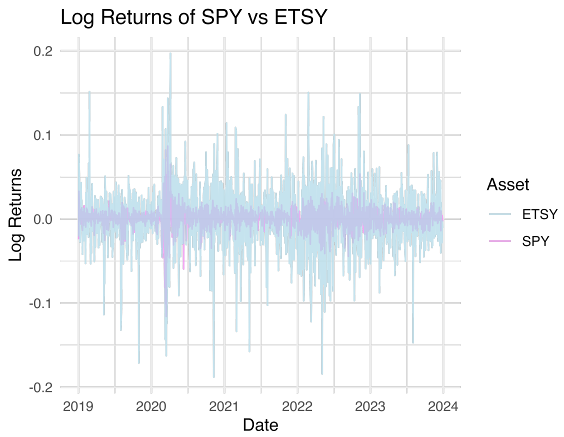
Daniel Devaris (z5480856) ACTL 2131 Assignment Part 2

Question 1:



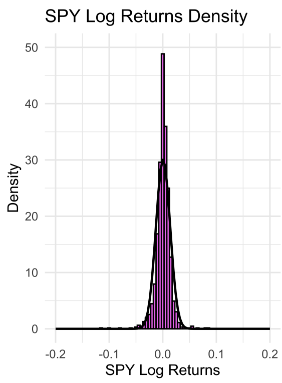
By inspecting the plot of S&P 500 (SPY) and Etsy's(ETSY) log returns over time we can observe the large negative spike during mid 2020, indicative of Covid-19's drastic impact on global financial markets. Directly following the negative spike, there is a positive spike for both SPY and ETSY, reflective of substantial fiscal stimulus packages delivered by the US government and several other economies in aim to boost economic activity in the wake of Covid-19.

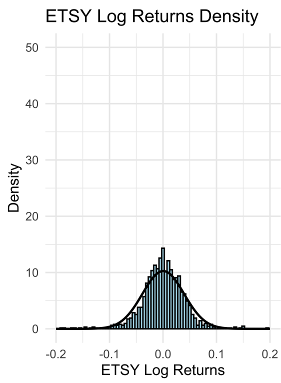
Question 2:

	SPY	ETSY
Mean	0.0005755506	0.0004335073
Variance	0.0001763828	0.0015099106
Skewness	-0.8193294638	-0.2064672044
Kurtosis	15.3389918179	6.0148156196

After calculating summary statistics for SPY and ETSY we can observe some interesting phenomena. Both means are very low and quite similar, around 0.058% and 0.043% for SPY and ETSY respectively. Further, the variances are extremely low, however, ETSY's variance is roughly 10x larger than SPY. Since $0.5 < |\text{Skew}_{\text{SPY}}| < 1$ we can say SPY is slightly negatively skewed, and since $|\text{Skew}_{\text{ETSY}}| < 0.5$ we can say it is approximately symmetrical. Further, both SPY and ETSY are lepokurtic (Kurtosis > 3) and therefore have a higher peak and heavier tails than the normal distribution. However, SPY's kurtosis is much more extreme than ETSY (≈ 2.5x larger), demonstrating that SPY has a much higher peak and heavier tails than ETSY.

Question 3:



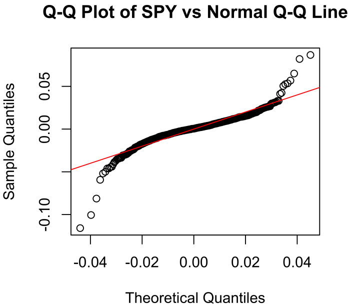


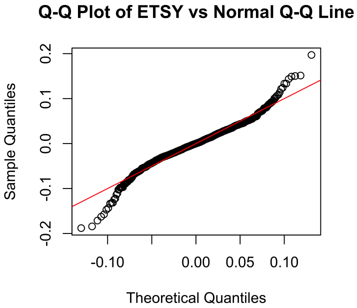
Both respective histograms seem to fit a "bell-shaped curve", however, it is not fully reflective of a normal distribution. The statistics gathered in Question 2 confirms this observation. The high levels of kurtosis reflects a much greater peak in both histograms compared to the normal distribution PDF with same mean and variance. The slight negative skew for ETSY is also noticeable in the histogram.

Question 4:

Numerical: The excess kurtosis ($\kappa_{\text{excess}} = \kappa - 3$) of any normal distribution is always 0, the excess kurtosis for SPY and ETSY are ≈ 12.3 and ≈ 3.01 respectively. These observed statistics numerically demonstrate that both of the data sets do not fit a normal distribution, especially SPY due to its relatively extreme kurtosis.

Visual: We can use further data visualisation to assess whether both SPY and ETSY log returns fit a normal distribution. Using a Q-Q plot, which plots the theoretical quantiles of a distribution, in this case the normal distribution, against the sample quantiles of a data set we can assess whether the data fits a normal distribution. Another visual tool we can use to test the similarity between SPY and ETSY and the normal distribution is a Q-Q plot, which plots the theoretical quantiles of their respective normal distribution and the observed quantiles from the data. If the Q-Q plot forms a perfectly diagonal line, it reflects that that dataset is distributed in the theorised way.

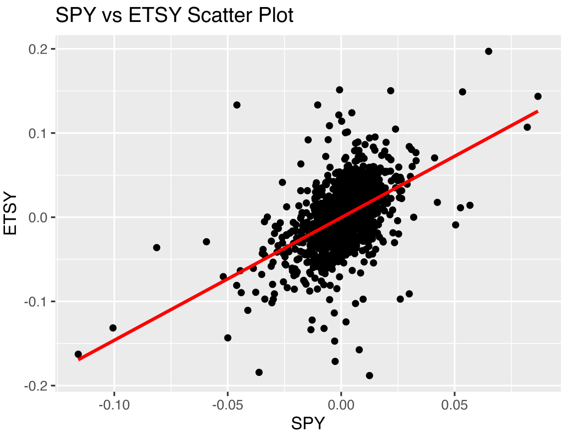




Examination of the Q-Q plot is further evidence that both sets of data do not fit a normal distribution. The main discrepancy between both SPY and ETSY and their respective normal distributions are in the tails. This is a consequence of the kurtosis in each sample, noting that the tails of SPY Q-Q plot fall off the theoretical values to a greater degree than ETSY, symptomatic of the greater kurtosis in SPY than ETSY.

Due to the shown non-normality, we can suggest a lepokurtic distribution such as the Laplace distribution for ETSY to better fit our data. Since SPY is slightly negatively skewed we can refine this suggestion to the Asymmetrical Laplace distribution. Reflecting more suitable distributions for SPY and ETSY than the normal.

Question 5:



Numerical: To explore the dependency between to sets of data, we can examine the correlation between two data sets. Using R, we can calculate the following correlation between SPY and ETSY as $\rho = 0.4981196$, indicating a low correlation between SPY and ETSY since $0.3 < |\rho| < 0.5$. However, since ETSY is a constituent of the S&P500, we can attribute this low correlation to ETSY having the 484th largest weighting on the S&P500, reflecting an insignificant weighting on SPY and negligible dependency.
Graphical: For graphical examination, we can create a scatter plot (figure on left) with a line of best fit between SPY and ETSY as seen to the left. The data seems to be centered around zero but with very little correlation on the rest of the scatter plot. Thus, the graph demonstrates our numerical findings, demonstrating that SPY and ETSY have a low dependency.

Question 6:
To test for statistical difference between SPY and ETSY's average annualised returns, we can formulate the following hypothesis test, where X = SPY and Y = ETSY:

$$H_0 : \mu_X - \mu_Y = 0 \quad \text{vs.} \quad H_1 : \mu_X - \mu_Y \neq 0,$$

for the significance level of $\alpha = 0.05$. Now we can formulate the following test statistic:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \sim N(0, 1).$$

To apply this test statistic, we have to assume that $X \perp\!\!\!\perp Y$, all observations within X and Y are independent and that the population variance is known. Since the population variance is unknown, with support from the Law of Large Numbers, since $n_X = n_Y = 1257$, we can assume that $s_X \approx \sigma_X$ and $s_Y \approx \sigma_Y$. Note that the value of $T|H_0$ will be an approximate value. Hence,

$$T|H_0 \approx \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}} = 0.1226371.$$

We can use this result as a quantile of the standard normal distribution to calculate the p-value such that,

$$p_{value} \approx 2(1 - \Phi_Z(T|H_0)) = 0.9023945.$$

Since our $p_{value} > \alpha$ we do not have enough evidence to reject H_0. However, we must note that the assumptions made to use this test are inherently flawed. It is impossible to say that daily log returns are not dependent of previous realisations due to the highly speculative nature of the stock market. Further, X and Y cannot be truly independent since Etsy is a component of the S&P500, having direct influence. Thus, our test statistic is ineffective in generating a valid critical region, compromising the whole hypothesis test.