Advanced Econometrics: Homework 1

October 23, 2019

Instructions[1]:

- **Submit one file for each problem. Submit by e-mail to `martin.hronec@fsv.cuni.cz` with the following subject:**
  "AdvEcox HW1 2018: Group surname1, surname2, surname3 ".
- Form groups of three yourself.
- As a solution, provide 3 Jupyter Notebooks with R source-code. Code should be properly commented, interpretations of results as well as theoretical derivations[2] should be written in markdown cells.
- Use "set.seed()" function, so I can replicate your results.
- Be concise (no lengthy essays please). Although, be sure to include all important things as I cannot second-guess your work.
- The empirical problems do not necessarily have a unique solution in terms of numbers, you are assessed based on execution of the analysis not on the *right numbers that you should get from output*. The emphasis is put mainly on meaningful presentation and extent of your knowledge.
- **The problem set is due on 6th November. Late submission automatically means 0 points.**

Problem 1. *For this problem, use data in file hw1_data.csv.*

Columns containing Y in their names correspond to dependent variables (4) and columns containing X are independent variables (2). Using following pairs of independent and dependent variables,(X1,Y1), (X1, Y2), (X1,Y3) and (X2,Y4) do the following:

1. *Estimate beta coefficients using OLS and MLE. Compare estimates from both methods, explain differences (if any). Discuss validity of necessary assumptions in individual cases.*
2. *Show diagnostic plots (for OLS) and interpret them.*

[1] The contact person for this homework is Martin Hronec, the same mail as for submission of homeworks.

[2] If you prefer not to write formulas in LaTeX, you can send PDF with your derivations and interpretations in additional file and R code in Jupyter Notebook.

3. *Use LAD to estimate conditional median and compare it with the estimate of conditional mean from OLS for each of the pairs. Explain differences between them.*

4. *Use quantile regression to estimate conditional quantiles (for tau 0.05,0.25,0.5,0.75,0.95) for each of the pairs and discuss differences across quantiles.*

Problem 2. *On the second seminar, we have talked about heteroskedasticity a lot and had an example where data was generated according to the equation*

$$income_i = \alpha + \beta \cdot educ_i + educ_i \cdot \frac{\epsilon}{10},$$

where $\epsilon \sim N(0, 200), \alpha = 4000, \beta = 200$.

1. *Illustrate theoretically that this heteroskedasticity implies varying slopes in quantiles and elaborate how to distinguish this kind of quantile dependency that arises purely from heteroskedasticity.*

2. *Simulate data from such model and show empirically, that your finding holds.*

   Hint: Derive the quantiles of income for given education first, i.e. $q_\tau(income|educ)$ *and then think how this relates to slope of quantile regression.*

Problem 3.

1. *Simulate 42, 168 and 672 data points from exponential distribution* ($\lambda e^{-\lambda x}$) *with* λ *of your choice.*

2. *Using all 3 samples, fit the* λ *parameter using MLE and exponential distribution as your assumed distribution.*

3. *Then, again using all 3 samples, fit the gamma distribution* ($\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$) *using MLE again. Report both estimates and their standard errors.*

4. *Finally, as the exponential distribution is a special case of gamma distribution, use the three likelihood-based tests you've seen during the lecture as well as seminar to test the null-hypothesis that the data come from exponential distribution (again using all 3 samples).*

5. *Compare results of your tests across all 3 samples and discuss whether they match your expectations.*