Advanced Econometrics – Lab 01

1 Undergraduate Econometrics

Exercise 1
Let's use the data file *cps_small.csv* that comes from the book *Principles of Econometrics*[1]. We are interested in estimating an econometric model for logarithm of hourly wage (*lnWAGE*). List of regressors consists of:

- *educ* - years of education,
- *female* - dummy variable coding sex, 1 for women and 0 for men,
- *black* - dummy variable, 1 for blacks,
- *femaleXblack* - interaction between *female* and *black*.

Estimate the model and interpret the output.

Exercise 2
Use data on houses sold in Stockton, California, to estimate a model for logarithm of price (*lnPRICE*). As regressors use total square feet of living area (*sqft*), number of full baths (*baths*), info on whether it is vacant Yes = 1 No = 0 (*vacant*), whether it has two floors (*stories2*), and *vac_sto2* that is interaction term between *vacant* and *stories2*.

a) Estimate the model:

$$lnPRICE = \beta_0 + \beta_1 sqft + \beta_2 baths + \beta_3 vacant + \beta_4 stories2 + \varepsilon$$

b) Estimate the model:

$$\begin{aligned} lnPRICE &= \beta_0 + \beta_1 sqft + \beta_2 sqft^2 + \beta_3 baths + \beta_4 vacant + \\ &+ \beta_5 stories2 + \beta_6 vacant \times stories2 + \varepsilon \end{aligned}$$

Exercise 3
Use data on houses sold in Stockton, California, to estimate a model for logarithm of price (*lnPRICE*). As regressors use total square feet of living area (*sqft*), number of full baths (*baths*), info on whether it is vacant Yes = 1 No = 0 (*vacant*), whether it has two floors (*stories2*), and *vac_sto2* that is interaction term between *vacant* and *stories2*. Let's reconsider the model from Exercise 2 (a). Verify assumptions of the linear regression model for the model $lnPRICE = \beta_0 + \beta_1 sqft + \beta_2 baths + \varepsilon$.

a) Verify the functional form of the model.
   i. Ramsey's RESET test

b) Check the assumption of homoscedasticity
   i. Breusch's and Pagan's test
   ii. White's test

c) Is the error term normally distributed?

[1] Hill, R.C., Griffiths, W.E., Lim, G.C., 2008, *Principles of Econometrics, Third Edition*, Wiley.

Exercise 4

The file *budgets.dta*[2] contains observations on 31901 households in Poland. An economist built an econometric model for alcohol expenditures (*w02*). As independent variables she used: household's income (*income*), total expenditures (*expend*), number of pearsons in the household (*nop*), alcohol price (*price*), and type of location (*location*) (codes: 1 for cities with at least 500 thousand of people, 2 for 200-500 thousand, 3 for 100-200 thousand, 4 for 20-100 thousand, 5 for less than 20 thousand, and 6 for countries).

a) Produce a similar table to the one given below
b) Test hypothesis of insignificance of location in modelling alcohol consumption
c) Verify the hypothesis $H_0: \ \beta_{location5} = -5$

	model1	model2	model3	model4	model5
income	0.007*	0.007*	0.007*	0.007*	0.007*
expend	0.008*	0.008*	0.008*	0.008*	0.008*
loc2	1.861	2.426	3.006*	2.479	
loc3	-1.486				
loc4	-1.776	-1.206			
loc5	-5.406*	-4.829*	-4.231*	-4.824*	
loc6	0.101	0.685	1.293		
nop	2.176*	2.162*	2.141*	2.240*	2.154*
price	163.687*	163.583*	163.399*	163.649*	162.973*
Constant	-166.554*	-167.019*	-167.417*	-167.247*	-166.806*
F	219.545	246.922	282.091	328.862	488.625
Prob > F	0.000	0.000	0.000	0.000	0.000
R-squared	.0584364	.0584199	.0583977	.0583551	.0578315
Adj R-squared	.0581702	.0581833	.0581907	.0581777	.0577131
N	31847	31847	31847	31847	31847

* p<0.05

Exercise 5

„[Let's] use a dataset (`fertil2.csv`) that contains data on 4361 women from a developing country. We want to model the number of children ever born (*ceb*) to each woman based on their *age*, their age at first birth (*agefbrth)*, and an indicator of whether they regularly use a method of contraception (*usemeth*)."[3]

a) estimate the model
b) verify the homoscedasticity assumption with Breusch's and Pagan's test
c) apply White's robust estimator (White's-Huber's)
d) apply White's robust estimator MacKinnon's and White's
e) assume that there are clusters in our data (*children* – number of children in the household); apply appropriate robust matrix
f) present results from (a), (c), (d), and (e) in one Quality Publication Table

[2]This file is of STATA format. Use `read.dta` or `read.dta13` functions from *foreign* or *foreign13* packages accordingly.

[3]C. F. Baum, *Introdution to Modern Econometrics Using Stata*, Stata Press, page 137