

ACTL2131 Assignment

Tadhg Xu-Glassop - z5480859

2024T1

Chosen stock: Amazon (NASDAQ: AMZN).

Q1. See Figure 1.

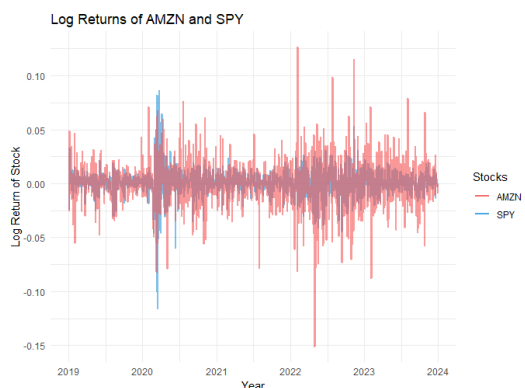


Figure 1: Daily Log Returns of AMZN and SPY between 2019 and 2024

Observing the graph, we notice a decrease in both log returns from 2019 to 2020, largely being the result of the COVID-19 Pandemic which caused significant economic downturn and market uncertainty and thus lower stock returns. Furthermore, in the first half of 2021, both log returns spiked significantly, likely due to numerous fiscal and monetary policies implemented by the US Government to increase consumer spending and economic activity. In the latter stage of the graph, AMZN displays periods of very high returns, significantly outperforming SPY. This is likely a result of AMZN experiencing unprecedented growth in FCF and earnings throughout this period, causing investors to favour the stock.

Q2. See the below table for key summary statistics of daily log returns.

	SPY	AMZN
Mean	0.0005755506	0.0005411658
Variance	0.0001763829	0.0004920690
Skewness	-0.8203029608	-0.1179294853
Kurtosis	15.3633758751	7.3121044619

Comparing the statistics of both log returns, we notice that despite having similar means, AMZN has significantly higher variance than SPY. This is expected, as AMZN is a single stock while the S&P500 is a diversified portfolio and thus should experience less volatility through diversification. AMZN is relatively symmetrical, with low negative skewness, while SPY has a moderate degree of negative skew. AMZN also has half the kurtosis of SPY, meaning SPY has significantly more returns around its mean and in its tails than AMZN.

Q3. See Figure 2. Note that the normal curves have mean and variance taken from the table in **Q2** for each respective stock.

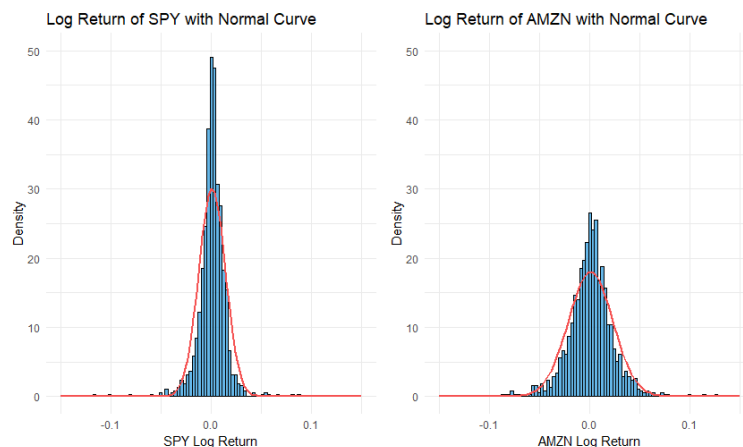


Figure 2: Density Histogram of both Log Returns with Normal Curve Overlay

With our density histograms, we can visualise the findings we noted in **Q2**. We see that SPY log returns do have significantly more kurtosis, with more scores concentrated around the mean and tails. The distribution of AMZN appears to be more spread than SPY, reflecting the higher variance. We also can see the significant excess kurtosis both log returns have, with the peak of both distributions considerably well above the peak of the overlaid normal curve.

Q4. Beyond the histograms, we can use two methods to assess whether either set of log returns fit a normal distribution.

Numerically, we can consider the excess kurtosis of both log returns, given by,

$$\kappa_{\text{excess}} = \kappa - 3.$$

Using the values found in **Q2**, we find that,

$$\kappa_{\text{excess, SPY}} \approx 12.36, \quad \kappa_{\text{excess, AMZN}} \approx 4.31.$$

By definition, a normal distribution has excess kurtosis of 0, meaning the log returns have significantly fatter tails and sharper peaks which cannot be modelled by a normal curve. In addition, SPY is moderately negatively skewed, which also cannot be modelled by a normal curve as again, a normal distribution by definition has 0 skew. Thus, the log returns do not fit a normal distribution.

Graphically, we can create a QQ plot with the theoretical quantiles from a normal distribution (with mean and variance from **Q2**) and visually see if a normal curve can fit the data. See Figure 3 for both plots.

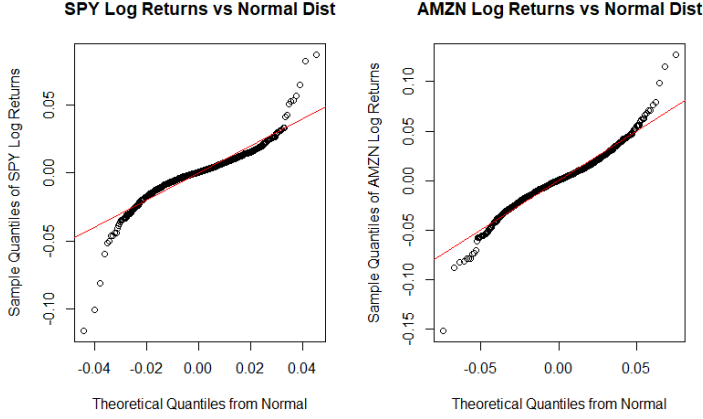


Figure 3: QQ Plots for SPY and AMZN with theoretical quantiles from $\text{Normal}(\bar{x}_{\text{SPY}}, s_{\text{SPY}}^2)$ and $(\bar{x}_{\text{AMZN}}, s_{\text{AMZN}}^2)$ respectively.

Observing the QQ Plots, we see that the quantiles deviate significantly from the reference lines, hence visually showing that both log returns do not fit a normal distribution.

In suggesting an alternative distribution, we can revisit our histogram in **Q3** and note our curve has a bell shape with excess kurtosis. We also seek a model that can model the negatively skewed nature of SPY. Thus, we can propose the Skew- T distribution as an alternative model, as this distribution (given the appropriate parameters) can model data with excess kurtosis and negative skew. Considering this, this distribution may be a better model than a normal distribution.

Q5. We can explore potential dependencies numerically and visually.

Numerically, we can calculate the Pearson correlation coefficient of the two log returns to measure linear dependency,

$$\rho_{\text{SPY}, \text{AMZN}} = 0.65.$$

0.65 is considered moderately significantly positive correlation, indicating that the two stocks have some significant degree of linear dependency.

Additionally, we can generate a scatter plot with both log returns to visualise any dependencies they may have (see Figure 4).

Observing our plot, we see that both returns are heavily clustered around 0 with some correlation around the line. Thus, we have shown that there is some linear dependency between the two stocks. This is as expected, as AMZN is currently the 4th largest constituent of the S&P500 and has remained in the top 25 throughout our time frame, meaning the return of the S&P500 is indeed influenced by AMZN.

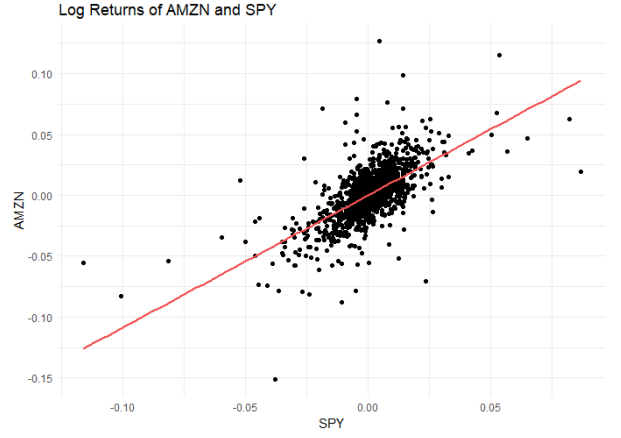


Figure 4: Scatter Plot of SPY and AMZN Log Returns with Line of Best fit

Q6. To test whether the expected annualised log returns of SPY and AMZN are statistically different, we can perform a hypothesis test on the difference of our sample average annualised returns, taking our null hypothesis to be that the true difference is 0 and our alternate hypothesis to be that the true difference is non-zero. That is,

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{v.s.} \quad H_1 : \mu_1 - \mu_2 \neq 0.$$

Consider the following test statistic, which is distributed as a standard normal by the Central Limit Theorem, for our test (assuming independence between our data sets and each data point is independent and identically distributed (IID)),

$$T = \frac{\bar{X}_{\text{SPY}} - \bar{X}_{\text{AMZN}} - (\mu_{\text{SPY}} - \mu_{\text{AMZN}})}{\sqrt{\frac{\sigma_{\text{SPY}}^2}{n_{\text{SPY}}} + \frac{\sigma_{\text{AMZN}}^2}{n_{\text{AMZN}}}}} \sim \mathcal{N}(0, 1),$$

where n is the sample size of either data set and the statistics are annualised (σ^2 and \bar{X} taken from **Q2** and multiplied by 250).

Using the approximation that $s^2 \approx \sigma^2$ for large n by Law of Large Numbers, we can find the test statistic to be $T \approx 0.04717057$. As our test is two-tailed, we have, $p = 2(1 - \Phi(0.04717057)) = 0.9623773$, i.e. the p -value is 0.9623773.

As our p value is significantly high, we lack sufficient evidence to reject our null hypothesis, and it is very likely that there is no statistical difference in our annualised log returns.

This test is predicated upon the assumption that the data sets are independent and each data point is IID. This first assumption is unlikely to be true, as in **Q5** we concluded that the datasets have a significant level of dependency. In addition, considering our data is log returns of stocks, it is an extreme assumption that each daily return is independent to each other, as the price of a stock is heavily influenced by speculation which directly involves considering past and expected future performance. Thus, as the two assumptions we made earlier are virtually impossible to justify, this test has limited application here and our result may be invalid.