General instructions for the Exam:

Advanced Econometrics I

March 21st, 14:00 h

Hörsaal XXIV

- The datasets are available via the Ilias Platform (folder: Advanced Econometrics (Theory)). If you have problems to download the data from the Ilias platform, you may ask the responsible personal for a USB stick.
- A printout of the problem set (Klausur) and some blank pages of paper will be made available at the beginning of the exam.
- Please make available your responses (solutions) of the cases via Word, Open-Office or some other text editor. You may copy the output of programs like R, Gretl or Stata in your text editor. Please save the complete content of the text editor as a PDF-File.
- If necessary, you can also submit a handwritten solution on the paper that was distributed at the beginning of the exam.
- Start your text file with indicating your name and your student ID.
- The name of your PDF-file should be just your student ID (no names).
- Please upload the PDF-File with your solution to the Ilias-Folder "Submitted files" 60–70 minutes after the begin of the exam. You cannot assume that the exam will be accepted if you upload it later than 70 minutes.
- If you encounter technical problems during the uploading process, you may also submit your PDF-file via email attachment. But you need to make sure that you send the email (see next bullet point) within the time limit of 70 minutes.
- Good luck!

Case 1: Time series regression

The data set `milk.csv` contains time series data of a consulting firm from sales of milk packages in all stores of a super market chain

PRICE: price of the own brand

PRICE_FM: average price of the competitors

PROMOTION: number of stores with promotion

SALE: total sales of milk packages in all stores

The data runs from period 1 up to period 247.

Tasks

Create a new variable that results from multiplying `SALES` with the factor = `MatNo`/7123456, where `MatNo` is your 7-digit student ID. Take the logarithm of the variable and denote it by `log_Sales`.

a) Estimate a demand function with `log_Sales` as the dependent variable and **the logarithms** of `PRICE`, `PRICE_FM` and `PROMOTION`. How much (in percent!) does the
8 sales change if the price of the own brand increases from 0.80 to 1.00 Euro?

b) Why is it problematical to assess the importance of some variable for explaining the dependent variable by its coefficient β? Why is the beta-coefficient a more suitable
10 statistic to assess the importance of the regressor instead? Show that in a simple *standardized regression* the residual variance is equal to the R^2 of the regression.

c) Test the assumption that the error term in the regression in a) is (i) serially uncorre-
12 lated and (ii) normally distributed. Is there evidence for an outlier with $e_i^2 > 10 \cdot s^2$, where s^2 is the estimator for the residual variance?

d) Test for a structural break for the model estimated in a) at period $t = 90$ by using
10 the Chow test. Repeat the test without specifying the break date (Quandt-Andrews test).

e) Estimate the demand equation as a dynamic ADL(2,0) regression model. Estimate
10 the long-run coefficient (elasticity) if the own price increases by 10 percent. Is the long-run effect significant at a significance level of $\alpha = 0.05$?

Case 2: Cross-section regression and nonlinearity

The data set `IVwages.csv` contains data for 1500 U.S. employees that is used to estimate a wage regression. The variables are defined as follows:

nr: Personal number of individual

lwage: logarithm of hourly wage

educ: years of schooling

exper: years of job experience

Inst1: first instrumental variable

Inst2: second instrumental variable

Tasks

Divide the first 3 digits of your student ID by 10,000 and add the result to the variable `lwage`. This variable is called `lwage_new`.

a) Run a pooled OLS regression of `lwage_new` on `educ`, `exper` and a constant. Compare the regression with an alternative specification, where `exper` is replaced by
10 log(`exper`). Which model specification do you prefer? How large is the change in the wage (in U.S. Dollar) given a change from `exper`=10 to `exper`=11 in both model specifications?

b) Apply a test for heteroskedasticity to the second model specification with log(`exper`).
10 Do you have an explanation why the result is quite different for the Breusch-Pagan and the Koenker test?

c) The theoretical model suggests that the coefficient of `exper` depends on `educ` in the following way:

10

$$\text{wage}_i = \beta_0 + \beta_{1i}\ \text{exper}_i + \beta_2\ \text{educ}_i + \varepsilon_i$$
$$\textit{where}\ \ \beta_{1i} = \gamma_0 + \gamma_1\ \text{educ}_i + v_i$$

Estimate the coefficient γ_1 by constructing a suitably interaction term and show that the errors of the resulting regression are heteroskedastic by construction.

d) Your colleague argues that `educ` is endogenous as this variable is related to the skills and intelligence that also affect the career perspectives (and therefore the
10 wages) of the employees. She therefore suggests to include `Inst1` (an indicator for practical skills) and `Inst2` (IQ - 100) as additional regressors. Perform this proposed regression and comment on the results. Under what conditions the inclusion of the two variables solve the problem of estimating the coefficient of `educ` consistently?

e) Compute the 2SLS estimator for the model specification used in b) by estimating
10 both stages separately. What is the R^2 of the first-step regression? Why are the coefficients of the second stage identical to the IV estimator, whereas their standard errors differ?