

Data Sources

Bernt Arne Ødegaard

24 November 2021

1 Data

1.1 Obtaining data

Where can you go to get data?

Data needs to be based on the economic question to be investigated, but some obvious sources.

Split into

- Public data
- Proprietary data

Public data In this internet age more and more data becomes publicly available, but same old truths should guide your search for data.

Unless there is some reason to trust the data source, be very wary of how data is collected and its quality.

Why is the data provider putting out data?

- Public organizations (central banks, statistic organization)
- Academic researchers
- Companies doing business

Proprietary data

Tends to be expensive....

But for many purposes necessary to have high quality data.

Some well known databases

- CRSP/Compustat/WRDS
- EUROFIDAI (not so well known yet)
- Datastream
- Oslo Stock Exchange – OBI

1.2 Acknowledging data sources

Data, tell people where you got it

In any empirical work you should always be very clear about what data you used and where you got it from.

Ideally: allow other people to replicate

At least: let others understand the limitations of the data.

1.3 Data transforms

Data, working with and transforming

Most data must be transformed and adjusted before analysis is done.

Hard Work, eats up most of the empirisist time.

Computing tools: Lots of

- General tools,
 - Spreadsheets(Excel, Gnumeric)
 - Database programs
 - Matrix tools (Matlab, Octave, Ox)
 - General purpose programming languages: C, C++, Perl, Python, Java...
- Econometric programs, typically also allow data manipulation
 - R, Stata, SPSS, SAS ...

Exercise 1.

S&P 500 [2]

Gather data for the stock market index S&P 500, for example from Datastream or Yahoo Finance, for as long time period as you can.

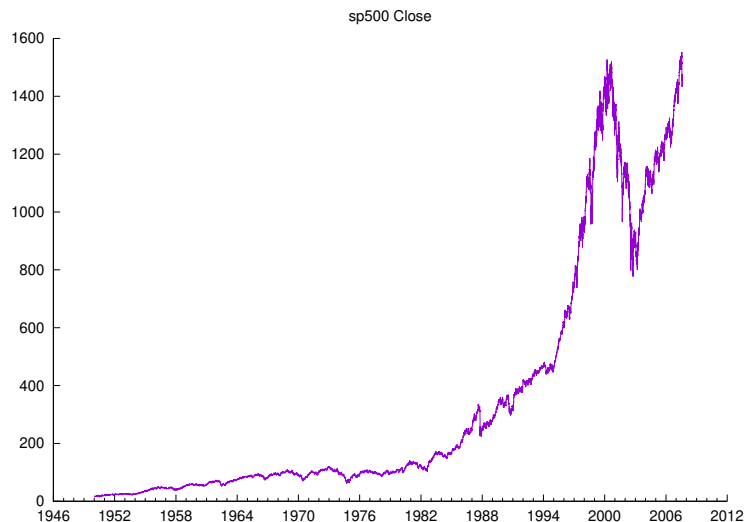
- Plot the time series of the stock market index.
- How would you use the stock market index to estimate the monthly return from investing in the stock market?
- Suppose you are an investor in Norway. Does these returns express the return you would have had from investing in the US stock market?

Solution to Exercise 1.

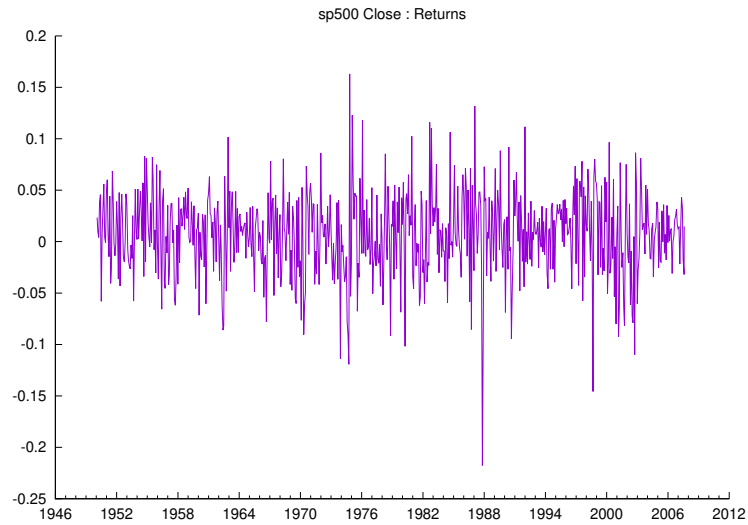
S&P 500 [2]

This data was pulled from Yahoo Finance.

- Time series plot of the S&P 500 index.



- The return of the index between dates t and $t + 1$ can be thought of as buying the index at time t and selling it at time $t + 1$. If I_t is the index level at time t , the return is $R_{t+1} = \frac{I_{t+1} - I_t}{I_t}$ for each observation. To find monthly returns, need to have the end-of-month levels of the stock price index.



- A Norwegian investor also has a currency component, because the index is “bought” at time t with the then relevant exchange rate and “sold” at time $t + 1$ at which time the exchange rate has changed.

Exercise 2.

IBM [2]

Yahoo Finance provides long historical time series for US stocks listed at the NYSE. Get the historical data for IBM. Among the data series are two columns, the closing prices, and a set of adjusted prices.

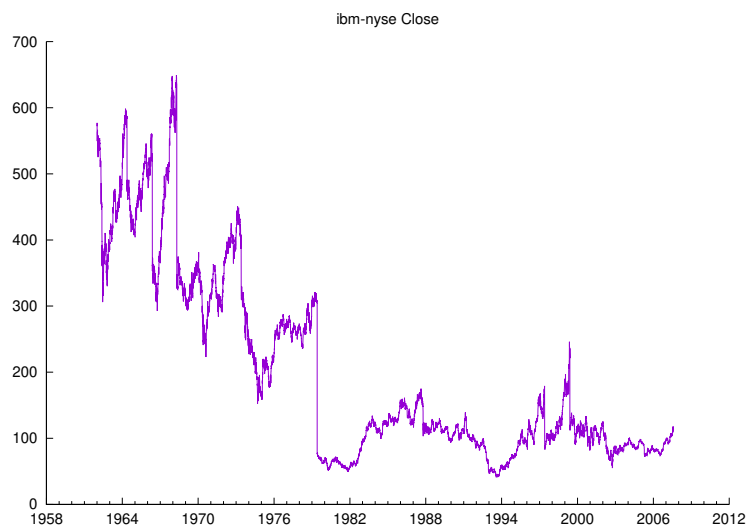
- Plot the two series.
- Which of the two series is relevant if you want to estimate the return to holding IBM stock?

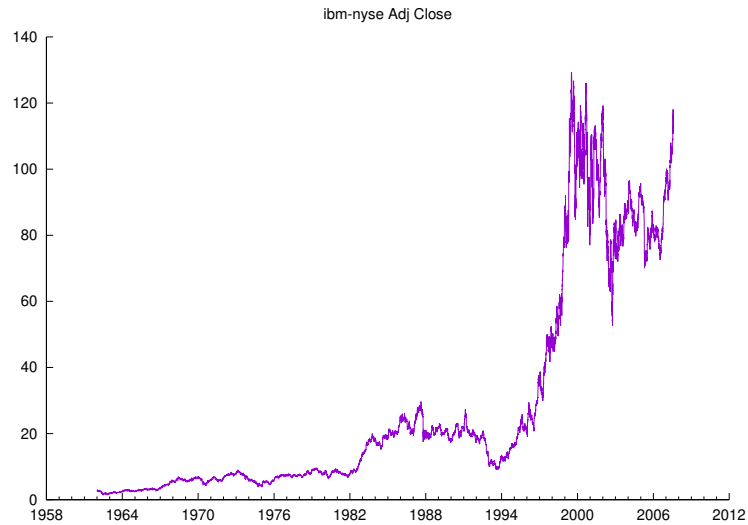
Solution to Exercise 2.

IBM [2]

This data was pulled from Yahoo Finance.

- Plotting the series.



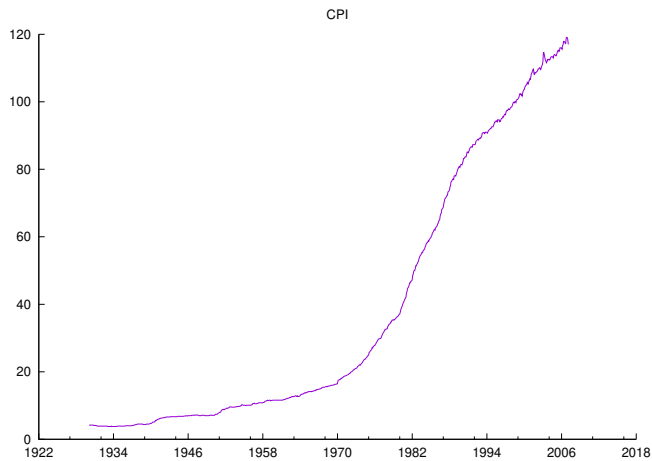


- The adjusted price series is the relevant one for returns calculations, since it accounts for such things as stock splits

1.4 Data classification

- Time series
- Cross sectional
- Panel data (both time series and cross sectional)

Example: Time series: The consumers price index for norway



Example: Cross sectional: Economists Big Mac index

Example: Panel Data: Stock returns for the cross section and time series.

The distinction Quantitative/Qualitative data

Quantitative: Observations that are in numerical units.

Qualitative: Observations that are not in numerical units. To work with econometrically necessary to do transformations.

Index numbers.