COMP0051 Algorithmic Trading Coursework 1

February 21, 2024

Time Series [10 Points]

1. Download two ETF time series using an API. The length of the time series T, with T=300 and a daily resolution.

SPY (SPDR S&P 500 ETF Trust) and QQQ (Invesco QQQ Trust) are selected for their prominence in reflecting significant segments of the US stock market, making them excellent subjects for financial analysis and comparison. Both ETFs are highly liquid, trade on major US stock exchanges, and offer insights into different, yet complementary, aspects of the market. **SPY** broadly represents the US economy through the S&P 500 Index (tracking the stock performance of the 500 biggest companies that are listed on US stock exchanges), while **QQQ** tracks the Nasdaq-100 Index, which is currrently composed of 101 securities issued by 100 of the biggest non-financial businesses listed on the Nasdaq stock exchange, which results in The Nasdaq-100 Index being a very tech heavy index.[11] Some of the companies in these indexes overlap, making them ideal for comparison.

Because SPY and QQQ have different sector focuses and similar market relevance and liquidity, they are perfect for comparative analysis because they provide different perspectives on market dynamics and sector performance. Their roles as benchmarks for different market sectors provide valuable insights for economic and investment strategies.

2. Plot the price time series

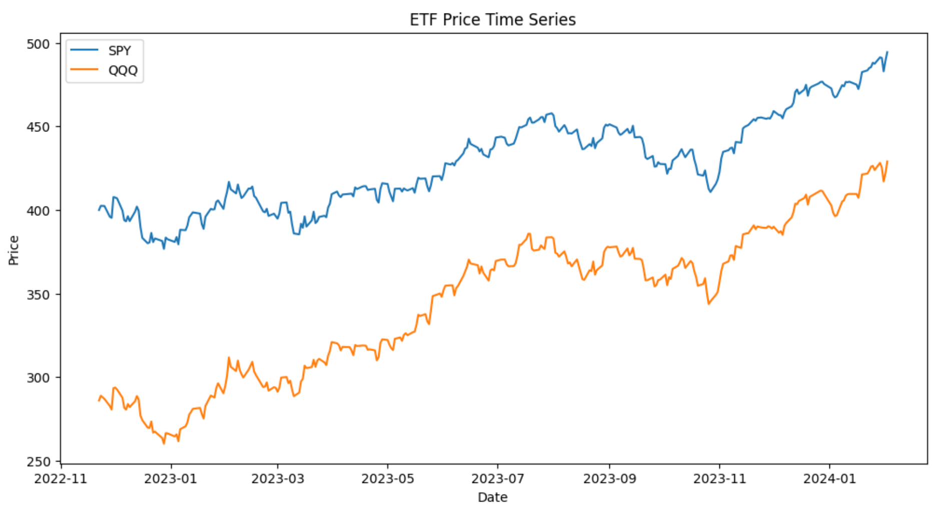


Figure 1: SPY and QQQ Price Time Series

Moving averages [20 Points]

3. Define mathematically the moving average of the price time series with an arbitrary time-window t

The moving average is calculated as the mean of the present value in the time series and the values immediately preceding it.[3]

- The formula for the moving average y_t is given by:

$$y_t = \frac{1}{3}(\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2})$$

where ε_t follows a normal distribution with mean 0 and variance σ^2, denoted by $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.

To incorporate an arbitrary time window for the purposes of this task we can modify the formula to look like this:

- Let $P(t)$ represent the price at time t.
- The moving average $MA(t)$ at time t for a time-window τ is given by the formula:

$$MA(t) = \frac{1}{\tau} \sum_{n=0}^{\tau-1} P(t-n)$$

4. Compute three moving averages of the price time series, with time-windows t=5,20,60

```
Ticker             QQQ         SPY     SPY_MA5     QQQ_MA5    SPY_MA20  \
Date
2024-02-13  428.549988  494.079987  498.536005  433.143994  489.407503
2024-02-14  433.220001  498.570007  498.630005  433.389996  490.721503
2024-02-15  434.510010  502.010010  499.368005  433.733997  491.997504
2024-02-16  430.570007  499.510010  499.030005  432.438000  492.851505
2024-02-20  427.320007  496.760010  498.186005  430.834003  493.517004

Ticker        QQQ_MA20    SPY_MA60    QQQ_MA60
Date
2024-02-13  425.282498  473.436168  407.542166
2024-02-14  426.582999  474.241835  408.329999
2024-02-15  427.659000  475.095502  409.137832
2024-02-16  428.128500  475.849668  409.801666
2024-02-20  428.408000  476.574502  410.449166
```

Figure 2: Latest SPY, QQQ prices, and their computed 5, 20, 60-day moving averages. Last 5 rows of the computed dataframe.

5. Plot the moving averages against the price time series

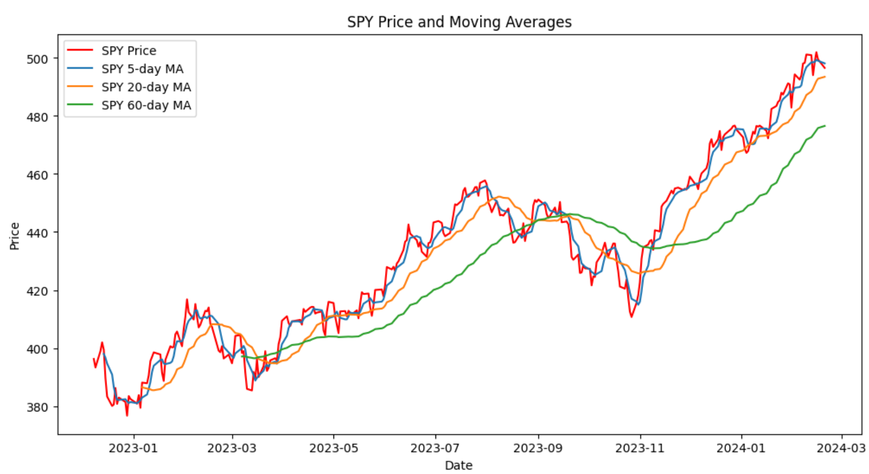


Figure 3: Plotted SPY prices and 5, 20, 60-day moving averages.

Figure 3 displays SPY's price movement along with its 5, 20, and 60-day moving averages over time. The moving averages smooth out the price data to identify trends; the shorter moving averages (5-day) follow the price more closely, while the longer ones (60-day) show longer-term price movements. In the period after November 2023, the SPY ETF exhibited a positive trend as evidenced by its price consistently remaining above the 60-day moving average. This positioning above the long-term moving average is typically interpreted as a bullish market sentiment, suggesting a robust confidence level among investors regarding the future performance of the S&P 500 index.

The sustained position above a moving average indicator usually signals a strong upward momentum, which may be utilized by traders as a confirmation for maintaining or initiating long positions. The 60-day MA also serves as a dynamic support level, providing potential entry points during price retracements in an overall uptrend. Such technical behaviors are crucial in developing strategic trading decisions, complementing fundamental analysis.[8]

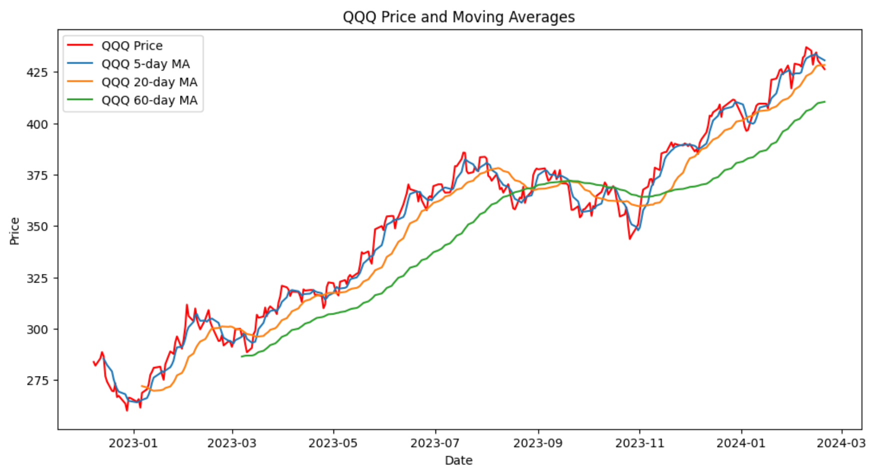
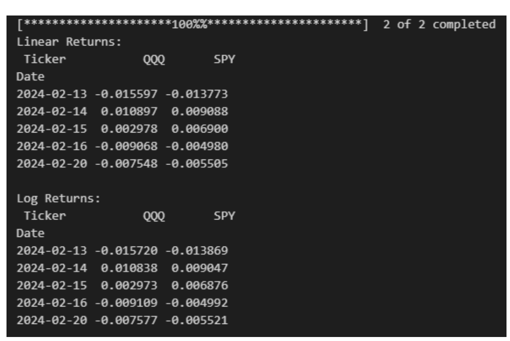


Figure 4: Plotted QQQ prices and 5, 20, 60-day moving averages.

As can be seen in Figure 4, trends observed in the QQQ ETF, which tracks the NASDAQ-100 Index, often mirror those in the SPY ETF due to the interconnectivity of the markets and the presence of overlapping companies in these indexes. Even though QQQ is more tech-oriented, the tech sector's significant weight in the S&P 500 means that movements in technology stocks can significantly influence both ETFs. In addition, major economic trends and monetary policies can affect the performance of both SPY and QQQ at the same time, leading to parallel trends in their price and moving averages.

6. Compute the linear and log-return of the price time series



```
[*********************100%%**********************]  2 of 2 completed
Linear Returns:
 Ticker           QQQ       SPY
Date
2024-02-13 -0.015597 -0.013773
2024-02-14  0.010897  0.009088
2024-02-15  0.002978  0.006900
2024-02-16 -0.009068 -0.004980
2024-02-20 -0.007548 -0.005505

Log Returns:
 Ticker           QQQ       SPY
Date
2024-02-13 -0.015720 -0.013869
2024-02-14  0.010838  0.009047
2024-02-15  0.002973  0.006876
2024-02-16 -0.009109 -0.004992
2024-02-20 -0.007577 -0.005521
```

Figure 5: Computed linear and log return of the price time series for QQQ and SPY. Last 5 rows of the computed dataframe.

7. Plot the linear return against the log-return time series

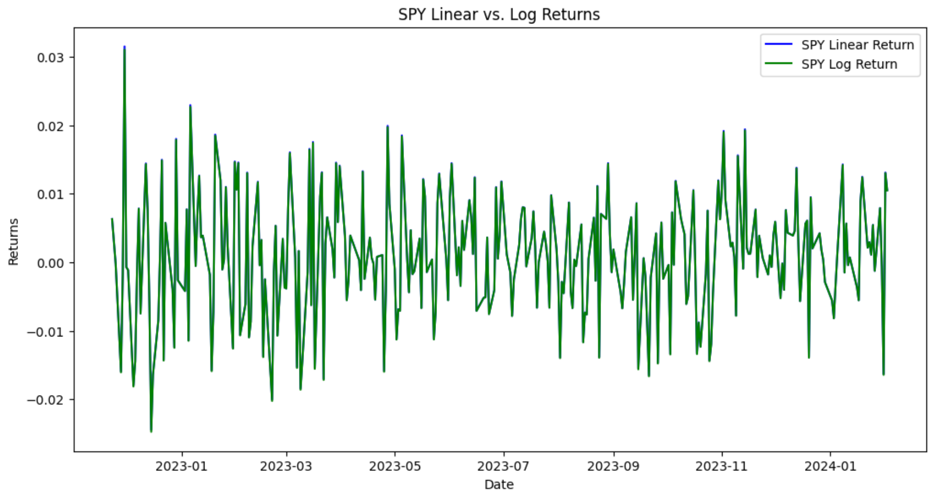


Figure 6: SPY linear returns plotted against SPY log-returns.

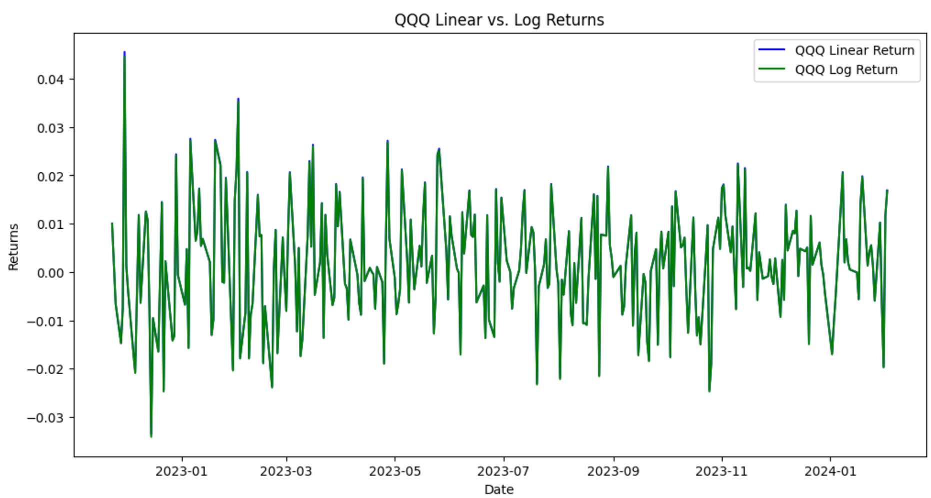


Figure 7: QQQ linear returns plotted against QQQ log-returns.

The presented graphs for the SPY and QQQ ETF show a notable similarity between the log and linear returns, indicating low volatility and steady price fluctuations over the studied time. This convergence highlights the stable market behaviour of both SPY and QQQ.

Time Series Analysis [20 Points]

8. Define the auto-correlation function (for a stationary time-series)

For a stationary time series X_t, the auto-correlation function (ACF) at lag k is defined as follows:[9]

- Let μ be the mean of the time series X_t and σ^2 be its variance.
- The auto-correlation function $\rho(k)$ at lag k is given by the formula:

$$\rho(k) = \frac{E[(X_t - \mu)(X_{t-k} - \mu)]}{\sigma^2}$$

where:

- E is the expectation operator.
- X_t is the value of the time series at time t.
- X_{t-k} is the value of the time series at time $t - k$.
- μ is the mean of the time series.
- σ^2 is the variance of the time series.
- k is the lag, a non-negative integer.

$\rho(k)$ measures the degree of correlation between the values of the given time series separated by a lag of k periods. The ACF output value is between -1 and 1, where 0 indicates no correlation, 1 signifies perfect positive correlation, -1 indicates perfect negative correlation.

9. Compute the auto-correlation functions (ACF) of the price time series

10. Plot the price ACFs

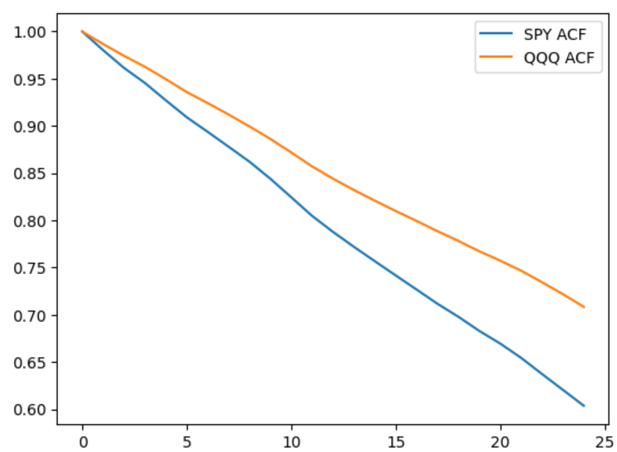


Figure 8: ACF for SPY and QQQ

The ACF plots for SPY and QQQ ETFs reveal distinct autocorrelation characteristics. With both starting at an autocorrelation of 1 at lag 0, SPY's autocorrelation declines more slowly than QQQ's, implying that past SPY returns have a more prolonged effect on future values. In contrast, QQQ's faster decline in autocorrelation suggests its returns become uncorrelated more quickly, hinting at a market that more rapidly accounts for new information.

11. Compute the partial auto-correlation functions (PACF) of the price time series

12. Plot the price PACFs

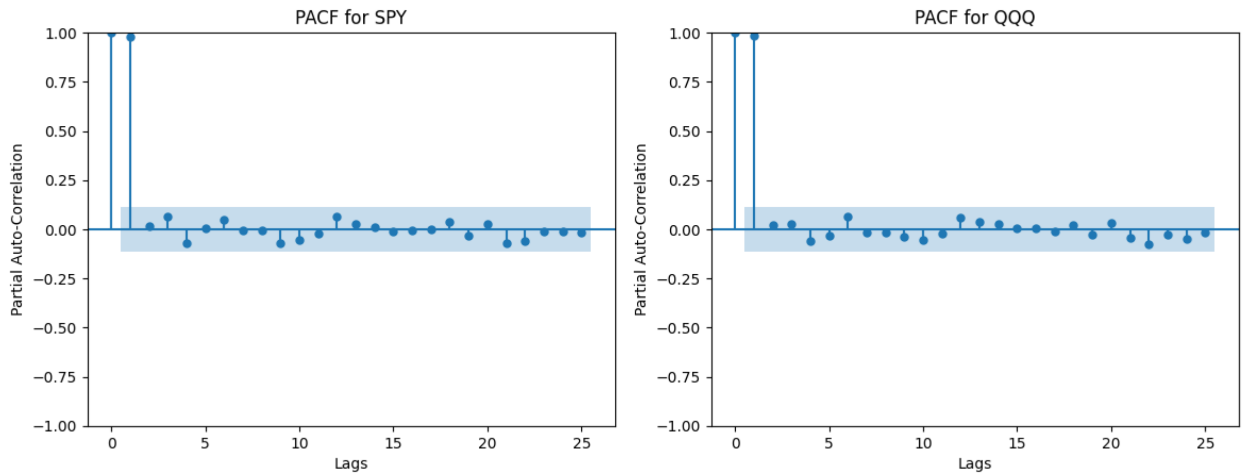


Figure 9: Plotted PACF of SPY and QQQ

The PACF graphs for SPY and QQQ in Figure 9 suggest that both ETFs exhibit a strong partial autocorrelation at lag 1, with subsequent lags showing no significant correlation. This indicates an AR(1) characteristic for both time series, where current values are primarily influenced by their immediate past values.[7]

13. Compute the auto-correlation function (ACF) of the return time series

14. Plot the return ACFs

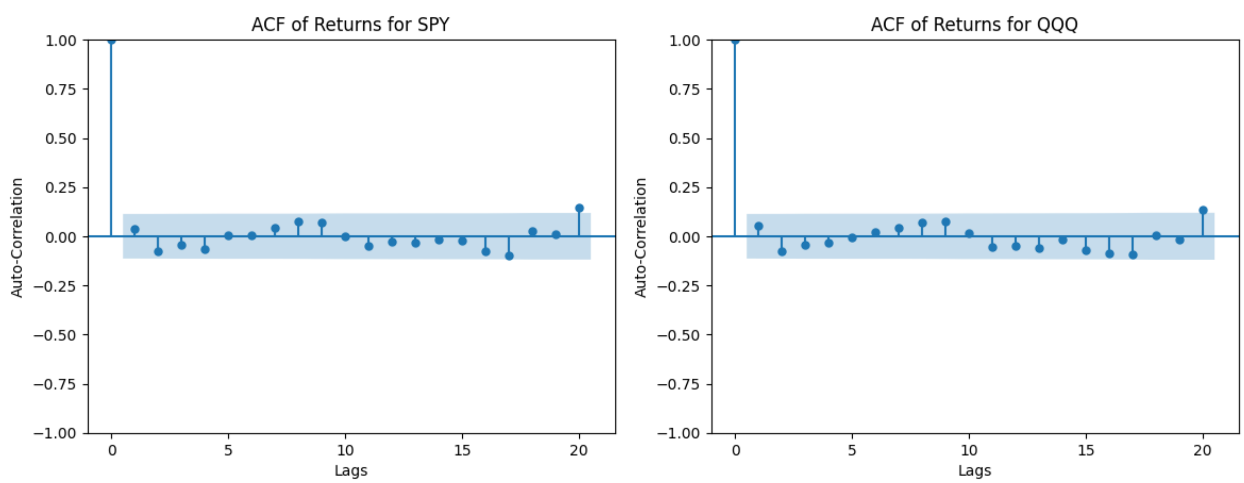


Figure 10: Plotted ACF of SPY and QQQ returns.

The autocorrelation of the return series of the SPY and QQQ ETFs shows a rapid fall after the first lag, dropping to almost zero levels and staying within the confidence ranges for the successive lags, according to the ACF plots in Figure 10. This pattern suggests that returns are largely random and do not exhibit strong serial dependence, aligning with the Efficient Market Hypothesis which posits that asset prices fully reflect all available information.[2] This may indicate, from a financial standpoint, that historical returns are not very predictable or useful for predicting returns in the long run.

15. Compute the partial auto-correlation functions (PACF) of the return time series

16. Plot the return PACFs

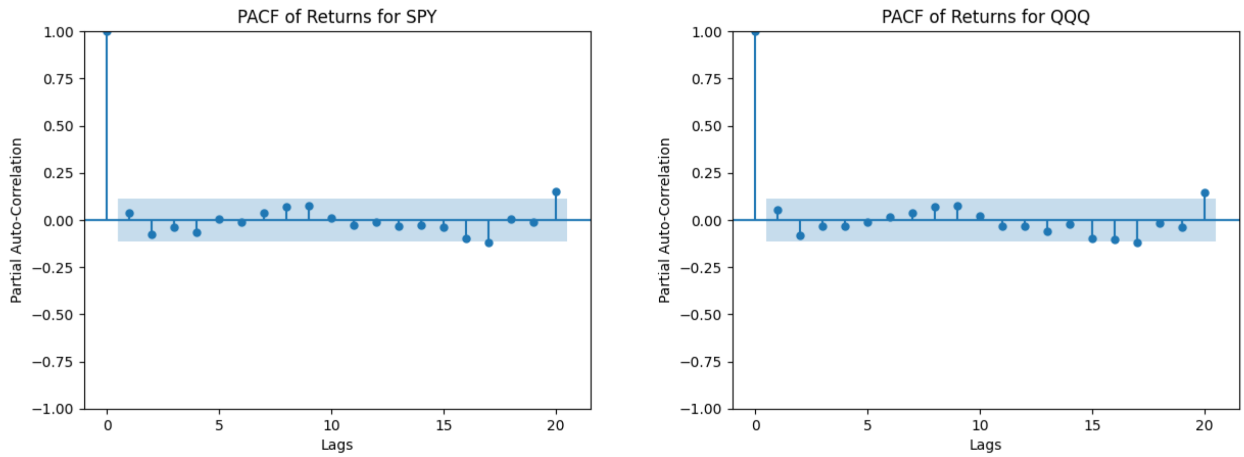


Figure 11: Plotted PACF of SPY and QQQ returns.

Similar to previous plots, the PACF plots for SPY and QQQ ETFs in Figure 11 show the vast majority of partial autocorrelations within the confidence interval after the initial lag. This pattern suggests that any linear predictability from past values does not persist beyond the first difference. The initial significant spike at lag 1, particularly for SPY, indicates a potential AR(1) process,[7] which is common in financial time series. After accounting for the first lag, subsequent values do not seem to offer additional predictive power. This behavior aligns with efficient market conditions where prices reflect all available information up to the present, rendering past data less useful for predicting future returns.[2]

Gaussianity and Stationarity test [20 Points]

17. Introduce mathematically a Gaussianity test

The Shapiro-Wilk test calculates a statistic, W, which can be used to assess the normality of a distribution.[6] The test hypothesis for the Shapiro-Wilk test [10] is as follows:

- Null hypothesis (H_0): The data originates from a population with a normal distribution.
- Alternative hypothesis (H_1): The data does not originate from a population with a normal distribution.

The test statistic W is defined as:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

where:

- n is the sample size,
- $x_{(i)}$ are the ordered sample values ($x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$),

- a_i represents constants that were generated from the covariances, variances, and means of the order statistics of a normally distributed sample,
- x_i represents sample values,
- $\bar{x}$ is the mean of the sample values.

The value of W ranges from 0 to 1, where a value close to 1 indicates that the data is likely normally distributed. The significance level (usually denoted as alpha, α) is chosen, and the p-value from the test is compared against this α to decide whether to reject the null hypothesis. For our test, we selected alpha to be 0.05.

18. Perform a Gaussianity test of the return time series

```
[*********************100%%**********************]  2 of 2 completed
SPY - Statistics=0.994, p=0.244
SPY returns look Gaussian (fail to reject H0)
QQQ - Statistics=0.997, p=0.831
QQQ returns look Gaussian (fail to reject H0)
```

Figure 12: Shapiro-Wilk test computation output for SPY and QQQ returns.

As shown in Figure 12, the Shapiro-Wilk test results for the SPY and QQQ ETF returns, with p-values of 0.244 and 0.831 respectively, fail to reject the null hypothesis of normal distribution. This implies that the returns for these ETFs are consistent with a Gaussian distribution within the examined timeframe. This suggests that traditional statistical methods and models that assume normality may be appropriate for analyzing and forecasting the returns of these ETFs.

19. Introduce mathematically a stationarity test

A common stationarity test used in time series analysis is the **Augmented Dickey-Fuller (ADF) test**. The ADF test is utilized to determine whether a time series is stationary, meaning its statistical properties do not vary over time. The test specifically examines the null hypothesis that a unit root is present in an autoregressive model of the time series, implying non-stationarity.[4]

The ADF test statistic is derived from the following regression model [14]:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \ldots + \delta_p \Delta y_{t-p} + \epsilon_t$$

where:

- y_t is the time series at time t,
- $\Delta y_t = y_t - y_{t-1}$ is the first difference of the time series,
- α is a constant,
- βt is the coefficient on a time trend,
- γ is the coefficient on y_{t-1},
- $\delta_1, \delta_2, \ldots, \delta_p$ represent weights for the initial p lagged differences in the series, indicating the influence of past changes on the present value.,
- ϵ_t is the stationary error process.

The null hypothesis (H_0) of the test is that $\gamma = 0$, indicating the presence of a unit root and thus non-stationarity. The alternative hypothesis (H_1) depends on the specific form of the test but generally suggests that the time series is stationary.

The ADF test statistic is calculated as:

$$\tau = \hat{\gamma}/\mathrm{SE}(\hat{\gamma})$$

where $\hat{\gamma}$ is the estimated value of γ from regression calculation, and $\mathrm{SE}(\hat{\gamma})$ is the standard error of $\hat{\gamma}$. This statistic is then compared to critical values for the ADF distribution. If τ ends up being smaller than the critical value, we reject H_0, showing evidence against the presence of a unit root and in favor of stationarity.

20. Perform a stationarity test of the return time series

```
[*********************100%**********************]  2 of 2 completed
SPY - ADF Statistic: -16.856
SPY - p-value: 0.000
SPY returns appear to be stationary (reject H0)
QQQ - ADF Statistic: -16.454
QQQ - p-value: 0.000
QQQ returns appear to be stationary (reject H0)
```

Figure 13: ADF computation output for SPY and QQQ return time series

The ADF test produces highly negative statistics for SPY (-16.856) and QQQ (-16.454) with p-values at 0.000, leading to the rejection of the null hypothesis of a unit root. This implies that the return series for both ETFs do not follow a random walk and are mean-reverting, which is a key assumption for many time-series forecasting models. The stationarity of these series implies that past return behavior can be informative for predicting future behavior.[1]

Cointegration tests [30 Points]

21. Define mathematically a cointegration test

One widely used cointegration test is the **Engle-Granger two-step method**.[5][13]

Engle-Granger Two-Step Method:

Step 1: Regression

First, one non-stationary time series is regressed on another to obtain the residual series:

$$Y_t = \alpha + \beta X_t + \epsilon_t$$

where:

- Y_t and X_t are the non-stationary time series,
- α and β are coefficients,
- ϵ_t is the residual of the regression at time t.

Step 2: Stationarity Test of Residuals

Then, the residuals (ϵ_t) are tested for stationarity using a unit root test like the Augmented Dickey-Fuller (ADF) test.[12] If the residuals are found to be stationary, then Y_t and X_t are considered to be cointegrated.

Interpretation:

- **Null Hypothesis (H_0)**: There is no cointegration, implying the residuals are non-stationary.
- **Alternative Hypothesis (H_1)**: There is cointegration, implying the residuals are stationary.

If the test statistic is less (more negative) than the critical value, or if the p-value is less than a chosen significance level (in our case 0.05), H_0 is rejected, indicating that the series are cointegrated.

22. Perform a cointegration test of the two ETF price time series

```
[*********************100%%**********************]  2 of 2 completed
SPY and QQQ are not cointegrated
p-value: 0.1842395272378916
```

Figure 14: Output of cointegration computation for SPY and QQQ price time series.

As can be seen in Figure 14, a cointegration test between the SPY and QQQ ETF price time series has resulted in a p-value of approximately 0.184. This p-value is higher than the set significance level of 0.05, which suggests that there is not enough statistical evidence to conclude that SPY and QQQ are cointegrated.

In a financial context, this means that despite both ETFs being major trackers of US equities, they do not necessarily move together in a long-term equilibrium relationship. In other words, the price movements of SPY and QQQ are not tied together in a way that one could reliably predict the long-term movement of one based on the other. This could be due to the different compositions of the indices they track, with SPY focused on a broad range of large-cap U.S. stocks and QQQ concentrated on tech-heavy and non-financial stocks from the Nasdaq-100 index.[11]

23. Perform a cointegration test of the two ETF return time series

```
[*********************100%%**********************]  2 of 2 completed
SPY and QQQ returns are cointegrated
p-value: 7.132852368949067e-18
```

Figure 15: Output of cointegration computation for SPY and QQQ return time series.

As can be seen in Figure 15, the provided output indicates that the SPY and QQQ ETF returns have passed a cointegration test, with a highly significant p-value near zero (7.13e-18). This result suggests a strong cointegration relationship, implying that the returns of SPY and QQQ move together in the long run despite potential short-term deviations.[13]

In a financial context, the existence of cointegration between these two ETFs could mean that while they may individually experience price fluctuations, their returns are bound by an equilibrium relationship over time. This could be useful for investors interested in portfolio diversification, pairs trading, or long-term investment strategies that capitalize on the relationship between the broader market and the technology sector, as represented by SPY and QQQ, respectively.

References

[1] Affek, S. P. [2024], 'Stationarity in time series analysis', https://towardsdatascience.com/stationarity-in-time-series-analysis-90c94f27322. Accessed: 2024-02-9.

[2] Baldridge, R. [2022], 'What is the efficient market hypothesis?', https://www.forbes.com/advisor/investing/efficient-market-hypothesis/. Accessed: 2024-02-14.

[3] COMP0051 [15 January 2024*a*], 'Algorithmic trading slides - lecture 2b'. Slide 5.

[4] COMP0051 [22 January 2024*b*], 'Algorithmic trading slides - lecture 3a'. Slide 24-25.

[5] COMP0051 [22 January 2024*c*], 'Algorithmic trading slides - lecture 3b'. Slide 6.

[6] COMP0051 [29 January 2024*d*], 'Algorithmic trading slides - lecture 4 part 1'. Slide 27-28.

[7] Miller, S. J. [2021], 'Autoregressive models', https://sjmiller8182.github.io/LearningTS/build/autoregressive_models.html. Accessed: 2024-02-10.

[8] Mitchell, C. [2023], 'How to use a moving average to buy stocks', https://www.investopedia.com/articles/active-trading/052014/how-use-moving-average-buy-stocks.asp. Accessed: 2024-02-05.

[9] Penn State Online Statistics Course [2023], 'Lesson 1.2 - introduction to time series', https://online.stat.psu.edu/stat510/lesson/1/1.2. Accessed: 2024-02-17.

[10] Ramachandran, K. M. and Tsokos, C. P. [2021], Chapter 11 - categorical data analysis and goodness-of-fit tests and applications, *in* K. M. Ramachandran and C. P. Tsokos, eds, 'Mathematical Statistics with Applications in R (Third Edition)', third edition edn, Academic Press, p. 484.
URL: *https://www.sciencedirect.com/science/article/pii/B9780128178157000117*

[11] Sanchez, J. [2023], 'Spy vs qqq: How do they compare?', https://www.physicianonfire.com/spy-vs-qqq/#:~:text=SPY%20is%20an%20ETF%20that,companies%20in%20the%20financial%20sector. Accessed: 2024-02-12.

[12] *statsmodels.tsa.stattools.coint - Statsmodels Documentation* [2023], https://www.statsmodels.org/stable/generated/statsmodels.tsa.stattools.coint.html. Accessed: 2024-02-06.

[13] Wang, Y. [2021], 'An introduction to cointegration', https://hudsonthames.org/an-introduction-to-cointegration/. Accessed: 2024-02-15.

[14] Y. Verma [2024], 'Augmented dickey-fuller (adf) test in time-series analysis', https://analyticsindiamag.com/complete-guide-to-dickey-fuller-test-in-time-series-analysis/. Accessed: 2024-02-18.