Utrecht University
Utrecht University School of Economics

Midterm exam Econometrics (wisb377)

Tuesday, 6 October 2020, 13:00-15:00 CET

For those who have permission from the Board of Examiners for an extension of time, they can hand the answer sheet in on 15:40 CET ultimately.

- Please do not post copies on the Internet.

Remarks:
- This entrance test consists of 9 sub-questions on 5 numbered pages (included front page).
- Write your name and registration number on each page of your exam.
- You should write down your answers in full, as if it were an on campus exam.
- Please make pictures of all pages of your answers (with a camera, copy the pictures in **ONE Word document**, and upload the document on Blackboard.
- You should upload the Word document **on 15:00 CET ultimately** (or mail it to W.H.J.Hassink@uu.nl).
- Please do not post copies of this exam on the Internet.

By taking the midterm exam of Econometrics of October 6th 2020, you automatically confirm that

"*By taking this exam I declare that I will formulate the answers myself, without the help of others, and without using unauthorized tools, and take the exam according to its instructions. Violation of these rules is regarded as fraud / plagiarism.*"

Questions

In the questions below – unless otherwise stated – we consider the linear regression equation

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + u_i \qquad i=1,\ldots,n$$

for which y is the dependent variable. $\boldsymbol{\beta}$ is a $(k+1)$ dimensional column vector. $\mathbf{x}$ is the $(k+1)$ dimensional column vector of explanatory variables, and u is an error term. Subscript i refers to the i-th individual. n is the number of observations of the data set.

a) We are interested in minimizing the loss function

$$L(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \boldsymbol{\beta}'\mathbf{x}_i)^2$$

Please carefully explain why the Hessian of this minimization procedure is a positive definite matrix, by making use of the definition of a positive matrix.

b) Compare the two loss functions

$$L_1(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \boldsymbol{\beta}'\mathbf{x}_i)^2 \ .$$

and

$$L_2(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left|y_i - \boldsymbol{\beta}'\mathbf{x}_i\right|$$

for which the loss is the absolute distance of y_i and $\boldsymbol{\beta}'\mathbf{x}_i$

Explain <u>why</u> and <u>how</u> the influence of outliers in the observations have a different effect on the vector of estimated parameters when minimizing both loss functions. Please motivate your answer.

c) Could you please calculate the minimand $\hat{\boldsymbol{\beta}}$ of the following optimization problem?

$$\min_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{C}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

For which $\mathbf{C}$ is a symmetric (k+1) x (k+1) matrix of constants. $\mathbf{X}$ is an n x $(k+1)$ matrix. Which necessary assumptions have you applied to derive $\hat{\boldsymbol{\beta}}$?

d) Three subquestions:
d1) Please calculate the covariance between the fitted value $\hat{y}$ and the residual $\hat{u}$ that were both obtained by an Ordinary Least Squares regression procedure.
d2) For this result, is it necessary to include an intercept in the regression equation?
d3) What does this result imply for the covariance between y and u ?

e) Please calculate $\mathrm{tr}\left(\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}-\boldsymbol{\beta}\right)'\left(\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}-\boldsymbol{\beta}\right)$, for which tr refers to the trace operator. Which necessary assumptions are required for your solution?

f) Three sub-questions.

f1) What is wrong with the notation of the following estimated regression equation, for which the parameters were estimated with Ordinary Least Squares with information of a sample of firms?

$$Log(Costs_i) = \hat{\beta}_0 + \hat{\beta}_1 \log(Firmsize_i) + \hat{\beta}_2 (Productivity_i) \qquad i=1,\ldots,n$$

for which the log is the natural logarithm, *Costs* is the costs of a firm in thousands of euros, *Firmsize* is the number of employees and *Productivity* is the value of the production per worker in thousands of Euros.

f2) For the following population regression equation, please give a precise interpretation of the regression parameters β_1 and β_2

$$Log(Costs_i) = \beta_0 + \beta_1 \log(Firmsize_i) + \beta_2 (Productivity_i) + u_i \qquad i=1,\ldots,n \qquad (1)$$

f3) Do the parameters β_1 and β_2 of equation (1) reflect a causal impact on the dependent variable? Please motivate your answer.

g) For the linear regression equation

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + u_i \qquad i=1,\ldots,n$$

the regression parameters are estimated by Ordinary Least Squares for four different datasets.
a) A random sample of 20 observations.
b) A sample of 20 observations.
c) A random sample of 1000 observations.
d) A sample of 1000 observations.

<u>Question:</u> Please formulate carefully the necessary assumption of exogeneity between the error term *u* and the right-hand side variables in $\mathbf{x}$ for each of the four data sets. Motivate your answer.

h) How would you reconsider your answers to question g) above, for the following specific linear regression equation: The dependent variable *y* is the transaction price of the house, and the two right-hand side explanatory variables in $\mathbf{x}$ are the number of rooms and the plot size of the house? Please motivate your answer.

i) The variance of the error term $Var(u) = \sigma_u^2$ of the linear regression equation can be estimated by $\hat{\sigma}_{u,n}^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2$, which depends on the sample size *n*.

Please apply the strong law of large numbers (and some other theorems) to demonstrate that the estimator $\hat{\sigma}_{u,n}^2$ is weakly consistent for σ_u^2

$$\text{plim}\ \hat{\sigma}_{u,n}^2 = \sigma_u^2$$

Carefully describe the full list of necessary assumptions that you have applied.

< End of the exam >