Computational Statistics
Bootstrapping non i.i.d. data

1 Bootstrapping regression

We consider again the data in the file `investment.txt`, which contains 15 yearly observations of U.S. investment data for the period 1968-1982. The variables are

- `Year` = Date,
- `GNP` = Nominal GNP,
- `Invest` = Nominal Investment,
- `CPI` = Consumer price index,
- `Interest` = Interest rate,
- `Inflation` = rate of inflation computed as the percentage change in the CPI.

We consider the linear regression model $\mathbf{y} \sim \mathcal{N}_n(X\boldsymbol{\beta}, \sigma^2 I_n)$ with the dependent variable `Invest/(10*CPI)` and, as covariates, the time trend (a vector of integers from 1 to 15), `GNP/(10*CPI)`, `Interest` and `Inflation`.

1. Compute the least-squares estimates of the regression coefficients as well as 95% confidence intervals based on the normal theory (use the function `confint`).

2. Install the package `boot`. Using the functions `boot` and `boot.ci`, compute 95% bootstrap confidence intervals on the regression coefficients by case-based resampling. (Use the normal-theory and percentile methods to construct the confidence intervals).

3. Compute 95% bootstrap confidence intervals on the regression coefficients using the model-based approach (bootstrapping the residuals).

4. Compare the different confidence intervals obtained and draw some conclusions.

2 Moving-block bootstrap

We consider consider the wage-productivity data `wages.txt` from the book "Basic Econometrics" by D. N. Gujarati, McGraw-Hill, 4th edition, 2003. These data consist in indexes of real compensation per hour (Y) and output per hour (x) in the business sector of the U.S. economy for the period 1959 to 1998. The base of the indexes is 1992=100. We consider the following model:

$$Y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \epsilon_t. \quad (1)$$

1. Plot the data.
2. Compute the least-squares estimates of the regression coefficients as well as 95% confidence intervals based on the normal theory (use the function `confint`).
3. Plot the residuals. Using the function `dwtest` in the package `lmtest`, apply the Durbin-Watson test. What do you conclude?
4. Using the function `tsboot` in the package `boot`, resample the residuals using the moving-block bootstrap method. Compute 95% bootstrap confidence intervals on the coefficients. Compare them to the conventional ones.

3 Maximum-entropy bootstrap

Install the package `reboot`. Using the command `data("USconsum")`, upload the time-series data about the US consumption and disposable income in the period 1948–1998. Let c_t denote the logarithm of the US consumption and y_t the logarithm of disposable income. We consider the following regression equation

$$c_t = \beta_1 + \beta_2 c_{t-1} + \beta_3 y_{t-1} + \epsilon_t$$

and we wish to test the null hypothesis $\beta_3 = 0$.

- Plot the data.
- Using the function `dynlm` in the package `dynlm`, compute the least-squares estimates of the regression coefficients and a 95% confidence interval on β_3 (use the function `confint`).
- Using the function `meboot`, resample the times series (c_t) and (y_t) by the maximum-entropy bootstrap method. Compute a confidence interval on β_3 using the percentile method, and compare them to the previous conventional intervals.