Coding Exercise

Model Risk Management & Control

Summary Instructions

This coding exercise asks you to perform four tasks related to a regression. You may use any language (no spreadsheet program) that you prefer. The only requirement is that the analysis is written in a manner such that it can be executed to reproduce your output. The code needs to be sufficiently commented so that another person can understand the thought process behind the analysis.

Description of Data

The data set contains quarterly time series on 12 U.S. macroeconomic variables for the period from 1950 until 2000. The variables are:

- gdp: Real gross domestic product
- consumption: Real consumption expenditures
- invest: Real investment by private sector
- government: Real government expenditures
- dpi: Real disposable personal income
- cpi: Consumer price index
- m1: Nominal money stock
- tbill: Quarterly average of month end 90 day treasury bill rate
- unemp: Unemployment rate
- population: Population
- inflation: Inflation rate
- interest: Ex-post real interest rate

The data is contained in a pipe separated txt file, i.e. the symbol — indicates different columns. The first column contains the dates, the remaining columns are the time series. NA is used to indicate missing values.

Description of Tasks

Assume that you want to challenge the following regression specification to model the change in consumption:

$$D(\text{Consumption})_t = \alpha + \beta_1 \; D(\text{DPI})_t + \beta_2 \; D(\text{Unemployment})_t + \epsilon_t$$

Hereby, $D()$ denotes first differences, α the intercept, and β_1 and β_2 the slope coefficients of the respective independent variables. ε_t are the residuals. The estimation results of the model are:

Table 1: Fitted linear model

	Estimate	Std. Error	t value	$\Pr(>\|t\|)$
D(DPI)	0.3557	0.04778	7.444	2.844e-12
D(Unemp)	-16.01	3.792	-4.223	3.657e-05
(Intercept)	16.28	1.911	8.522	3.785e-15

Task 1: Replication of Results

Verify the results of the regression estimation using your own implementation; i.e. load the data from the txt file which was provided with this exercise, perform any necessary variable transformations, and compute the regression coefficients, standard errors, t-statistics, and p-values. Evaluate the accuracy of the regression.

Task 2: Outlier Detections

Examine the time series of the residuals for outliers. Use Tukey's test for this, i.e. observations are considered outliers if they fall outside of the following interval:

$$[Q_{0.25}(x) - 1.5 \times \text{IQR}, Q_{0.75}(x) + 1.5 \times \text{IQR}]$$

where $Q_p(x)$ denotes the pth quantile of x and IQR denotes the interquartile range. Comment on the number of outliers and what this implies for the interpretation of the regression results.

Test 3: Autocorrelation of Residuals

Calculate the Durbin-Watson (DW) statistic of the residuals. The DW statistic is defined as:

$$DW = \frac{\sum_{t=2}^{T} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{T} \varepsilon_t^2}$$

where T is the length of the sample and ε_t are the residuals of the regression above. Comment on the impact of autocorrelation of the errors on the regression.

Task 4: Bootstrapping of Standard Errors

Recalculate the standard errors of the regression coefficients using bootstrapping. For the bootstrap, generate 10'000 samples, each of size T, from the original data, using random sampling with replacement. When you compare the bootstrapped standard errors to the default standard errors, what differences can you observe? What are the implications?