

Data, where to get it?

Some pointers

- ▶ Public data

Internet, lots of stuff there, but always question quality.

Who is the data provider?

Public organizations (Central Banks, Statistics agencies, ...)

- ▶ Proprietary data

Quality controlled data can be very expensive....

- ▶ CPSP/Compustat - WRDS
- ▶ EUROFIDAI
- ▶ Datastream.
- ▶ Oslo Stock Exchange – OBI

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.

Data, working with and transforming

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

Computing tools: Lots of

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

General advice: Know at least one tool very well, but be able to switch when the situation demands.

Exercise *S&P 500*

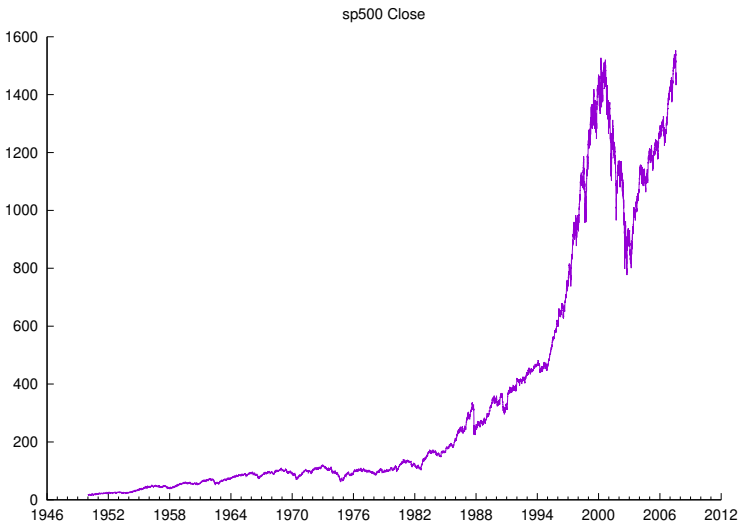
Gather data for the stock market index S&P 500, for example from Datastream or Yahoo Finance, for as long time period as you can.

- ▶ 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

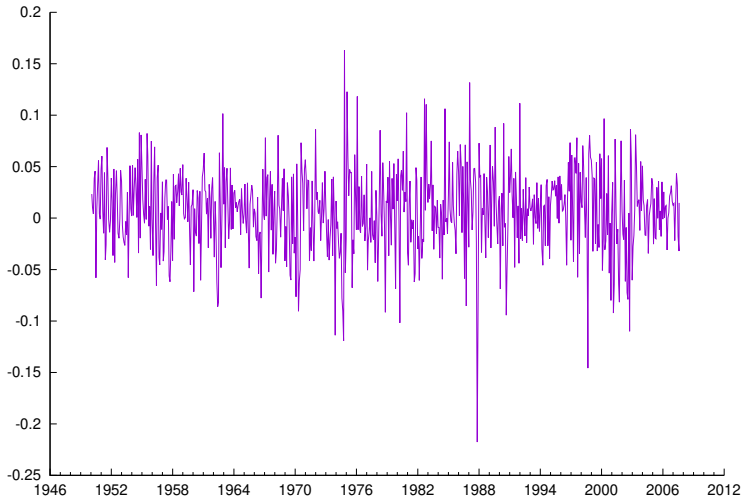
This data was pulled from Yahoo Finance.

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.

sp500 Close : Returns



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 *IBM* [2]

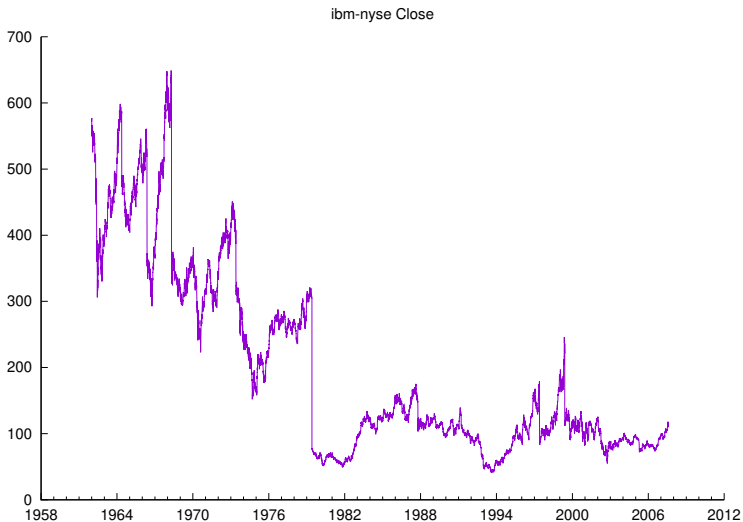
Yahoo Finance provides long historical time series for US stocks listed at the NYSE. Get the historical data for IBM. Among the data series are two columns, the closing prices, and a set of adjusted prices.

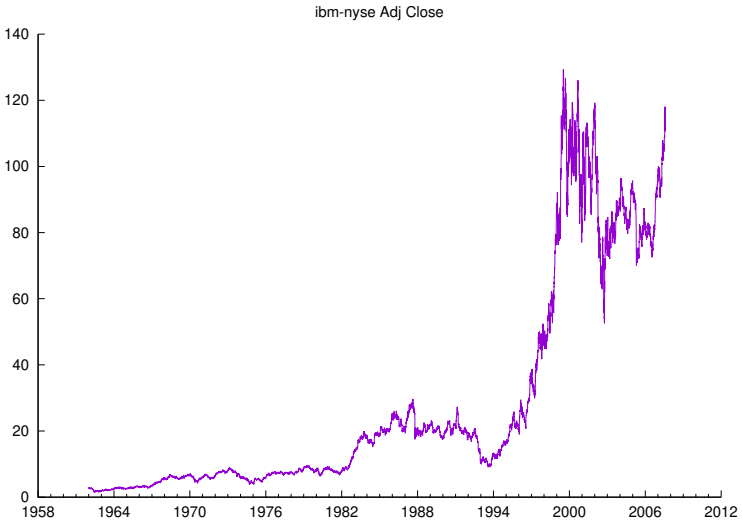
- ▶ Plot the two series.
- ▶ Which of the two series is relevant if you want to estimate the return to holding IBM stock?

Solution *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

Classifying data

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

The distinction Quantitative/Qualitative data
Index numbers

CPI Norway

