# Quant Take Home Assignment 2025

Abil Nurgaliyev

December 3, 2025

## 1 Overview

This is the final report for the BCI Quant Equity take home assignment. As instructed, I have included the code with which I produced the figure that will be shown in this report.

## 2 Q1. Normality

I used yfinance to obtain the closing price of AAPL from Jan 1st 2012 to Dec 31 of 2021. Next I generated 1, a plot displaying the daily returns. To test the normality assumption, we can compute the mean ($\mu$), and standard deviation ($\sigma$) of the data, and plot $\mathcal{N}(\mu, \sigma)$ over the histogram of the data. This is seen in 2. Clearly the normal distribution does not fit the data well. It fails to capture the peak and heavy tails of the data.
Another popular method for testing normality is plotting a Quintile to Quintile (Q-Q) 3 plot with respect to a normal distribution. If returns followed a normal distribution, one would see the data on a $45°$ straight line. This is not the case; in fact we observe a significant deviation in the tails.
We can say with certainty that daily returns are not normally distributed.

## 3 Q2. OLS

Since the closing price between consecutive days is usually very close. We can rephrase this as

$$P_t \sim P_{t-1} + \epsilon_t, \epsilon_t << 1$$

As the closing price on day $t$ is going to be very close to the closing price on day $t-1$. If we use OLS, we will end up with $a \simeq 0, b \simeq 1$. This can be seen empirically in 4. All the model has predicted is that $P_t \simeq P_{t-1}$, not very informative for forecasting purposes. For these reasons, OLS is not a good model for predicting the future price of AAPL. Although, it is a good way for finding the parameters $a$ and $b$ for illustrative purposes.

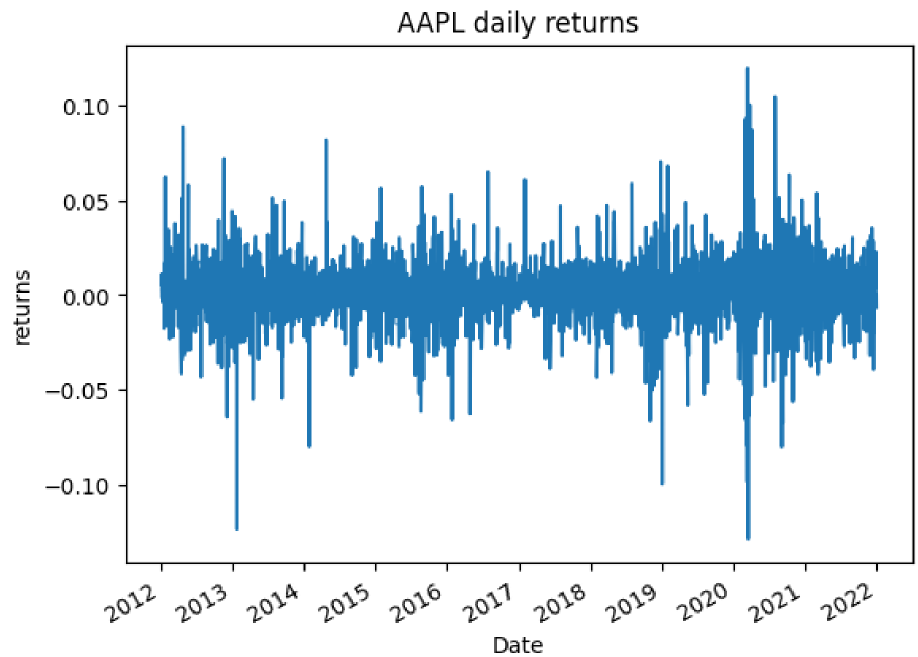


Figure 1: Daily returns of AAPL from 2012-2021

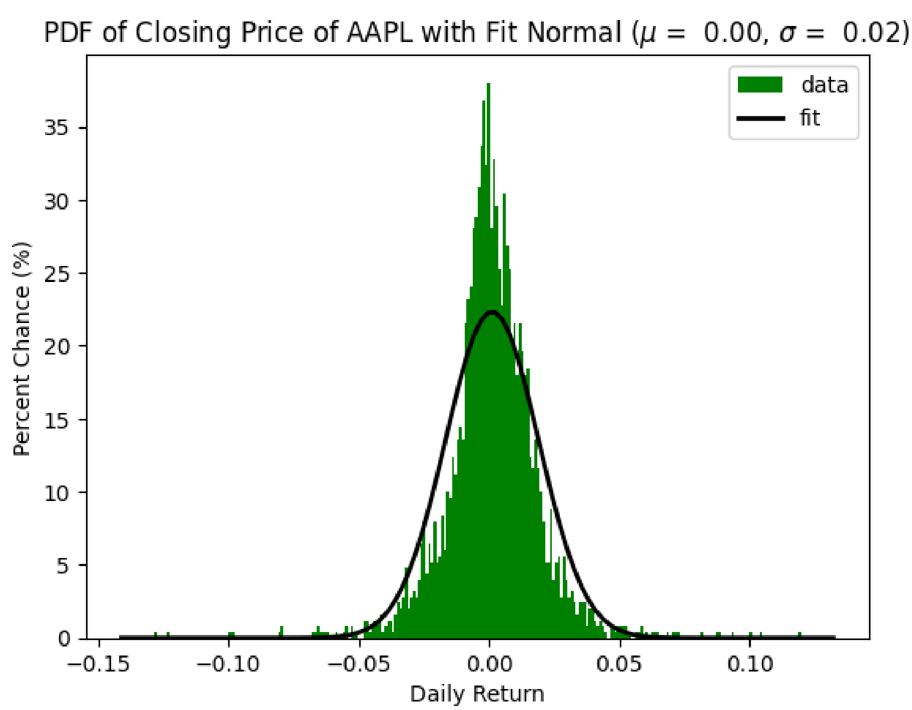


Figure 2: Histogram of Daily Returns of AAPL, with a fit Normal distribution

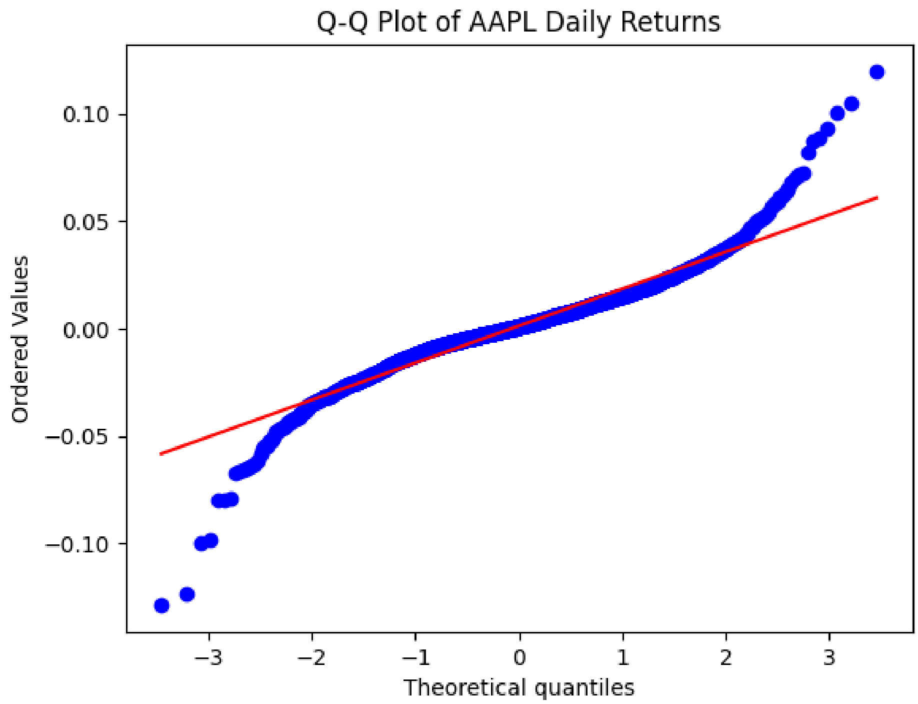


Figure 3: Q-Q plot of AAPL

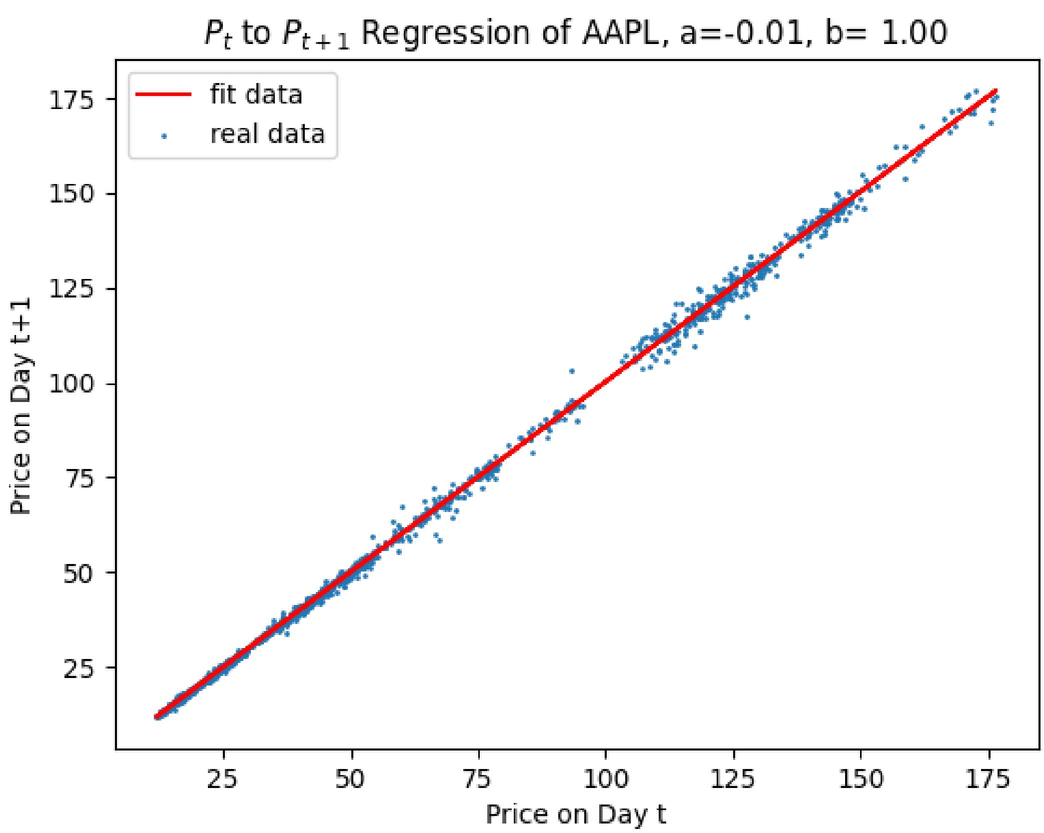


Figure 4: OLS with daily price of AAPL

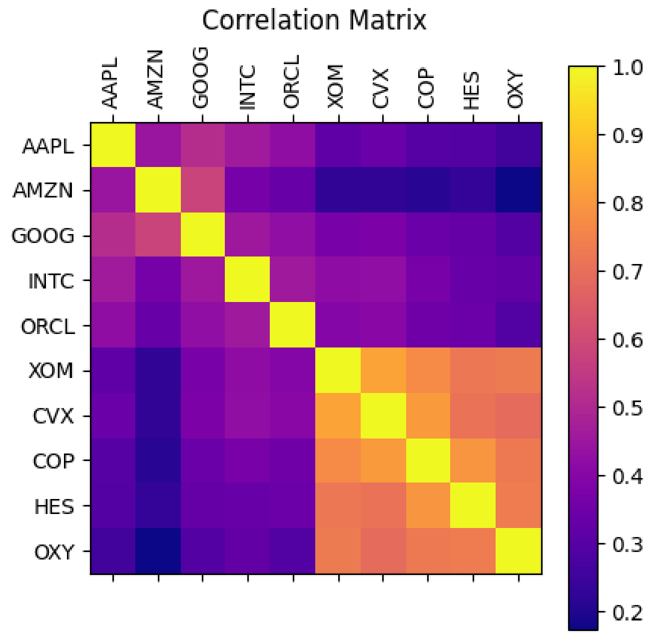


Figure 5: Correlation Matrix

## 4 Q3. Correlations

The plot of the correlation matrix can be found in 5. As expected the correlation between the energy stocks (OXY, HES, COP, CVX, XOM) are high ($> 0.7$), indicating that they tend to move together. In contrast, the correlation between the energy and tech stock has close to no significance ($< 0.4$), suggesting little co-movement between the sectors. The correlation among the tech stocks is moderate ($\sim 0.5$), suggesting some sector specific co-movement but not to the degree of the energy stocks. In general, these patterns align well with the expectation that stocks in the same sector are more closely correlated than stocks across sectors.

## 5 Q4. K-means

To investigate whether the 10 stocks exhibited similar behavior during the pandemic, I computed quarterly returns for each stock from Q2 2020 to Q4 2021. Each stock is represented by a 7-dimensional feature vector corresponding to its 7 quarterly returns over this period. I then applied K-means clustering with $k = 2$ to identify groups of stocks exhibiting similar performance patterns. The resulting cluster assignments are shown in the table below.

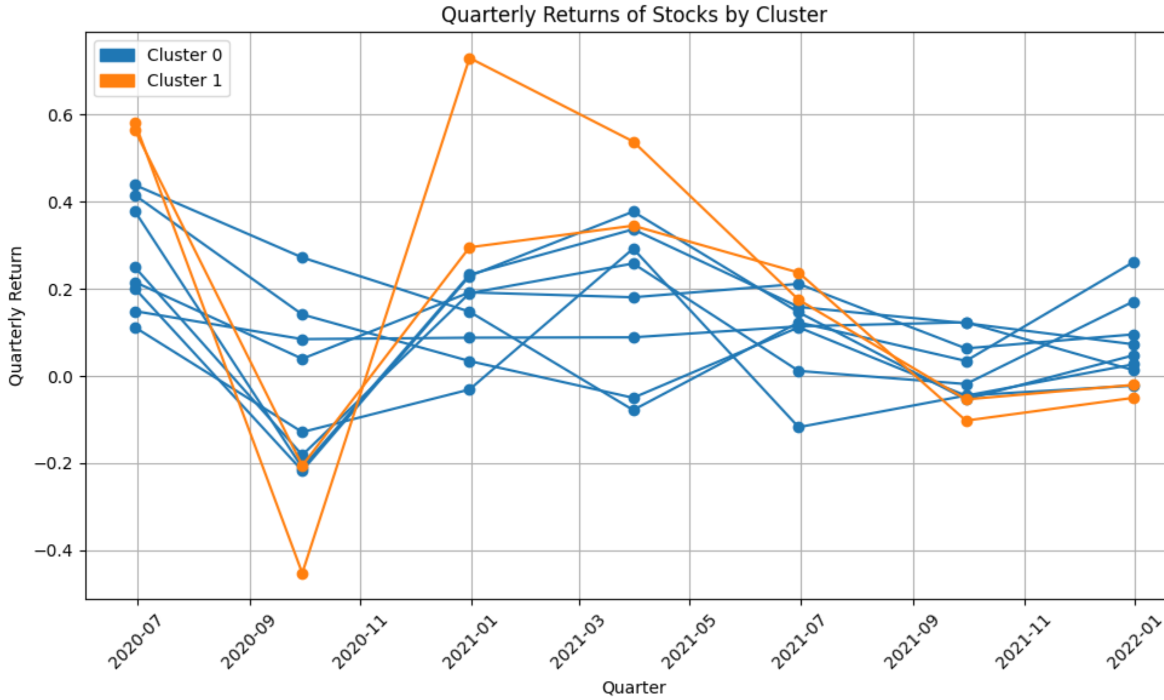


Figure 6: Quarterly Returns, seperated by clusters

| Ticker | Group |
|---|---|
| AAPL | 0 |
| AMZN | 0 |
| GOOG | 0 |
| INTC | 0 |
| ORCL | 0 |
| XOM | 0 |
| CVX | 0 |
| COP | 0 |
| HES | 1 |
| OXY | 1 |

The clustering separates the stocks into two groups. Group 0 consists of technology and large-cap energy companies (AAPL, AMZN, GOOG, INTC, ORCL, XOM, CVX, COM), which generally exhibited relatively stable performance during the pandemic. Group 1 contains smaller, more volatile energy stocks (HES, OXY), which experienced larger swings in quarterly returns. This confirms our hypothesis that the performance of these 10 stocks differed across two groups during the pandemic, reflecting company-specific sensitivities to COVID-19. The plot of the quarterly returns is show in 6.

## 6 Q5. GMVP

Using numerical optimization via `scipy.optimize.minimize`, I obtain the GMVP weights, shown in the table below.

| Ticker | Optimal Weight (%) |
|---|---|
| AAPL | 13.01 |
| AMZN | 21.76 |
| GOOG | 4.66 |
| INTC | 11.89 |
| ORCL | 15.80 |
| XOM | 0.0 |
| CVX | 0.04 |
| COP | 6.26 |
| HES | 7.30 |
| OXY | 19.28 |

The Global Minimum Variance Portfolio (GMVP) allocates weights to minimize total portfolio risk, given the constraints that all weights are non-negative and sum to 100%. Moreover, if a weight turned out to be less than $10^{-12}$ it was set to 0 to offset floating point errors.
As you can see, the allocation favors moderate-volatility stocks like AAPL, AMZN, ORCL, and OXY, while assigning near-zero weights to very high-volatility or highly correlated stocks like XOM and CVX. This reflects the GMVP's focus on risk minimization over expected return, concentrating investments in assets that reduce overall portfolio variance.
Overall, the GMVP provides a risk focused allocation across the 10 stocks.

## 7 Q6. Predictive Model for Stock Returns

To estimate future stock returns, I constructed a predictive model for the 10 stocks using publicly available data. The features included:

- Recent return history: daily return, 5-day, 10-day, and 20-day moving averages
- Volatility: 20-day rolling standard deviation of returns
- Fama-French 5 factors: Mkt-RF, SMB, HML, RMW, CMA

The target variable was the next-day excess return, defined as the return of the stock minus the risk-free rate.

In order to test the results of the model, I created a train-test split, with the model training from 2012-2017 and test between 2017-2021. I employed an **XG-Boost Regressor** with the following hyperparameters: 500 trees, max depth 4, learning rate 0.05, and subsampling and column sampling of 0.8.

#### Results

The predictive performance was evaluated using directional accuracy (correct prediction of return sign) and $R^2$ scores. Table below summarizes the results:

| Ticker | Directional Accuracy | $R^2$ Score |
|---|---|---|
| AAPL | 0.512 | -0.113 |
| AMZN | 0.485 | -0.174 |
| GOOG | 0.523 | -0.118 |
| INTC | 0.505 | -0.070 |
| ORCL | 0.479 | -0.197 |
| XOM | 0.495 | -0.151 |
| CVX | 0.500 | -0.130 |
| COP | 0.503 | -0.087 |
| HES | 0.519 | -0.118 |
| OXY | 0.509 | -0.097 |

The average directional accuracy across all stocks was approximately 50.3%, and the average $R^2$ score was -0.126. Sample plots for GOOG, OXY and INTC are shown in Figure 7, 8, 9, illustrating predicted versus actual excess returns.

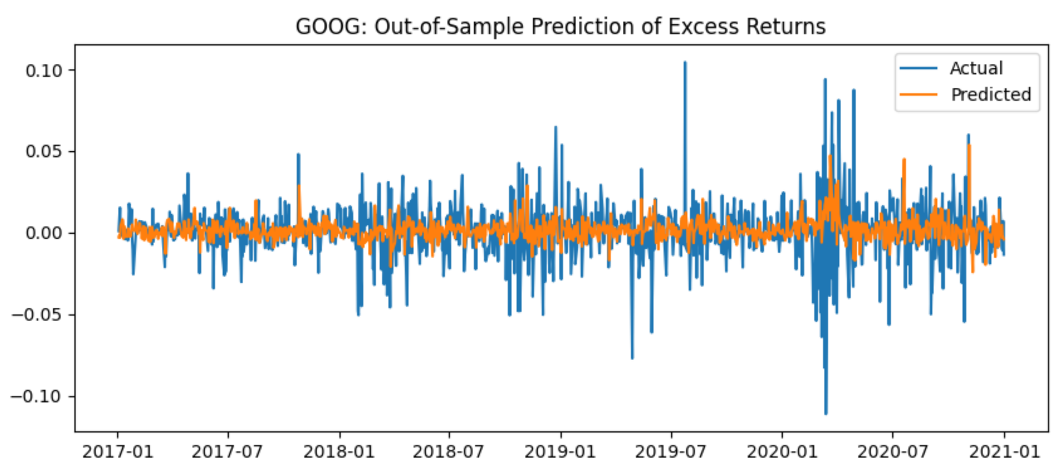


Figure 7: Out-of-Sample Prediction of GOOG Excess Returns

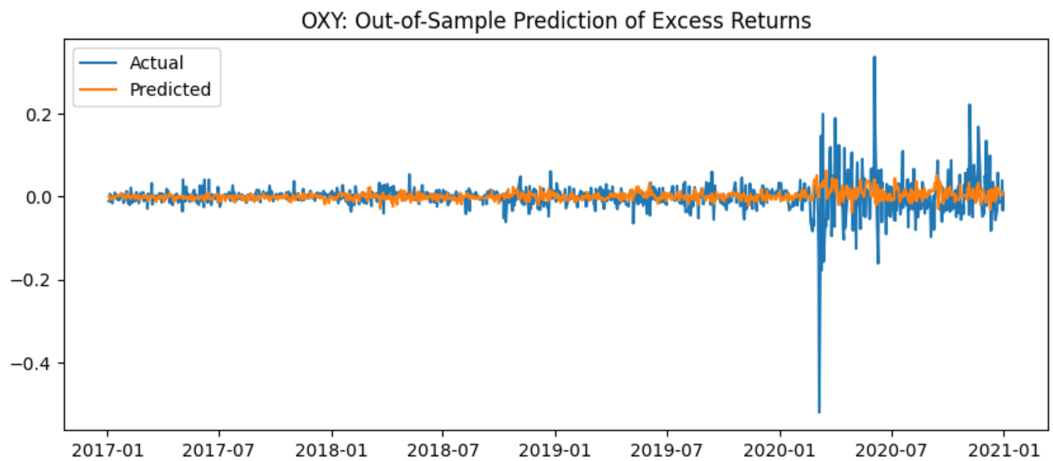


Figure 8: Out-of-Sample Prediction of OXY Excess Returns

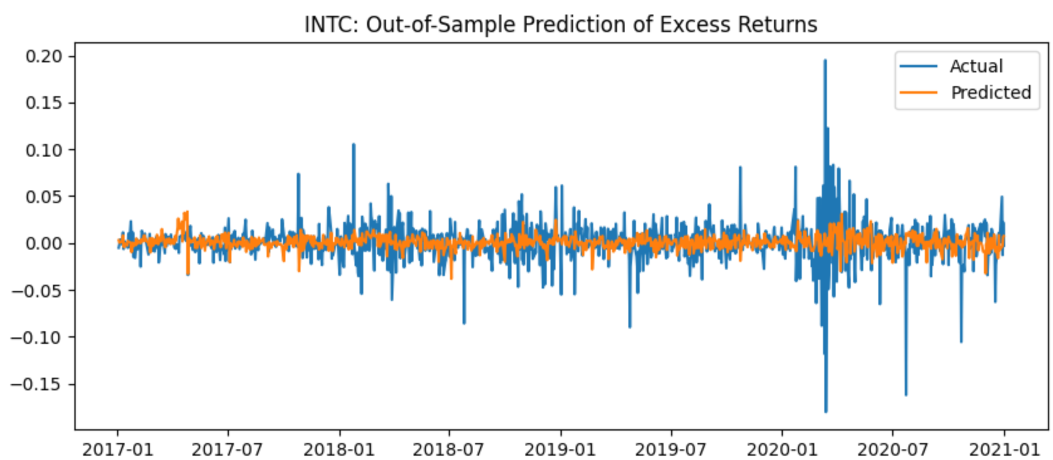


Figure 9: Out-of-Sample Prediction of INTC Excess Returns

**Discussion**

The directional accuracy of 50% indicates that the model performs only slightly better than random guessing in predicting the sign of returns. Negative $R^2$ values across all stocks suggest that the model does not capture the magnitude of returns well and is not predictive. Despite including both technical features (momentum, volatility) and fundamental factors (Fama-French 5 factors), short-term stock returns remain extremely noisy and challenging to predict. In particular, the Fama-French factors are fundamental features and are usually more informative over longer time horizons, which explains their limited usefulness for daily return prediction. This exercise showed the difficulty of predicting daily stock returns even with a powerful non-linear model like XGBoost. However, directional accuracy may still provide marginal signals for short-term trading strategies.