Advanced Econometrics: Homework 1
Suggested solution for problematic parts and comments

2023

Problem 1

- You should not be adding the simulated ϵ into regressions.
- In part 2, the specification of the functional relation that you have used in estimation is wrong. This is because there should be mileage^2 instead of mileage. Because of that, coefficients α_0, α_1 will be biased, but others will not be biased. The unbiasedness is due to the fact that all the regressors are uncorrelated. Your conclusion can be easily checked by increasing the number of observations so you can see more clearly which coefficients are biased and which are not. Some of you used RESET test, which will detect the misspecified relationship. This is fine, but it does not exactly answer the question.

Regarding part 5 (Quantile regression):

- You need to take into account confidence intervals when interpreting differences between OLS and quantile regression
- In case you ran the specification without heteroskedasticity: The quantile regression does not provide more information simply because there is no dependence in the quantiles. It is not surprising because the data were generated that way. Otherwise: The mileage^2 or mileage should hint at a difference in quantiles precisely because of the heteroskedasticity.

Problem 2

For the interpretation of OLS and quantile regression results, let's consider urban and total expenditure variables. The dependent variable is log of yearly health expenditures ($ln(y)$).

$\hat{\beta}_{urban} = -0.385, \%\Delta y = 100 \cdot (e^{\hat{\beta}_{urban}} - 1) = -31.95\%$ People living in urban areas are expected to have lower healthcare expenditures by 31.95%. For cases when $\hat{\beta}_t$ is close to zero (between -0.1 and 0.1), then we can also use the approximation and say the effect is $100 \cdot \hat{\beta}_t\%$.

$\hat{\beta}_{0.75,urban} = -0.408, \%\Delta y = 100 \cdot (e^{\hat{\beta}_{0.75,urban}} - 1) = -33.53\%$. This reduction in healthcare expenditures for urban households will be more than 33.53% for approximately 75% of people. This disparity seems smaller in the higher quantiles - for 25% of people living in urban areas, the disparity between urban and farm households' healthcare expenditures will be less than 33.53%.

$\hat{\beta}_{log\_total\_expenditures} = 0.724$. Both the independent variable and dependent variable are in logs. For 1% increase in total expenditures, it is expected that the expenditures on healthcare will be increased by 0.72%.

$\hat{\beta}_{0.25,log\_total\_expenditures} = 0.578$. For 25% of people, the additional 1% increase in total expenditures will increase healthcare spending by less than 0.58%. For 75% of people, it is estimated that they spend more than 0.58% on healthcare.

Similar conclusions apply to other variables. To conclude and be sure that the OLS estimates are statistically different from quantile regression, the confidence intervals cannot overlap.

In part f), some of you missed that you should be splitting the data into deciles (based on the response variable) and performing linear regression for each decile separately.

Problem 3

- True values for this data sample are $\alpha = 0.06, \beta = 1$.
- Lagrange multiplier test statistic:

$$\mathrm{LM} = \left(\frac{\partial \ln L\left(\hat{\theta}_R\right)}{\partial \hat{\theta}_R} \right)' \left[\mathbf{I}\left(\hat{\theta}_R\right) \right]^{-1} \left(\frac{\partial \ln L\left(\hat{\theta}_R\right)}{\partial \hat{\theta}_R} \right)$$

which follows χ^2 with one degree of freedom under the null hypothesis. $\hat{\theta}_R$ are estimates from the constrained model.

- A Good sanity check of your results is that the tests should give you approximately the same p-value. If they don't, you probably have something wrong, as the tests are asymptotically the same in the case of linear restrictions.

General comments

- Common problems: lack of interpretation or incorrect interpretation, using functions from packages without knowing the necessary conditions for proper usage.
- The structure of your work should be clear, logical and readable, with the outputs of your code being informative.

In case you have any questions regarding the first homework, please let me know (lenka.nechvatalova@fsv.cuni.cz), and we can arrange a consultation.