Utrecht University

Utrecht University School of Economics

Midterm exam Econometrics (WISB377)

Thursday, 13 October 2022, 11:00-13:00 CEST

For those who have permission from the Board of Examiners for an extension of time, they can hand in the answer sheet on 13:40 CEST ultimately.

Exam instructions

At the start of the exam

- Candidates who arrive 30 minutes after the time scheduled for the start of the examination will not be permitted entry to the examination room.

During the examination

- Nobody is allowed to leave the room within the first 30 minutes after the start of the exam.
- You are not allowed to go to the restroom unless you have permission of the Examiner or an invigilator.
- **MOBILE PHONES AND OTHER COMMUNICATION DEVICES ARE ONLY ALLOWED WHEN SWITCHED OFF AND STORED IN CLOSED BAGS**.
- *It is a closed book exam. It is* ***not*** *allowed to use any study aids such as books, readers, (preprogrammed) calculators*
- You may use a simple calculator and a dictionary (without any [handwritten] notes in it).
- The exam form is **NOT** allowed to be taken home by the candidate

Results/Post-examination regulations:

- The results of the examination will be announced on Blackboard within two weeks of the exam date. At the same time the time & place of the exam inspection will be announced.
- We do not discuss exam results over the phone or by email.
- After the announcement of the exam results in OSIRIS you have four weeks within which to lodge an appeal against your grade.
- Four weeks after the results of this exam are published, the original exam is available to you, when a declaration is signed, stating that no appeal has been made or will be made.

You can request a photocopy of your answers at the Student Desk up and until four weeks after publication of the results

Questions (sub-questions a – f: 1 point; g, h : 1.5 points). In total 9 points.

a) Using a random some of *n* observations, Ordinary Least Squares (OLS) is applied for the estimation of the vector of regression parameters $\boldsymbol{\beta}$ of the linear regression equation

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

Demonstrate that $\hat{\mathbf{y}}\,'\hat{\mathbf{u}} = 0$. In this derivation, you also need to mention how you apply the necessary assumptions to obtain this result.

Solution: See Chapter 1, example 13.

b) What does the outcome $\mathbf{X}\,'\hat{\mathbf{u}} = \mathbf{0}$ imply about the assumptions for the unbiased estimator

$$E\left(\hat{\boldsymbol{\beta}} \mid \mathbf{X}\right) = \boldsymbol{\beta} \text{ ?}$$

Please carefully motivate your answer by referring to the relevant assumptions.

Solution: The outcome $\mathbf{X}\,'\hat{\mathbf{u}} = \mathbf{0}$ is based on the assumption $\text{rank}(\mathbf{X}) = k+1$. For $E\left(\hat{\boldsymbol{\beta}} \mid \mathbf{X}\right) = \boldsymbol{\beta}$, it is assumed that $E\left(\mathbf{u} \mid \mathbf{X}\right) = \mathbf{0}$. Thus the outcome of the residual is uninformative about the assumption for the error term $E\left(\mathbf{u} \mid \mathbf{X}\right) = \mathbf{0}$.

c) For the specification of sub-question a, could you please derive whether the unconditional expected value of the OLS estimator $\hat{\boldsymbol{\beta}}$ is the vector of parameters $\boldsymbol{\beta}$. Thus $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$. What are the necessary assumptions for this proof?

Solution: $E\left(\hat{\boldsymbol{\beta}} \mid \mathbf{X}\right) = \boldsymbol{\beta}$ is based on the three assumptions a) linear model in $\boldsymbol{\beta}$, b) $\text{rank}(\mathbf{X}) = k+1$ and c) $E\left(\mathbf{u} \mid \mathbf{X}\right) = \mathbf{0}$. Next we can apply the law of iterated expectations $E\left(E\left(\hat{\boldsymbol{\beta}} \mid \mathbf{X}\right)\right) = E\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$

d) For the linear regression equation

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

the explanatory variables in the random vector $\mathbf{x}_i$ are strictly exogenous.

- Could you please give a careful definition of strict exogeneity (in formula)? Make sure you carefully define all of the variables used.
- What does strict exogeneity mean (in words)?
- Give two examples of economic models in which the assumption of strict exogeneity is violated.

Solution: Formula: $E(\mathbf{u} \mid \mathbf{X}) = \begin{pmatrix} E\left(u_1 \mid \mathbf{x}_1 \ldots, \mathbf{x}_n\right) \\ E\left(u_2 \mid \mathbf{x}_1 \ldots, \mathbf{x}_n\right) \\ \vdots \\ E\left(u_n \mid \mathbf{x}_1 \ldots, \mathbf{x}_n\right) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$. It says that all error terms are mean independent of all explanatory variables, for all observations. Strict exogeneity is violated in a linear regression equation with a lag of the dependent variable as well as a linear regression equation with a feedback mechanism.

e) For the linear regression equation

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

we consider the covariance matrix of the OLS estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, for which the matrix $\mathbf{X}$ is of full column rank. Under the additional assumptions of homoskedasticity and strict exogeneity, could you please derive

$$Var(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1}$$

Please carefully mention how you are using the assumptions in your derivation (and please make the assumptions explicit).

f) For the covariance matrix

$$V\hat{a}r(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = \hat{\sigma}_u^2 (\mathbf{X}'\mathbf{X})^{-1}$$

please compute the variance of the predicted value of the dependent variable: $Var(\mathbf{x}'\hat{\boldsymbol{\beta}} \mid \mathbf{x} = \mathbf{a})$ for which $\mathbf{a}$ is some vector of constants.

Solution: $V\hat{a}r(\mathbf{x}'\hat{\boldsymbol{\beta}} \mid \mathbf{x} = \mathbf{a}) = V\hat{a}r(\mathbf{a}'\hat{\boldsymbol{\beta}} \mid \mathbf{x} = \mathbf{a}) = \mathbf{a}'V\hat{a}r(\mathbf{a}'\hat{\boldsymbol{\beta}} \mid \mathbf{x} = \mathbf{a})\mathbf{a}$

g) In this question we consider two specifications. For a sample of *n* sales of houses, a researcher wants to use Ordinary Least Squares to estimate the regression parameters of the linear regression equation

(1) $\log(Price_i) = \beta_0 + \beta_1 Size_i + \beta_2 Nrooms_i + u_i \qquad i = 1, ..., n$

For which *Price* is the transaction price of the house, *Size* is the size in square meters and *NRooms* is the number of rooms of the house. In particular, we are interested in the effect of the number of rooms on log(*Price*).

For the attractiveness of the location of the house, it is argued that

$$E(Location_i \mid Size_i, NRooms_i) \neq 0$$

and that the location of the house has positive effect on the dependent variable $\log(Price_i)$

Thus,

$$E(v_i \mid Size_i, NRooms_i, Location_i) = 0$$

for which $u_i = v_i + Location_i$

In addition, we consider the auxiliary regression (w is an i.i.d. error term) that

$$Location_i = \gamma_0 + \gamma_1 Size_i + \gamma_2 Nrooms_i + w_i$$

$$E(w_i \mid Size_i, NRooms_i) = 0$$

For which the R-squared has a value of 0.995.

Question: to estimate the effect of the number of rooms on log(*Price*), what would be your favourite specification? Please motivate carefully your answer, in which you pay attention to both the location (bias) and the variation (efficiency) of the OLS estimator.

Solution: This is an example of the trade-off between bias and efficiency. Without the location of the house, the parameter estimates of equation (1) are biased. However, when we include the location of the house, the standard errors of the estimated regression parameters will be very large, because of the strong multicollinearity between *Size*, *Nrooms* and *Location*. There is no perfect multicollinearity, since the R-squared of the auxiliary regression is not equal to one. Overall, the preferred specification is

$$\log(Price_i) = \beta_0 + \beta_1 Size_i + \beta_2 Nrooms_i + \beta_3 Location_i + v_i$$

h) Please give the proof for the following theorem.

Let

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

for which $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{ik})'$.

Assumption 2. $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ be a sequence of identically and independently distributed (k+1) dimensional random vectors $E\mathbf{x}_i\mathbf{x}_i{}' = \mathbf{C}$ where $\mathbf{C}$ is a finite matrix, for which the inverse of $\mathbf{C}$ exists.

Assumption 3. The *n* random variables $x_{1j}u_1, x_{2j}u_2, ..., x_{nj}u_n$, $j = 1, ..., k$ are i.i.d., with $Ex_{ij}u_i = 0$ (so that it is finite). The same holds for the *n* random variables $u_1, u_2, ..., u_n$, which are i.i.d with $Eu_i = 0$.

Result: $\hat{\boldsymbol{\beta}}_n \xrightarrow{p} \boldsymbol{\beta}$ as $n \to \infty$

In your proof, please mention how all of these assumptions are used.

Solution: See Chapter 6, theorem 6.9

< end of the exam >