

Data Analysis Assessment: Covid Restrictions

Assessment Format Instructions

Many questions will ask you to calculate average returns for a given period of time or country. Please read the instructions below to make sure that you answer in the correct format:

- All daily returns are defined using log returns, i.e. the return on day t is defined as:
$$r_t = \log(\text{price}_t) - \log(\text{price}_{t-1})$$
, where price_t is the close price on day t .
- Unless otherwise specified, please provide your answer in percentages and round to two-decimal precision: for example, a return of 4.3163% should be entered as "4.32".
- Use negative numbers for negative returns: for example, if you are asked to compute the market return and the market dropped by 6.152% you should enter "-6.15".
- Some of the questions ask about *cumulative* returns over several days. You can calculate the cumulative returns using the following simple formula since we are using log returns:

$$r_{cum} = \sum_t^{t+j} r_t.$$

- Some of the questions ask you to calculate a *simple average* of returns, while others ask for *weighted average* returns. It is often good practise to calculate weighted averages to account for the fact that bigger companies are more important to the economy. For this purpose, we include a file "wgt.csv" with weights derived from a \log_{10} of the market capitalisation. When computing a weighted average, note that the weights must sum to 1, so you may need to renormalise the weights when computing a weighted average for a subset of stocks.

Data Overview

The folder contains the following 6 csv files:

File	Description
companies.csv	Table with information on between 400 and 500 companies (see details below).
restrictions.csv	Table with information on Covid restrictions (see details below).
vaccines.csv	Table with information on Covid vaccines (see details below).
ret.csv	Table with daily log returns of companies, calculated as $\log \log(\text{price}_t) - \log \log(\text{price}_{t-1})$. The first column contains the dates, and the subsequent entries in the top row show the tickers. You can assume that all dates are valid trading days.
cap.csv	Matrix with market capitalisation of companies in USD. The first column contains the dates, and the subsequent entries in the top row show the tickers. Again, you can assume that all dates are valid trading days. Indeed, the dates (first column) and tickers (first row) will be the same as those in ret.csv.
wgt.csv	Matrix with weights, derived from $\log_{10}(\text{cap})$. The first column contains the dates, and the subsequent entries in the top row show the tickers. Again, you can assume that all dates are valid trading days. Indeed, the dates (first column) and tickers (first row) will be the same as those in ret.csv.

The companies.csv file contains the following columns:

Column Name	Description
ticker	The stock ticker of the company. For the purpose of this exercise, tickers are defined as numbers increasing from 1 that are unique to individual companies.
company	The name of the company.
country	The country where the company is listed.
industry	The industry classification of the company.

The restrictions.csv file contains the following columns:

Column Name	Description
country	Country that introduces the restriction.
measure	The name of the restriction that was taken.
definition	A brief description of the measure.
date_start	The first date that the restriction was in place. In all cases, assume the announcement of the restriction was made before the market open on that day. So if the restriction was announced on day X, the effect on the market will be seen in the return on day X.
date_end	The last date that the restriction was in place (if applicable).

The vaccines.csv file contains (amongst others) the following columns:

Column Name	Description
name	The name of the vaccine candidate.
sponsor	The funding source of the vaccine.
study	A brief description of the vaccine study.
resultsDate	The date that the first interim Results from Phase III trials were published. In all cases, assume the results were published before the market open on that day. So if the results were announced on day X, the effect on the market will be seen in the return on day X.