Advanced Econometrics: Homework 1

2020

Instructions

- **The solutions should be sent to Lenka Nechvátalová: 60374828@fsv.cuni.cz, with the following subject:**
  'JEM005 HW1 2020: Group surname1, surname2, surname3'
- Form groups of three yourself.
- The empirical problems do not necessarily have a unique solution in terms of numbers, you are assessed based on the execution of the analysis not on the right numbers that you should get from the output. The emphasis is put mainly on the meaningful presentation and the extent of your knowledge.
- Send me the solution as **one Jupyter notebook** (.ipynb), named **'JEM005_2020_HW1_surname1_surname2_surname3.ipynb'** which should contain the main analysis together with commented code. Do not forget to use full potential of Jupyter notebooks, show graphs and results as output of your code, write the reasoning and other text in markdown cells (which supports headers, Latex equations, etc.). It is also possible to send me one pdf file with the analysis and one commented R script (following the naming convention).
- Be concise (no lengthy essays please). Although, be sure to include all important things as I cannot second-guess your work.
- The problem set is due on **5th November**. A late submission automatically means 0 points.
- If you have any questions concerning the homework, do contact me by mail and we can set up a consultation. Do it rather sooner than later, I won't give any consultation concerning the homework from the 3rd.

Problem 1

Weibull distribution has the following density:

$$f(x) = \alpha\beta x^{\beta-1}e^{-\alpha x^{\beta}}, x \geq 0, \alpha, \beta > 0$$

(a) obtain log-likelihood function for a random sample of n observations

(b) obtain the necessary condition equations for α and β. Note that we get the solution for α in terms of data x and β. While for the second equation we get only an implicit solution for β. How would you continue to obtain the maximum likelihood estimators?

(c) obtain a Hessian matrix of the log-likelihood with respect to α and β.

You are provided with sample data generated by the Weibull distribution in 'problem_1_dataset.Rdata'.

(d) obtain maximum likelihood estimates of α and β, and estimate the asymptotic covariance matrix for the estimates.

(e) obtain maximum likelihood estimate of α under the hypothesis that $\beta = 1$.

(f) carry out Wald test with $H_0 : \beta = 1$

(g) carry out likelihood ratio test with $H_0 : \beta = 1$

(h) carry out Lagrange multiplier test with $H_0 : \beta = 1$

Problem 2

You are provided data 'problem_2_dataset.Rdata'. The dataset contains the results of World Banks 1997 Vietnam Living Standards Survey. The following variables in the dataset are of interest:

- *lhhexp1:* log of total expenditure over the year,
- *lhhex12m:* log of expenditures on health care,
- *farm:* dummy for living on a farm,
- *urban98:* dummy for living in an urban area,
- *age:* age of respondent,
- *sex:* sex of respondent.

Do the following:

(a) Describe the data statistically. Make some descriptive plots.

(b) Fit a linear model with *lhhex12m* as a dependent variable and the rest of the variables as regressors, include constant. Describe and interpret the results.

(c) Fit quantile regression of the same model specification for quantiles from 5% to 95% by 10% steps and plot the results. Interpret.

(d) Test on 95% confidence level for which quantiles are your coefficients different from OLS coefficients. (Do it visually from the pictures, the results are obvious from them.)

(e) Discuss what information did you get from the quantile regression. Contrast the linear and quantile regression.

(f) We know that quantile regression does not work by considering only some portion of the data (i.e. one decile). Demonstrate (if it is indeed the case) that splitting the data into deciles (based on the response variable) and performing linear regression for each decile separately (incorrect method) leads to different "decile" betas compared to the (correct) betas from quantile regression. For simplicity, consider only one independent variable, *lhhexp1* and intercept. Plot betas from both methods the same way as in the case of quantile regression (confidence intervals are not necessary). Reason why there is a difference between the two methods.

Problem 3

Consider the simple ($y_i = \alpha + \beta x_i + \epsilon_i$) OLS regression and coefficient estimates $\hat{\alpha}$; $\hat{\beta}$ computed from observations $i \in \{1, \ldots, n\}$. Now, consider coefficient estimates $\tilde{\alpha}$; $\tilde{\beta}$ computed on observations $i \in \{1, \ldots, n, n+1\}$.

(a) Derive the difference in the coefficient estimates as a function of sample statistics (moments) of the new sample and the ($\hat{e}_{n+1} = y_{n+1} - \hat{y}_{n+1}$)(the error of the new observation prediction from the old regression).

(b) What is the intuition? How does it behave if $n \to \infty$?

(c) Simulate some data in R and check numerically whether your formula works.