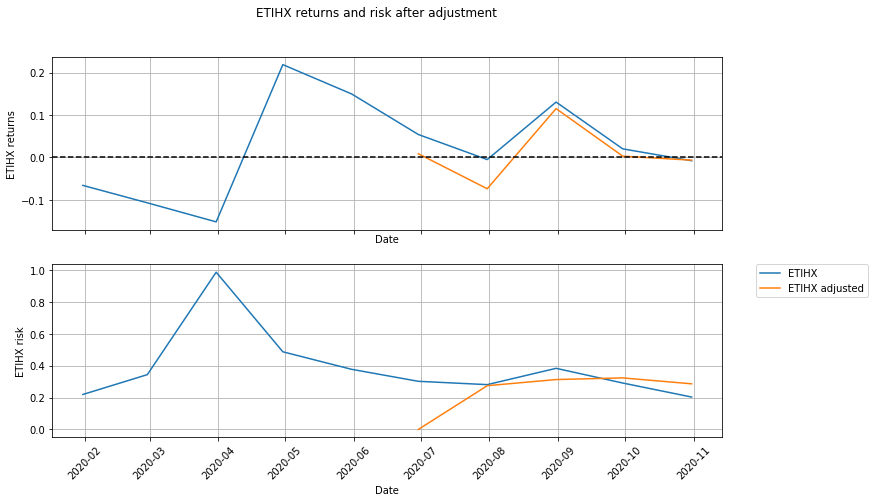
For our optimised portfolio, we increased the weights for GLPG and NBIX, namely because we expected a rising demand for the drug needed for patients with rheumatoid arthritis, and parkinson's disease, while we decreased the weights of GWPH and CRNX due to the uncertainty of the companies performance in this current pandemic. While our analysis of GWPH and CRNX was correct for the month of October as these assets returned losses, our analysis for GLPG and NBIX did not bear fruit as the price seemed to move sideways with no clear trend. These can be observed from the chart above.

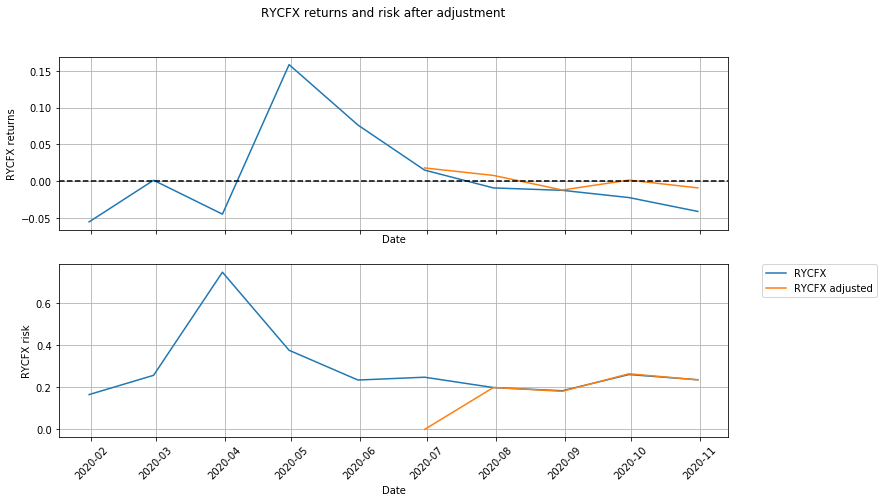
The performance of individual assets were also not consistent with our prior analysis in adjusting expected returns. The plot below compares the observed returns of the various assets in 2020 and the expected returns which we obtained from the Black-Litterman optimisation. We remark that the observed returns deviate greatly from the forecasts of expected returns. While there has undoubtedly been more volatility in the stock market this year due to the COVID-19 pandemic, it is also possible that our estimates were overly conservative.



It should also be noted that based on our estimates, only one of the assets were expected to return a loss in 2020. This shows the weaknesses in using analyst target prices as an investment outlook, as these views tend to be biased upward.

*Best / Worst Portfolio Adjustments*

******

**

The plots describe the returns and risk for our best and worst portfolios, after their weights were adjusted in June 2020. Visually analysing these graphs, it is difficult to form any meaningful conclusion as to whether the adjustment significantly affected the portfolio’s performance. For the best portfolio, the returns and risk plots are fairly similar despite having 10 new stocks introduced (and 10 removed) into the portfolio. For the worst portfolio, the adjustment of portfolio weights was not as drastic as none of the allocations changed by more than 1 percentage point. Similarly, only 6 new stocks were added to the portfolio, albeit being allocated a fairly small weightage. Graphs for the changes in weights can be found in Appendix A. These observations illustrate the different approaches to portfolio management when the composition of the portfolio leans toward larger, stable companies.

**Hedging**

*Value-at-Risk (VaR) analysis (Parametric Approach)*

We analyse our portfolio’s risk over the 3, 6 and 9 month horizons to gain an understanding of portfolio value that might be lost over these time horizons. To perform this analysis, we use the log returns of our hypothetical portfolio over 2019. Having checked that the normality assumptions for the log returns hold (see Appendix B), we proceed to estimate the 1 day VaR, expressed in simple returns, at 1% and 5% significance levels using the parametric approach. To scale the VaR to the respective time horizons, we multiply the 1 day VaR estimate by the square root of the respective time horizon. While this approach oversimplifies the longer horizon calculation of VaR, we observe that there does not appear to be any serial autocorrelation for lags up to 126 days at the 5% significance level (see Appendix C). The VaR estimates are reflected in the following table:

|  |  |  |
| --- | --- | --- |
| Horizon | Significance | Optimal Portfolio |
| 3 months | 1% | 25.3% |
| 5% | 18.4% |
| 6 months | 1% | 33.7% |
| 5% | 25.0% |
| 9 months | 1% | 39.6% |
| 5% | 29.7% |

In the midterm report, we had stated that Kris is willing to accept the VaR (5%) of his portfolio to be no greater than 20%. We extend this investor attitude to this report. Based on the VaR estimates, this threshold is first breached in the 6 month horizon. Thus, we will seek to hedge our portfolio in the 6 month horizon, limiting our portfolio loss to no more than 20%.

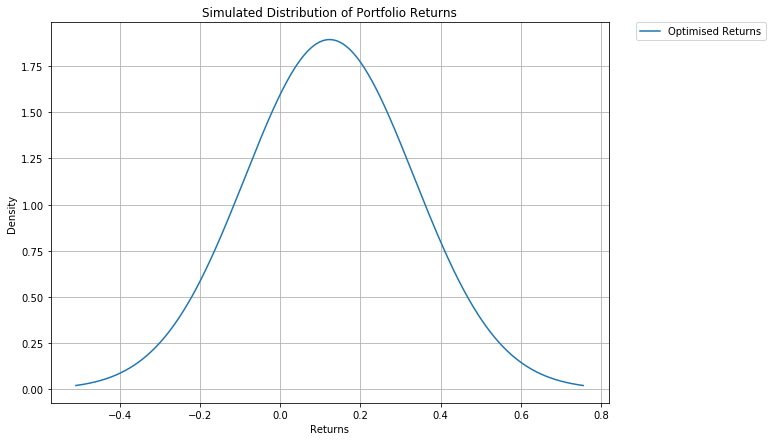
*Estimation of E(R) and Var(R)*

A weakness of the aforementioned method is that VaR estimates do not scale well for long-term estimates. Thus, we seek to calibrate our estimates for E(R) and Var(R) using the CAPM formula. We select the S&P 500 index (^GSPC) as a proxy for the market returns. To estimate beta, we regress our portfolio returns on the index returns. After which, we simply input the expected market returns (10%) and the risk free rate (2%) to obtain E(R).

To estimate Var(R), we first use the average implied volatility of at-the-money SPX options as an estimate of market volatility. Var(R) is then calculated by taking the square of market volatility multiplied by beta, divided by the r-square value to account for the amount of variance that is actually explained by the market returns. Our estimates for E(R) and Var(R) are 12.3% and 21.1% respectively.

After obtaining estimates for E(R) and Var(R), we proceed to simulate the distribution of our portfolio returns and obtain estimates for VaR at the 1% and 5% significance levels in a similar manner to the parametric approach taught in class. The values are reflected in the table below.

|  |  |
| --- | --- |
| Significance | Optimised Portfolio 1 year VaR |
|
| 1% | 38.8% |
| 5% | 24.4% |



While the VaR estimates are lower using this approach the losses at the 5% still exceed our investment attitude of tolerating a maximum 20% loss. In the next step, we describe how we attempted to use options to limit our downside risk for the adjusted portfolio.

*Options strategy*

First, we identified a suitable index to use as a cross hedge. Ideally, the index should have a high correlation with our portfolio so that any losses from the portfolio would be offset by a gain in the put options on the index. Among the various healthcare funds, we found that the **Fidelity MSCI Health Care Index ETF (FHLC)** had the highest r-square.

Next, we selected a suitable option strategy to take. The main options strategies considered were to employ a protective put, covered call or a collar. The protective put and collar strategy may be used to hedge the risk of a portfolio. However, a protective put strategy would be able to limit the downside risk of a portfolio without putting a cap on the profits earned. On the other hand, both the covered call strategy and the collar strategy cap the maximum profits that could be earned. Given our aggressive investment attitude, a protective put strategy would be best to help Kris achieve high returns while limiting some of his downside risk.

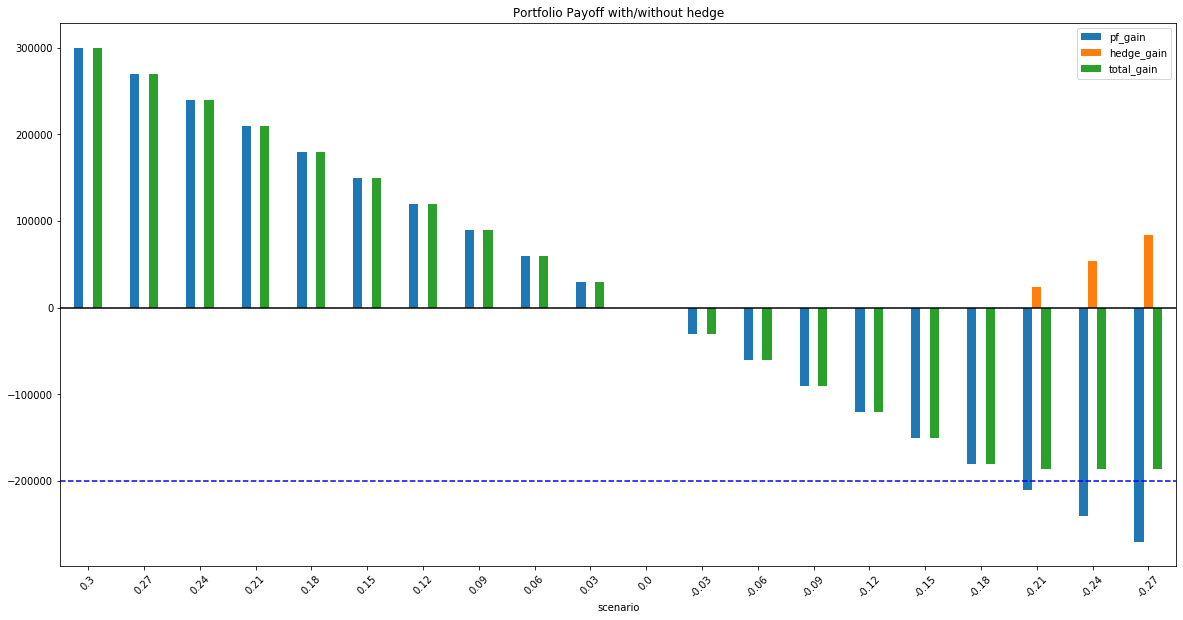
*Selection of strike price and calculation of weights to hedge*

Regressing our portfolio returns on FHLC returns, we obtained a beta coefficient of 1.29 and R-square of 0.72 . This means that a 1% movement in the market would approximately correspond to a 1.29% movement in our portfolio. Based on our target of limiting losses to 20%, we shortlist put options with strike price at close to 20 / 1.29 = 15.5% below the FHLC spot price on 31st December 2019. Using these criteria, we identify the put option expiring in June 2020 with a strike price of 42 as the put option to purchase.

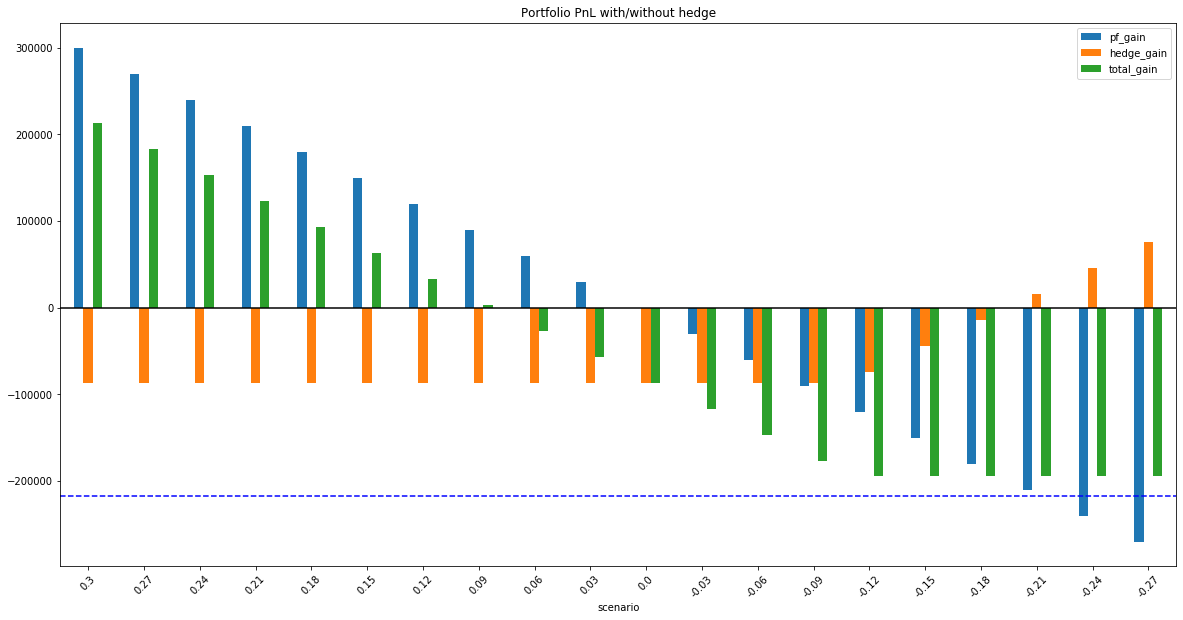
Assuming our portfolio’s initial value to be 1 million, we use the FHLC spot price on 31st December 2019 to calculate the number of shares required to cover our portfolio and thus the number of option contracts that should be purchased. This value is further scaled by beta so that the options position can offset potential portfolio losses.

*Analysis of hedge performance*

To analyse the performance of our hedge, we simulate a range of scenarios (shocks) to our portfolio value. In this analysis, we assume that the beta relationship holds and that the index moves accordingly. The plot below describes the portfolio payoff based on the various scenarios. As observed, portfolio losses greater than 20% are offset by gains from the puts, thereby limiting the entire portfolio loss to no greater than 20%.



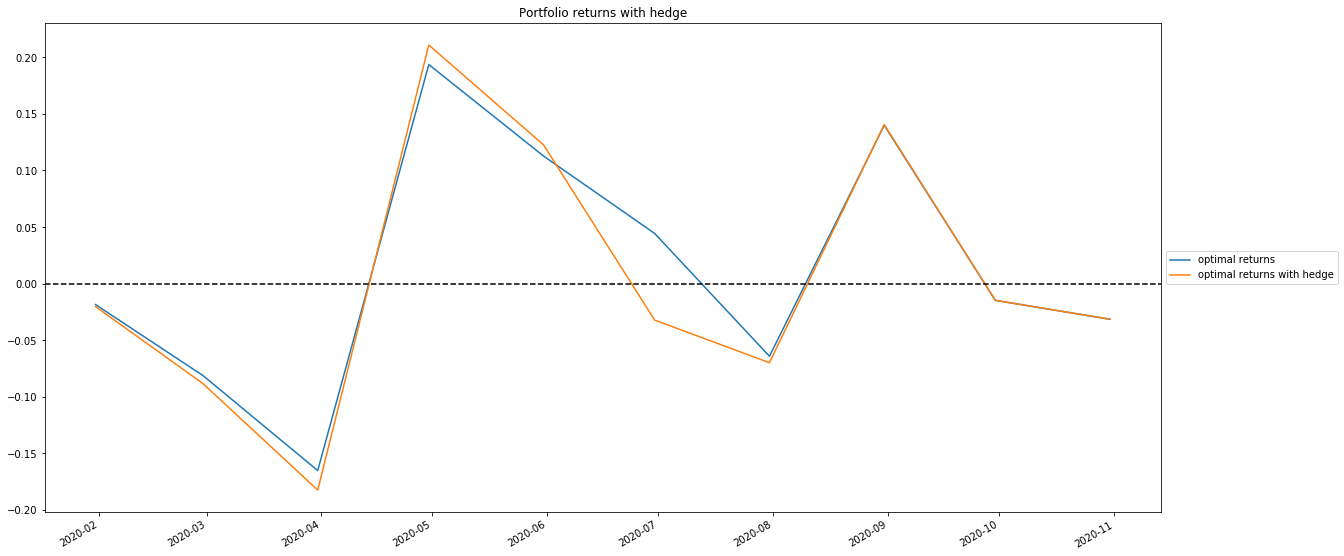
However, such an approach is unrealistic as it does not consider the premium paid to purchase the put options. We use the best offer listed in the table as the premium of the option contracts. Since the premiums listed are approximately 3, we adjust the strike price of the options purchased to 42 + 3 = 45. The premium for this option is 3.3. Based on this, we calculate the weight allocated to the hedge to be 7.99% (see Appendix D for details). The plot below describes the profits and losses (PnL) of the strategy based on a similar analysis as above.



While the maximum loss is still limited to roughly 20% of the initial portfolio value, we observe that the distribution of returns is highly asymmetrical around 0. If the portfolio loses less than 20% of its value and the options expire out of the money, our losses are amplified by the options’ position. Any portfolio gains are also muted for a similar reason. This shows that financing such a hedging strategy is expensive. Furthermore, allocating 7.99% of our portfolio to a hedge seems excessive when we consider that the expected returns are only around 12.3%.

To overcome this, we tried to use a collar options strategy by simultaneously writing the same number of calls as puts purchased. Rationalising that we were willing to forgo the top 25% of returns based on the simulated distribution, we calculated the equivalent price of FHLC in such a scenario to be 53.8. Thus, we chose the call option listed with the highest strike price at 51 to take a short position. In this strategy, we selected the put option at strike price of 47 so that the strike prices were ‘equidistant’ from the FHLC spot price. However, the strategy was unsuccessful as the premium collected from the calls was only 0.15 - a fraction of the premium of 3.3 paid for the puts. Thus, we decided not to pursue such a strategy.

*Portfolio Performance with Hedge*



In reality, the options expired out of the money as both FHLC and our portfolio had positive returns in June 2020. Thus, our portfolio actually had a net loss despite the stocks having positive returns. These findings emphasise the cost of financing a hedging strategy.

**Section 3: Stock Forecasting**

*Data preprocessing & feature engineering.*

The dependent variable in our problem is the monthly alpha of a given asset. This was calculated by taking a rolling window of the daily returns of a given asset and regressing it on the daily returns of the S&P 500 index for the same given month. The estimate for alpha would be the intercept term from the results of the regression. We chose to forecast the values for alpha as it would give us an understanding about whether the asset returns were simply driven by market returns or some other exogenous factor.

To train our model, we downloaded the historical data of over 700 stocks that were classified under the biotechnology industry according to yahoo finance. We reasoned that in only training the model on stocks from this industry, it would be able to pick up on the relationships unique to this industry between the feature variables and the dependent variable. Data was downloaded starting from 2018 to preprocess the features, and data from 2019 onward was used as training data.

In our XGBoost model, we made use of the closing price, daily returns, normalized volume and several common technical indicators as the features. The full list of indicators and their usefulness can be found in Appendix E. Data on ‘Open’, ‘High’, ‘Low’, and ‘Adj Close’ were not used since they can be approximated relatively well using EMA and other indicators. In addition, the volume was normalized because they typically correlate with price fluctuations.

To prepare the data for our model, we began by computing all the features required for each of the 758 stocks. The time series data is resampled by aggregating daily data to monthly, taking the last observation of each month. Individual stock’s dataframe is then combined into a single dataframe which is used as the dataset for our model. To make the alpha predictions for each month from January 2020 to October 2020, we use data from the past 12 months as our training and validation set, with inputs from month t-1. The train-test splits are not randomized here to avoid data leakage. The latest month is used for validation while the remaining 11 months are used for training our model. For instance, to predict the monthly alpha for January 2020, we use data from January 2019 to November 2019 to train our model and data in December 2019 to validate our model’s performance.

*Application to portfolio management*

In the midterm report, we used analyst target prices as input for our ‘views’ on the future performance of the various assets. We take a similar approach here to synthesise the predictions from our forecasting model with historical market-implied returns using the Black-litterman (BL) model.

The XGBoost model first outputs an estimate for the predicted alpha of an asset in the upcoming month. To generate a forecast (view) on the future returns of the asset, we first estimate the 1 year beta of the asset with the S&P 500 index. As an input to the CAPM equation, we use the average S&P 500 monthly return over the past 1 year, and the 1 month treasury yield rate as the risk free rate. Finally, we estimate the future returns of the risky assets using our model forecasts for alpha and the estimated beta to be input as ‘views’ in the BL model.

The BL model also allows users to specify their confidence about their views. Since this essentially describes our confidence in the model’s forecasts, we select a level of confidence based on the model’s performance on the validation set. We trained a classification model in parallel with the regression model to predict the sign of the alpha value. The performance of this classification model on the validation set was used as the level of confidence placed on the model forecasts.

*Results*

After we obtained the posteriors from the BL model, we ran the same optimisation process as we did for the midterm report to obtain portfolio weights. To benchmark the performance of our approach with the XGBoost model, we obtained analyst target prices in September 2020 as an alternative market outlook. After which, we compared the performance of the portfolios that resulted from these two optimisation processes.

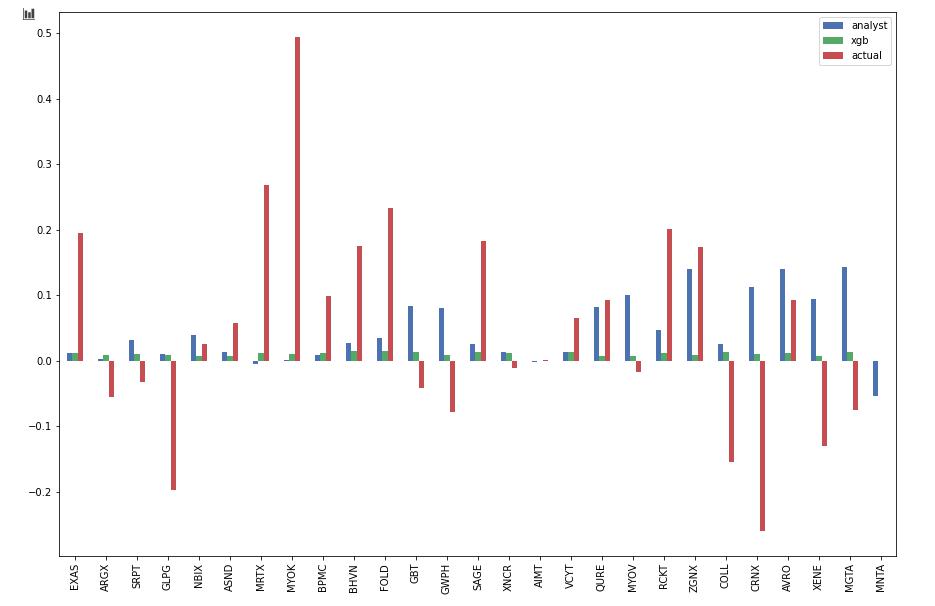
The table below describes the various metrics of the respective portfolio’s performance in October 2020. We remark that the approach using the analyst target prices appeared to give the best results. Although one month is not large enough of a sample size, the performance of the XGBoost portfolio suggests that there is considerable room for improvement.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Return** | **Standard Deviation** | **Sharpe Ratio** |
| Index | -3.54% | 2.62% | -1.35 |
| Analyst target prices | -3.12% | 2.47% | -1.26 |
| XGBoost predictions | -4.23% | 2.30% | -1.84 |

**Further Improvements and Conclusion**

In this project, we have demonstrated how an “optimal” portfolio may be constructed based on an investor’s personal preferences, risk appetite and other constraints. We have also shown how losses in excess of an investor’s risk tolerance may be limited through the use of options and the tradeoffs in terms of profits to protect one’s portfolio from such a loss. Lastly, we experimented with applying machine learning to the forecasting of an asset’s future performance and using the Black-Litterman model to apply these forecasts to portfolio management. As a guide for future work, we believe that the following areas could be explored further to improve the returns of the portfolio.

*Assumed Factors affecting returns*



Throughout this project, we have made the implicit assumption that the returns of our assets follow CAPM. While the model is widely used due to its simplicity, it is not without its drawbacks. Other factors, such as those prescribed by the various Fama-French models. Furthermore, the accuracy of the model may also depend on the selection of index as a proxy for the market. The S&P 500 is disproportionately weighted toward larger companies whereas the companies in our optimal portfolio tend to be medium to small cap companies. It may be useful to find a different index as a proxy for our market, especially when we compare the expected returns against the actual returns. We see that the expected returns using forecasted alphas were very small compared to the actual returns, suggesting that the S&P 500 may not be appropriate in this instance.

*Selection of Instrument for Hedging*

We used FHLC options to hedge our portfolio as it returned the highest R-square, indicating the strength of the relationship. However, the choice of options traded were somewhat limited as the deepest out of the money call was only $2 higher than the market price of the underlying asset. The bid-ask spread for the options was also very large (0.15 bid, 3.8 ask), suggesting that these contracts were relatively illiquid. This might have caused the estimates for setting up the hedge to be inflated. As an improvement, we could explore using other option contracts or financial instruments to hedge our portfolio.

*XGBoost*

In our XGBoost model, we made the assumption that all biotech stocks exhibit similar behaviour, which is why data for multiple stocks were combined to form a single training dataset. However, this generalisation may not necessarily hold true. To get better and more reliable predictions, we can possibly train multiple models for each stock using data within a longer time horizon. Each model can also be further fine-tuned to produce more accurate predictions.

**Tables**

Table 1: Monthly Returns

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **ETIHX (best)** | **RYCFX (worst)** | **Index** | **Optimised** | **Adjusted** |
| 31/1/2020 | -6.61% | -5.54% | -5.60% | -1.86% | - |
| 29/2/2020 | -10.71% | 0.13% | 0.30% | -8.13% | - |
| 31/3/2020 | -15.21% | -4.49% | -5.38% | -16.53% | - |
| 30/4/2020 | 21.86% | 15.86% | 14.97% | 19.36% | - |
| 31/5/2020 | 14.94% | 7.63% | 8.45% | 11.27% | - |
| 30/6/2020 | 5.39% | 1.50% | 1.61% | 4.43% | - |
| 31/7/2020 | -0.52% | -0.92% | -1.70% | -6.43% | - |
| 31/8/2020 | 13.04% | -1.25% | 0.90% | 13.99% | - |
| 30/9/2020 | 2.00% | -2.23% | -0.13% | -1.48% | - |
| 31/10/2020 | -0.79% | -4.12% | -3.73% | -3.15% | -2.24% |

Table 2: Monthly Volatility

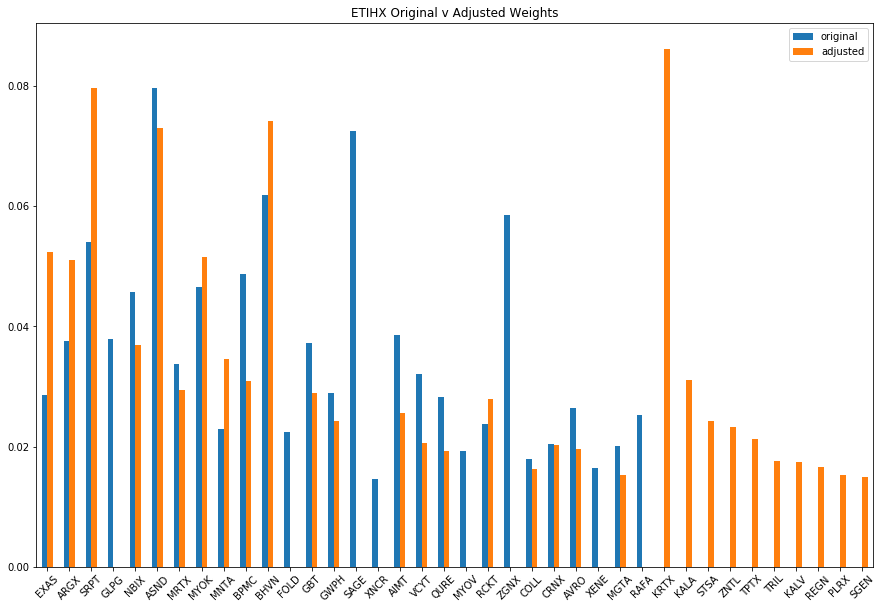
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **ETIHX (best)** | **RYCFX (worst)** | **Index** | **Optimised** | **Adjusted** |
| 31/1/2020 | 21.92% | 16.51% | 15.77% | 23.75% | - |
| 29/2/2020 | 34.36% | 25.62% | 25.86% | 31.21% | - |
| 31/3/2020 | 98.60% | 74.46% | 75.90% | 107.49% | - |
| 30/4/2020 | 48.69% | 37.48% | 40.14% | 48.57% | - |
| 31/5/2020 | 37.68% | 23.39% | 28.26% | 34.85% | - |
| 30/6/2020 | 30.18% | 24.73% | 25.45% | 37.80% | - |
| 31/7/2020 | 28.09% | 19.82% | 22.93% | 28.19% | - |
| 31/8/2020 | 38.31% | 18.33% | 20.28% | 56.02% | - |
| 30/9/2020 | 29.09% | 25.98% | 30.04% | 32.47% | - |
| 31/10/2020 | 20.31% | 23.52% | 26.20% | 27.11% | 24.54% |

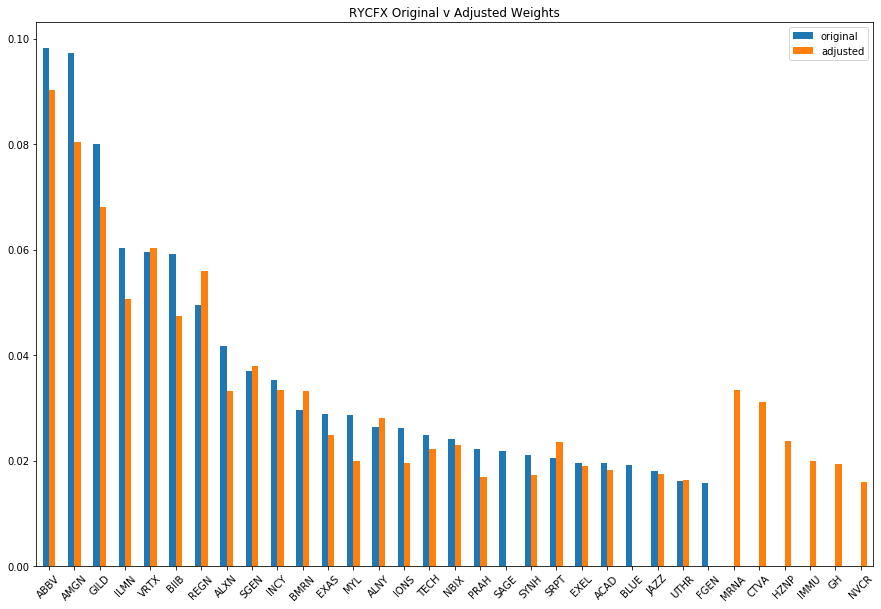
Table 3: Var / ES

|  |  |  |
| --- | --- | --- |
| **Portfolio** | **1-day VaR (5% significance)** | **1-day ES (5% significance)** |
| ETIHX (best) | 3.29% | 6.64% |
| RYCFX (worst) | 2.87% | 4.87% |
| Index | 2.56% | 5.02% |
| **Optimised** | 3.86% | 7.17% |

**Appendices**

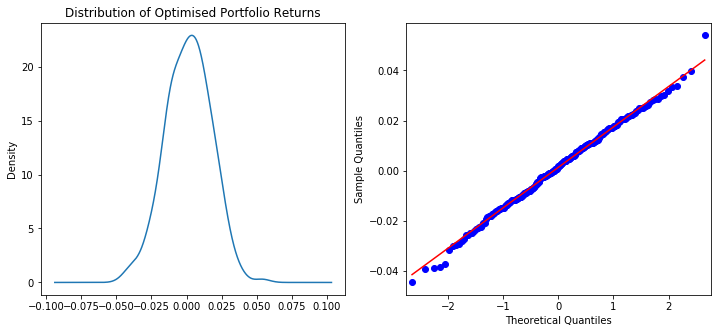
**Appendix A: Best / Worst portfolio adjustments**



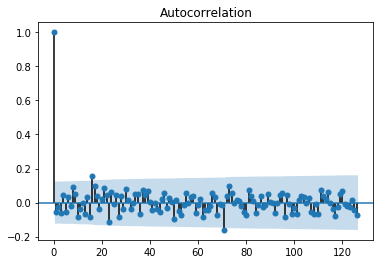


**Appendix B: testing normality of log returns**

The parametric approach of estimating VaR assumes normality of the return series. We test this assumption visually and mathematically. First, we plot the distribution of the returns and the Q-Q plot. The distribution of the returns appears to follow a normal distribution centered around 0. The Q-Q plot also generally does not suggest any deviation from the normal distribution.



We also performed the Shapiro test for normality, which tests the null hypothesis that the data is normally distributed. The test returned a p-value of 0.93. We conclude that the null hypothesis cannot be rejected and the normality assumption holds.

**Appendix C: checking for serial autocorrelation**

As mentioned by Joel, the multiplied volatility is usually modified by some autocorrelation factor (ACF) when scaling VaR to a longer horizon. In this plot, we analyse the ACF for the log returns of the optimised portfolio for lags up to 126 days (6 months). The blue region shows the 95% confidence level for the statistical significance of the ACF. Generally, the ACFs are not significant with the exception of lags around 18 and 70. Nonetheless, we make the assumption that results of the multiplied VaR would not be too different than if we had adjusted it with the ACF.

**Appendix D: calculation of hedge cost**

First, we estimate the number of contracts that are required for a perfect hedge.

For a cross hedge, we simply multiply the contracts required by the estimated beta coefficient obtained from OLS regression. After that, it is trivial to calculate the cost of the hedge and express it as a percentage of the total portfolio.

**Appendix E: Model Features**

1. Exponential Moving Average 'EMA'
   1. Describes the average price of the asset over the past window with more weightage to more recent observations
2. Moving Average Convergence Divergence 'MACD'
   1. Describes the relationship between the long term and short term relationship of the asset price. MACD is calculated as the difference between the longer period and shorter period moving averages
3. Market Momentum 'MOM'
   1. Describes the ‘ability’ of the market to sustain the current trend of the price movements. MOM is calculated by taking the difference between the most recent close price and the close price X days ago
4. Rate-of-Change 'ROC'
   1. Describes the percentage change in the asset price over the period of time. It can also be thought of as a momentum indicator. ROC is calculated by comparing the current price of an asset as a percentage of the price from an earlier period
5. Relative Strength Index 'RSI'
   1. Describes if an asset is ‘overbought’ or ‘oversold’. It can also be thought of as a momentum indicator.
6. Average True Range 'ATR'
   1. Describes the range of the asset prices over a period of time. It can be thought of as an indicator of asset volatility.
7. Vortex Indicator 'VORTEX'
   1. Typically used to spot trend reversals and confirm current trends.
8. Accumulation Distribution Indicator 'ADL'
   1. Cumulative indicator that provides insights on the strength of a trend.
9. Stochastic Oscillator %K 'STOCH'
   1. Describes the closing price of an asset relative to the range of prices over a period of time. Used as a momentum indicator
10. Money Flow Index 'MFI'
    1. Describes the flow in and out of an asset over a period of time. Used as a momentum indicator
11. On Balance Volume 'OBV'
    1. Cumulative sum of volume on days where the stock closed higher, minus the cumulative sum of volume on days where the stock closed lower. Used as a momentum indicator.
12. Commodity Channel Index 'CCI'
    1. Describes the relation between current price and the average price over a time period. Used to identify trend