STOCKHOLMS UNIVERSITET,
MATEMATISKA INSTITUTIONEN,
Avd. Matematisk statistik

Case study 1 of 2 – Econometric methods MT5014

- You should hand in a well-written and self-contained report, which means that one should be able to follow your report without having access to e.g. this document or the lecture notes. The reasoning should be well motivated and the structure clear and easy to follow.

- The case study should be solved individually.

- The report should be written in R Markdown with pdf as output option. Any tables, figures or results should be dynamically computed when compiling/"knitting" the document. Chunks of code should only be visible (echo = TRUE) in the pdf report when explicitly asked for. Both the compiled pdf and the source Rmd-file should be uploaded on the course home page.

- See the course web-page for information about grading. The deadline is available on the course web-page.

Exercise 1

The file AR.txt (available on the course web-page) contains data generated by an AR model. Your task is to figure out the number of lags and the parameters of the AR model. Use the AIC when solving this exercise.

Hint: the number of lags is not above 5.

Include the code for this exercise. Write short informative comments in your code. *It is not allowed to use any* **R** *packages etc in this exercise.*

Exercise 2

A series of U.S. quarterly real gross domestic product (GDP) from 1947 to 2010 (from the Federal Reserve Bank at St. Louis) is available in the file GDP.txt on the course webpage. Consider GDP growth rate i.e. the first difference series of log(GDP), and:

1. Test the null hypothesis $H_0 : \rho_1 = \rho_2 = ... = \rho_{12} = 0$ against the alternative hypothesis $H_1 : \rho_i \neq 0$ for some $i \in \{1, ..., 12\}$ (where we use the notation of Tsay). Interpret the results.

2. Build a suitable AR model. Perform model checking to justify your model; cf. Tsay around page 50.

Exercise 3

The file bwa.txt on the course webpage contains the wages and other data for 474 employees of US bank. Specifically, the file contains the following data for each employee:

Variable:	Description:
SALARY	current yearly salary in dollars
LOGSALARY	natural logarithm of SALARY
EDUC	education (number of finished years)
SALBEGIN	yearly salary at employee's first position at the same bank
LOGSALBEGIN	natural logarithm of SALBEGIN
GENDER	gender (0 for females, 1 for males)
MINORITY	minority (0 for non-minorities, 1 for minorities)
JOBCAT	job category (1 for administrative jobs, 2 for custodial jobs, 3 for management jobs).

Exercise 3.1

Consider a regression model with LOGSALARY as a response variable and EDUC and LOGSALBEGIN as explanatory variables. That is, consider:

$$Y_j = \beta_1 + \beta_2 E_j + \beta_3 B_j + \varepsilon_j, \;\; j = 1, ..., n,$$

where Y corresponds to LOGSALARY, E corresponds to EDUC and B corresponds to LOGSALBEGIN.

1. Estimate the parameters $\beta_i, i = 1, 2, 3$ and their standard deviations.

2. Calculate the coefficient of determination and the adjusted coefficient of determination. Interpret the results.

3. Test for heteroskedasticity.

4. Test for autocorrelation.

Exercise 3.2

Include gender and minority as additional explanatory variables to the model in the exercise above, and regress Y on a constant together with these four explanatory variables.

1. Observe that if you add variables, the coefficients of the variables E and B change in comparison to Exercise 4.1. Explain why.

2. Test joint significance of gender and minority.

3. Test whether the coefficient for education is different from zero. What conclusions can you draw based on this?

4. Test the significance of the regression.