COMP0051 Algorithmic Trading
Coursework 1

Yue He

February 2024

CONTENTS

I. TIME SERIES

1. SELECT AND DOWNLOAD ETFS

This analysis focuses on the SPDR SP 500 ETF (SPY) and the Technology Select Sector SPDR Fund (XLK) to assess the performance of the technology sector in comparison to broader market benchmarks. The time series data, spanning 300 trading days, was sourced from the Yahoo Finance API, covering the period from December 6, 2022, to February 16, 2024.
Closing price is the standard benchmark used to track financial performance over time. In this coursework closing price is used for price time series.

2. PLOT PRICE TIME SERIES

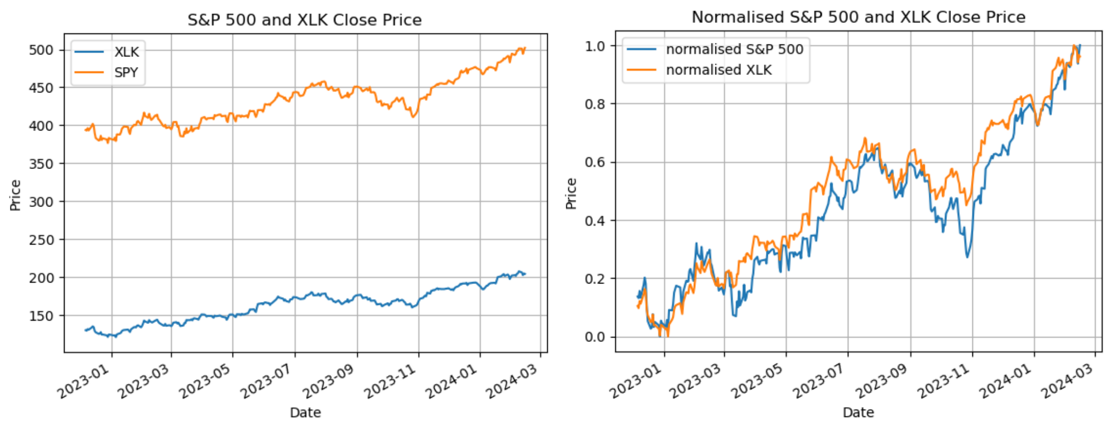

Figure 2.1: Comparison of the price time series of S&P 500 and XLK

Normalised prices of the S&P 500 and the Technology Select Sector SPDR Fund exhibit a comparable trend, creating a potential opportunity for statistical arbitrage through cointegration. Further statistical tests should be employed to validate and refine this observation.

II. MOVING AVERAGES

1. DEFINE MOVING AVERAGE

The Simple moving average of the price time series with an arbitrary time window τ can be defined as a unweighted mean of the current value of a time series and its τ previous neighbours in the past.

$$\text{SMA}_t(\tau) = \frac{1}{\tau}\left(P_{t-(\tau-1)} + P_{t-(\tau-2)} + \ldots + P_{t-1} + P_t\right)$$

$$\text{SMA}_t(\tau) = \frac{1}{\tau}\sum_{i=t-\tau+1}^{t} P_i$$

Simple moving average is adapted in this course work as it treats all prices in the window with equal weights which is suitable for capturing and exploring underlying trends in a price time series.

2. COMPUTE AND PLOT MOVING AVERAGE

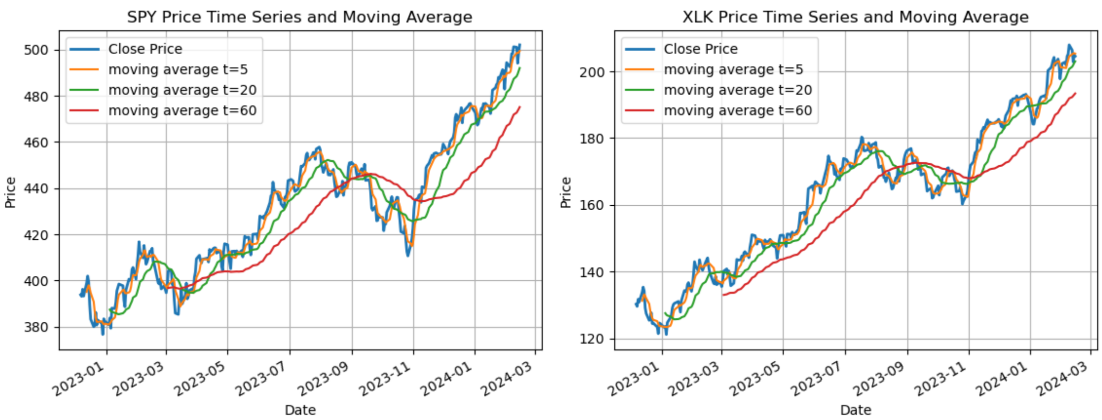

Figure 2.1: Moving averages of S&P 500 and XLK

As observed, a short-term moving average with a 5-day window is highly responsive to immediate price changes, whereas a long-term moving average with a 60-day window provides a more representative depiction of the longer-term trend.

A viable trading strategy for this moving average data is gold cross analysis. When the short-term moving average ($\tau = 5$) crosses above the long-term moving average ($\tau = 60$) from below, it signals a 'Golden Cross' which indicating a potential bullish trend and time to buy. In November 2023, signalling a favourable time to initiate a buy position.

3. COMPUTE AND PLOT LINEAR AND LOG RETURNS

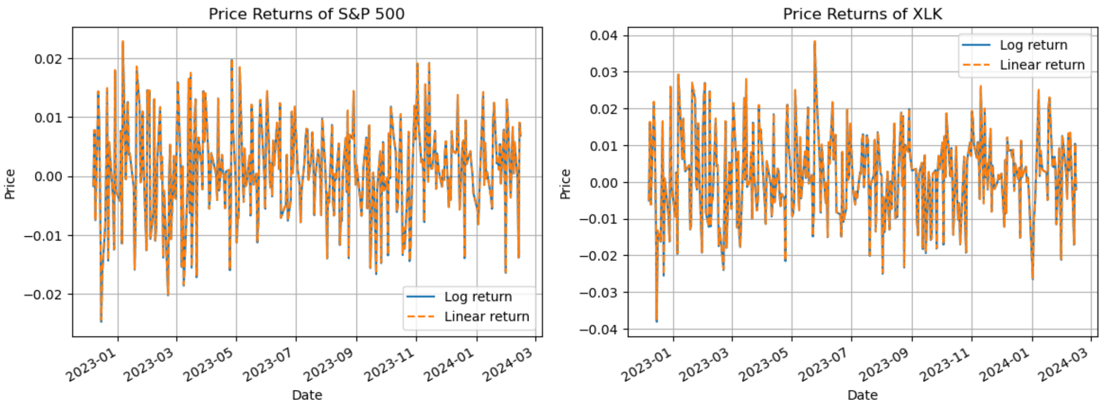

Figure 3.1: Price returns of S&P 500 and XLK

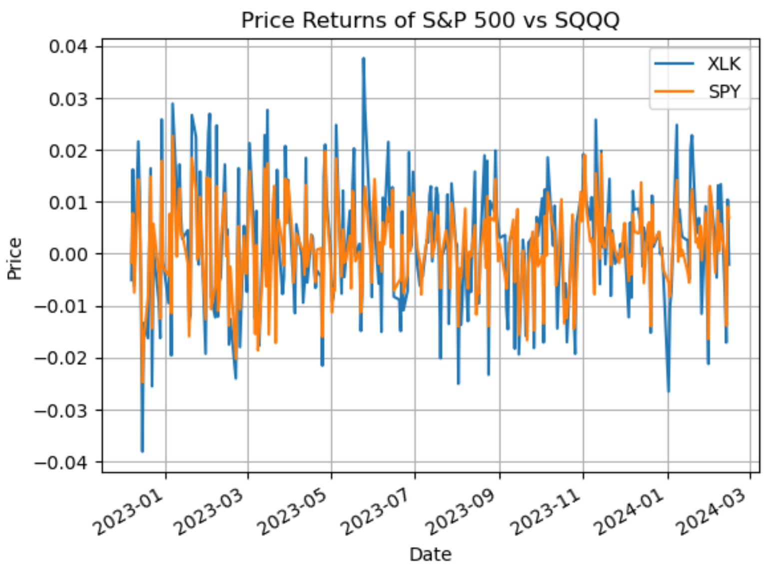

Figure 3.2: Comparision of log returns of S&P 500 and XLK

As observed, the log returns of the two ETFs are almost identical to its linear return. One potential reason would be due to the relatively small price changes.

III. CORRELATION ANALYSIS

1. AUTO-CORRELATION FUNCTION

1.1. Definition

The auto-correlation function (ACF) is used to quantify the correlation between a time series and its own lagged values.
For a stationary time series Xt, the auto-correlation function at lag k is denoted as Pk and is calculated between time series t and $t-k$:

$$\rho_k = \frac{\text{Cov}(X_t, X_{t-k})}{\sqrt{\text{Var}(X_t) \cdot \text{Var}(X_{t-k})}}$$

$Pk = 1$ indicates perfect positive auto-correlation while $Pk = 0$ and $Pk = -1$ indicates no auto-correlation and negative perfect auto-correlation respectively.

1.2. Auto-correlation of price time series

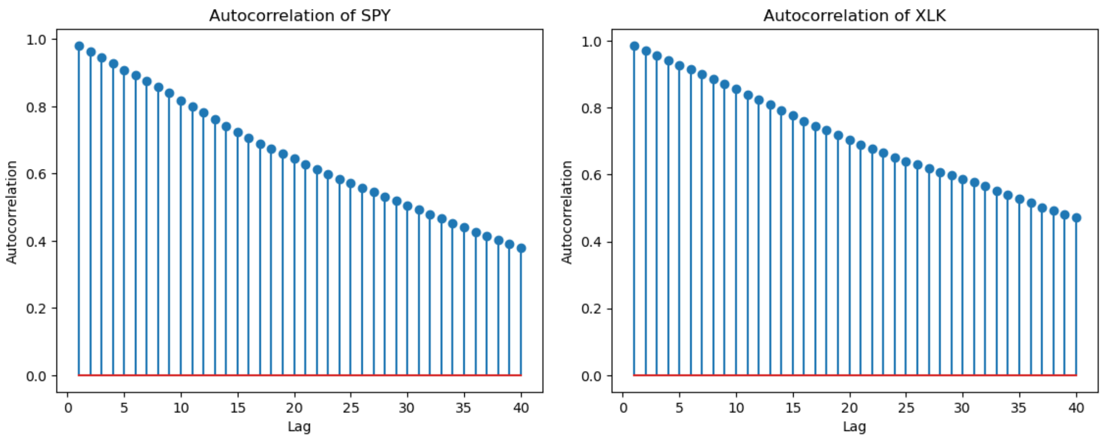

Figure 1.1: Auto-correlation of prices of S&P 500 and XLK lag = 40

Figure 1.1 indicates of non-stationarity rather than a sign of interesting auto-correlation.

2. PARTIAL AUTO-CORRELATION FUNCTION

Partial auto-correlation functions between time series Yt and $Yt + k$ is the auto-correlation between them that is not accounted for any smaller lag.
Figure 2.3 shows that auto-correlation when lag equal to 1 is statistically significant for both ETFs and correlation tends to zero past the order one of the model. Therefore a auto regressive model of lag 1 is best suit for modelling SP 500 and Technology Select Sector Fund.

The auto correlations of returns are insignificant. This is expected for most asset return, except for very small intraday time scales for which microstructure effects come into play.

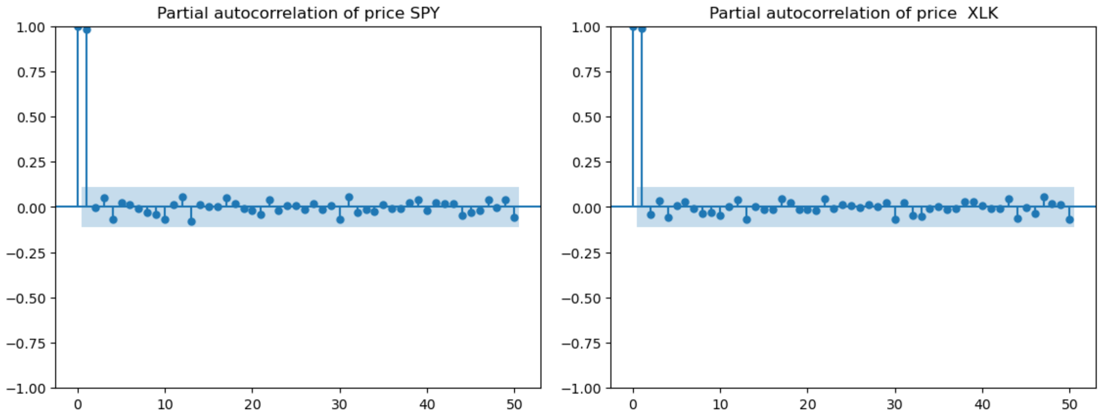

Figure 2.1: Partial auto-correlation of prices of S&P 500 and XLK

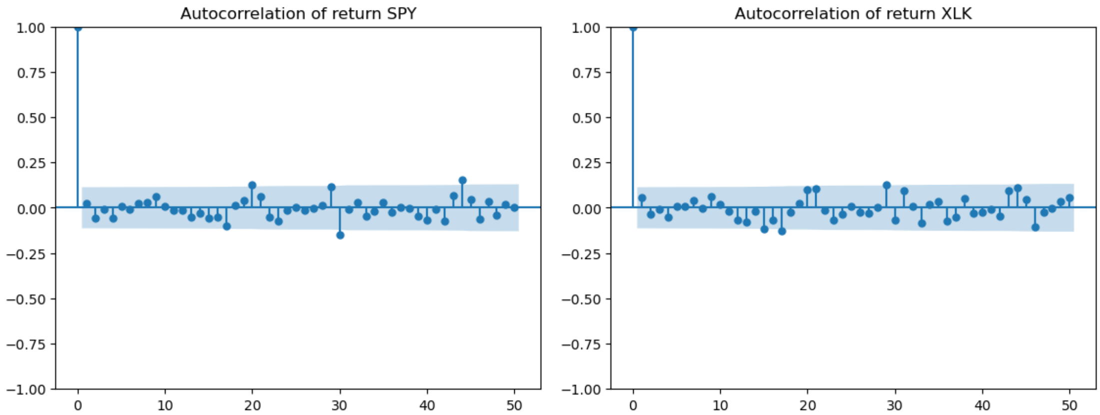

Figure 2.2: Auto-correlation of return of S&P 500 and XLK

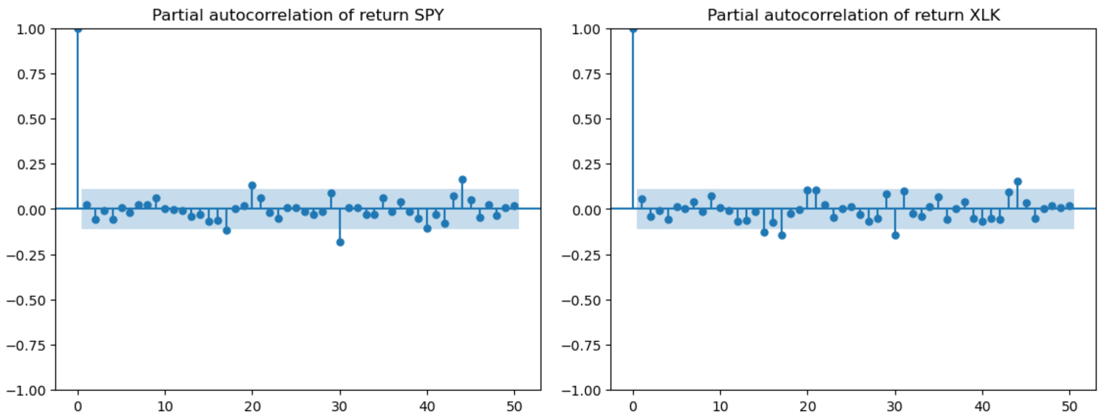

Figure 2.3: Partial auto-correlation of return of S&P 500 and XLK

IV. GAUSSIANITY AND STATIONARITY TESTS

1. GUASSIANITY TEST

1.1. Kolmogorov–Smirnov test

The Kolmogorov–Smirnov statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution,in this analysis, the Gaussian distribution.
The empirical distribution function Fn for n independent and identically distributed (i.i.d.) ordered observations Xi is defined as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\theta(x - x_i)$$

θ function returns 1 if argument is positive and returns 0 otherwise. The Kolmogorov–Smirnov statistic for a given cumulative distribution function $F(x)$ is:

$$D_n = \sup |F_n(x) - F(x)|$$

sup function is supremum of the distance.
$D - n$ converges to zero if Xi is sampled from $F(x)$ The null hypothesis is that at the sample comes from the hypothesised distribution $F(x)$. It rejects the null hypothesis at level α when $\sqrt{n}D_n > K\alpha$

1.2. Guassianity of EFTs

Kolmogorov–Smirnov test can be used to test gaussianity of return time series.
The log returns of S&P 500 and XLK are tested against the Gaussian distribution and under 95 percent confidence level, null hypothesis is rejected indicating both ETFs are not Gaussian distributed. Specifically, D value for S&P 500 is 0.480 and p value=0.0. D value of XLK is 0.486,p value=1.354e-65.

This result is expected as most asset returns are non Gaussian.

1.3. Q-Q plot

Q-Q plot is plotted to further explore the of return time series of chosen ETFs. The Q-Q plot compare the quantiles of the return time series of S&p 500 against the quantiles of the theoretical Gaussian distribution. To allow more comprehensive analysis of the distribution, The length of time series is increased to cover December 2017 to 2023.

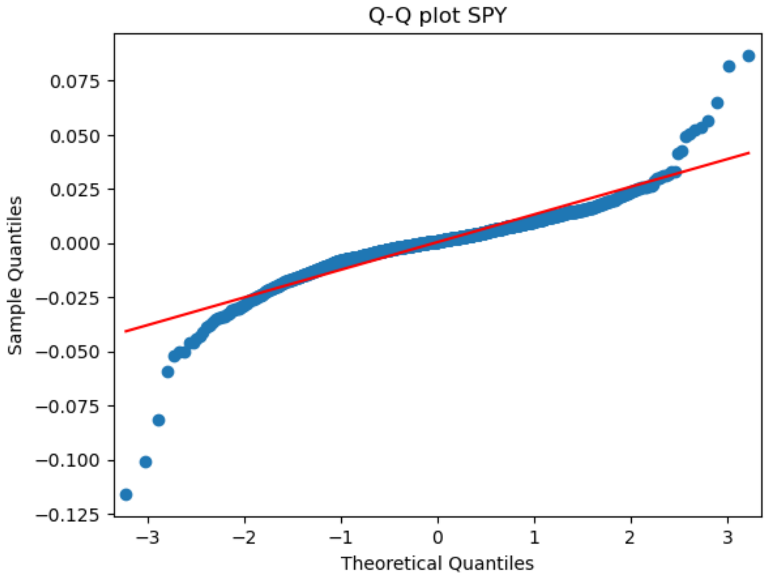

Figure 1.1: Q-Q plot of SP 500 return time series

As observed , the Q-Q plot suggests that return price series of S&P 500 follows a fat tailed distribution.

2. STATIONARY TEST

2.1. Augmented Dickey-Fuller

The Augmented Dickey-Fuller (ADF) test is a statistical test used to determine whether a unit root is present in a time series dataset. The presence of a unit root suggests that the time series is non-stationary. The ADF test look for significance of Mean-reversion coefficients γ in regression equation:

$$\Delta y_t = \alpha + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \ldots + \delta_k \Delta y_{t-k} + \varepsilon_t$$

The null hypothesis is that γ=0 means that the lagged value $y_t - 1$ does not have a significant effect on the change in the time series , indicating the potential presence of a unit root and non-stationary.

2.2. Stationary of return time series

ADF test is used to test the stationary of return time series. The p value for returns of S&P 500 is 3.53e-22 therefor the null hypothesis is rejected and The log return of SPY is stationary. The p value for returns XLK is 3.70438e-29 therefore the null hypothesis is rejected and the log return of XLK is stationary

V. COINTEGRATION TESTS

1. ENGLE-GRANGER TEST

Two time series are said to be co-integrated if a linear combination has a lower level of integration. Let 2 time series Yt and Xt be integrated of order one, or I(1), meaning both of them have unit root and are non stationary. Let there be a θ such that Z$_t$ =Yt$-\theta Xt$ and Z_t is integrated of order zero. This proves Yt and Xt is cointegrated.
To test cointegration, First find a suitable θ through ordinary least square. Then perform a unit root test on Z_t. If Z_t is integrated of order zero,The two time series is co integration.

1.1. Engle-granger test for price time series

First step is to test whether price time series of S&P 500 and XLK are integrated of order one. After performing Augmented Dickey-Fuller tests, both price time series are proved to be I(1). Next step is to compute the best θ coefficient using ordinary least square method. The numpy library function np.linalg.lstsq() is used to find ordinary list square. Than calculate the residual and perform a ADF test usding previously implemented method.
The p value of residual's ADF test is 0.9526. Since 0.9526 is larger than 0.05, there is not enough evidence to reject the null hypothesis and residual is of I(1)

This show no cointegration between S&P 500 and XLK. However,the ADF test is known to have low power (Type II error, a false negative, i.e., the null of unit root is false in reality, meaning that the times eries is known to mean-revert but the test tends to accept the null) Therefore for further investigation, a cross check with another type of stationarity test should be performed.

1.2. Engle-granger test for Return time series

Is it not possible to compute any cointegration test for return time series of S&P 500 and XLk as both of them are proved to be stationary by the ADF test. Cointegration is designed to address the issue of non-stationarity in time series. If variables are already stationary, there is no need for cointegration testing because they don't exhibit the long-term dependence that cointegration captures.