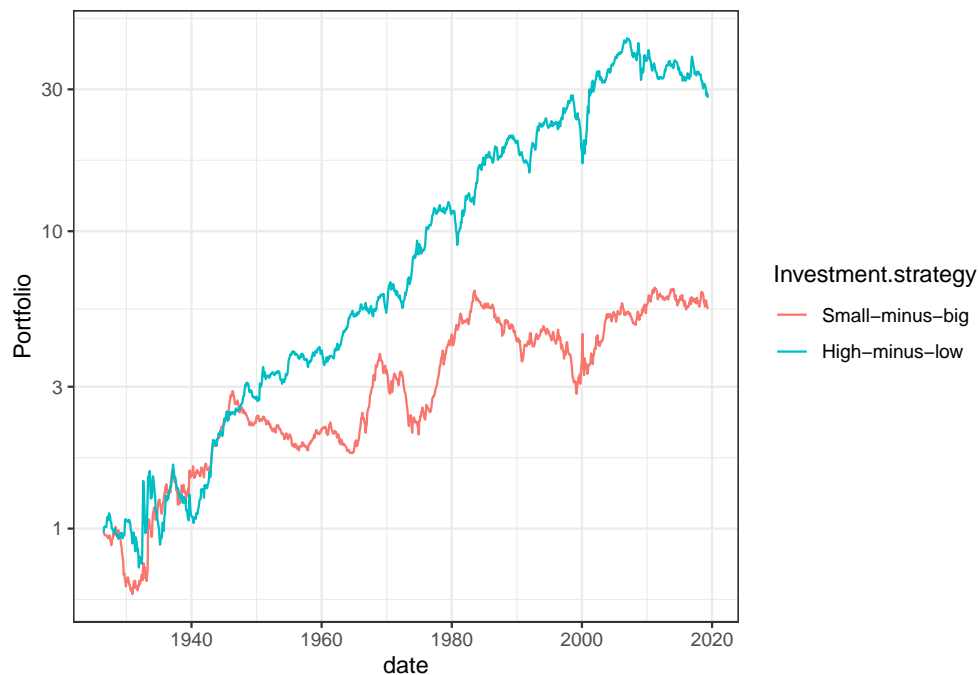


# Financial Econometrics

## Simple Regression

Darya Yuferova

There are very few as well established empirical facts in Finance as (1) small stocks earn a higher return than big stocks and (2) stocks with high book to market (BM) ratio earn a higher return than stocks with low book to market ratio. The plot below exemplifies this by showing the value of two portfolios from 1926 to today. The size of the portfolio at the beginning is 1 USD. The red (blue) portfolio invests in small (high BM) by shorting big (low BM) firms. Note that the plot uses log scales for the y-axis.



In this lab session, you will use data on S&P500 firms from 2010 to 2015 to revisit the aforementioned findings in this particular dataset. This is interesting because, as you may recognize in the plot, the pattern seemed to have disappeared or even reversed in the most recent times. In particular, using the fiscal year 2014, you want to estimate following two regression equations:

$$Return_{Next\ fiscal\ year} = \alpha_0 + \beta_1 \log(BM) + u$$

$$Return_{Next\ fiscal\ year} = \alpha_0 + \beta_1 \log(MKTV_{End\ of\ current\ fiscal\ year}) + u$$

This exercise takes you through the standard steps in econometric analysis: load data, construct variables, limit the sample, and interpret the result.

You are provided with a csv file *Data\_lab\_simple\_regression.csv* that includes following variables:

- lpermno: firm identifier
- datadate: date when the data was supplied to authorities
- fyear: fiscal year
- BE: book value of equity
- MKTV.june.of.current.fiscal.year, MKTV.end.of.current.fiscal.year, MKTV.end.of.subsequent.fiscal.year: market value recorded at different times
- FF.49: industry classification following Fama and French's 49 industries

Your analysis proceeds in following steps:

- Task 1: Load the data into R.
  - Investigate the structure of the data using the functions *str()*. Are all variables of the correct object type? If not, please readjust them [optional].
  - Summarise the data using the function *summary()*
- Task 2: Construct new variables
  - Book to market is defined book value of equity divided by market value as of June of the current fiscal year
  - Return during the next fiscal year is defined as the difference of log market capitalization of end of current and end of subsequent fiscal year
- Task 3: Estimate the regression equations as outlined above, using only the fiscal year 2014. Note that you can use the logarithm function *log()* within the regression function *lm()*.
- Task 4: Having estimated both models, combine them in a output table with the package *stargazer*. Google for additional information.
- Task 5: Complete following sentences for each model,
  - An increase in  $\log(\text{BM})$  of 1, leads to a ... of ... in returns at the ... significance level.
  - An increase in  $\log(\text{MKTV})$  of 1, leads to a ... of ... in returns at the ... significance level.
- Task 6 [optional]: You want to know, if the estimated coefficient  $\hat{\beta}_1$  for the link between BM and future return depends on the year the regression is estimated for. Use following coding strategy:
  - Declare a 5x4 matrix. Each row of this matrix will be populated by the regression output for each fiscal year.
  - Write a loop that iterates over the fiscal years 2010 to 2014
  - For each iteration estimate the regression in Task 3
  - Store the estimated coefficient  $\hat{\beta}_1$  on  $\log(\text{BM})$ .
  - Inspect the output. Does the relationship between BM and future return seem to be robust?