

Exercise 2

Machine Learning in Finance with Python (ECON5130)

Richard Foltyn
University of Glasgow

Deadline: November 17, 12:00

1 Daily returns of US stock market indices

In this exercise, we examine how the three major US stock market indices performed this year using data from Yahoo! Finance.

This exercise can be downloaded as a Jupyter notebook from [GitHub](#). You can use this notebook to get started when typing up your solution.

1.1 Download data from Yahoo! Finance

Use the `yfinance` library to download the time series of daily observations for the [S&P 500](#), the [Dow Jones Industrial Average \(DJIA\)](#) and the [NASDAQ Composite](#) indices. Restrict the sample to observations to the period from 2022-01-01 to 2022-09-30.

Hint: The corresponding ticker symbols are `^GSPC`, `^DJI`, `^IXIC`, respectively.

Rename the DataFrame columns to `'SP500'`, `'Dow Jones'` and `'NASDAQ'`.

Hint: You may need to first install `yfinance` as follows (in particular when running on Google Colab):

```
[1]: # Uncomment this to install yfinance in your Python environment
     #! pip install yfinance
```

1.2 Plot daily closing prices

For the remainder of the exercise, we focus on the price at close, i.e., the `'Close'` column. Create a DataFrame that contains only the values from this column (for each ticker symbol) and plot the three time series in a single graph. Label all axes and make sure your graph contains a legend.

Hint: You can directly use the `DataFrame.plot()` method implemented in pandas.

1.3 Plot normalised daily closing prices

The graph you created in the previous sub-question is not well-suited to illustrate how each index developed in 2022 since the indices are reported on vastly different scales (the S&P500 appears to be an almost flat line).

To get a better idea about how each index fared in 2022 relative to its value at the beginning of the year, normalize each index by its value on the first trading day in 2022 (which was 2022-01-03). Plot the resulting normalised indices.

1.4 Compute and plot the daily returns

For each index, compute the daily returns, i.e., the relative change vs. the previous closing price in percent.

Create a plot of the daily returns for all indices.

1.5 Plot the distributions of daily returns

Compute the average daily returns and the volatility (standard deviation) for each index. Create a histogram of daily returns for each index using 25 bins (i.e., create a figure with 3 panels). For each histogram, add the density of a normal distribution that has the same mean and variance.

Hint: You can either use `DataFrame.hist()` to plot the histogram, or Matplotlib's `hist()` function. In either case, you should add `density=True` such that the histogram is appropriately rescaled and comparable to the normal density.

Hint: Use the `pdf()` method of the `scipy.stats.norm` class to compute the normal density.

1.6 Compute and plot the pairwise correlations

Compute the pairwise correlations between the daily returns of each index pair. Create a 3-by-3 graph where each panel contains a bivariate scatter plot of the daily returns of one index vs. another.

Hint: You can use the function `scatter_matrix()` to accomplish this task. Alternatively, you can create a figure with 3-by-3 subplots using Matplotlib's `subplots()`, iterate over all rows and columns and add a `scatter()` plot to each axes object.

2 Long-run returns of the S&P 500

For this question, we restrict our attention to the S&P 500 and examine its long-run annual returns.

2.1 Download historical S&P 500 data

Download the historical time series of the S&P 500 up to 2021-12-31 using `yfinance` (the earliest available observation is from 1950). Rename the ticker symbol column name to `SP500`.

2.2 Aggregate to annual data

Discard all columns other than `Close`. For each calendar year, select the last observation to get an annualised time series.

Hint: You can use the `resample()` method to aggregate by year.

2.3 Compute and plot annual returns

Compute the annual returns of the S&P 500, defined as the relative change from the previous-year's closing price in percent.

Plot the resulting annual return time series. Indicate the average annual returns over this period using a horizontal line in this plot.

What are the mean and standard deviation of annual returns?