University of Duisburg-Essen Summer Semester 2019

Mercator School of Management

Prof. Dr. Marie Paul and Fabian Dehos

Exam in the Course:

Core Econometrics II for RGS Econ

August 2019

The exam lasts 90 minutes. 90 credits can be obtained.

You may use a non–programmable calculator and one sheet of paper with handwritten notes.

Please don't forget to write your name on all sheets you hand-in.

The exam consists of **4** pages including the cover page. Please check the completeness.

Good Luck!

1.) Suppose you want to estimate the effect (or rather descriptive relation) of x (primary school teachers' pay scheme) on y (student test scores) using a simple linear regression model. The data you use is a random sample of primary school students in Germany in 2018. The data includes indicators for students, schools, school-districts and states (Bundeslaender). The pay scheme is set by the state, thus the variable x varies between states but not within states.

   a) Shortly explain why you would cluster standard errors. [5 credits]

   b) At which level would you cluster and why? [5 credits]

   c) Do you expect the difference between the clustered standard errors and the normal standard errors to be large in this example or not? Give two reasons. [5 credits]

   —————————

   [15 credits]

2.) Suppose you consult a large firm offering translation services. You are asked to estimate the causal effect of telework (i.e. working from home instead of from the firm's offices) on translaters' performance. Translaters may choose whether they do telework or not. You measure employee's performance by number of translated lines per hour.

   a) First, you consider applying a multiple regression model to estimate the causal effect of telework on performance. From the process generated data of the firm you draw a random sample of translaters for 2017. For each employee in the sample you have information on the average number of translated lines per hour of work, whether he or she does telework or not, as well as gender and experience. Write down the model you would estimate. Give an example why the estimated coefficient on the dummy *telework* might be downward biased due to omitted variable bias. Give an example why this estimate might alternatively be upward biased due to omitted variable bias. [8 credits]

   b) Second, you consider applying a fixed-effects estimator to obtain the causal effect of telework on performance. The option of telework became first available in January 2016 and employees need to decide for a whole year whether they do telework or not. From the process generated data of the firm you draw a random sample of translaters for 2015 and add for each person in the sample the information for 2016 and 2017. For each person in the sample and each year you have the information listed in part (a). Discuss whether in this application the strict exogeneity assumption holds:

$$E[\varepsilon_{it}|\alpha_i, \mathbf{x}_{i1}, ..., \mathbf{x}_{iT}] = 0, \quad t = 1, ..., T.$$

[8 credits]

c) Third, you consider applying an instrumental variables estimator to esti-
mate the causal effect of telework on performance. You use the data from
part a). You add to the data a dummy variable whether a worker has
participated in a particular seminar on work-life balance. Your idea is
to use this variable as an instrument for telework, because (1) partici-
pants of this seminar had been randomly chosen, and (2) in the seminar
participants learned among other things about advantages of telework.
Consider the framework for *heterogenous* treatment effects. State the
following assumptions (either in formulas or in words) and discuss their
validity for the application: 1) Independence, 2) Exclusion, 3) Monoto-
nicity. [14 credits]

_____

[30 credits]

3.) Consider a dynamic panel data model with a lagged dependent variable,
and one additional regressor w_{it}, namely

$$\begin{aligned} y_{it} &= y_{i,t-1}\beta_1 + w_{it}\beta_2 + \alpha_i + u_{it} \\ &= x_{it}\beta + \alpha_i + u_{it}, \end{aligned}$$

where $x_{it} = (y_{i,t-1}, w_{it})$ and $\beta = (\beta_1, \beta_2)'$. We observe y_{it} and x_{it} for $i = 1, ..., n$ and $t = 1, 2$ as well as y_{i0}. We assume that $y_{i0} = \alpha_i + u_{i0}$ and that $u_{i0}, u_{i1}, u_{i2}, w_{i1}, w_{i2}$ and α_i are mutually independent, *iid* across i, and are all
normally distributed with mean zero and variance one. First differences are
defined by $\Delta y_{it} = y_{it} - y_{i,t-1}$, $\Delta w_{it} = w_{it} - w_{i,t-1}$, $\Delta x_{it} = x_{it} - x_{i,t-1}$ and
$\Delta u_{it} = u_{it} - u_{i,t-1}$. Consider the following estimator (the OLS estimator after
first differencing) for β:

$$\hat{\beta}_{FD-OLS} = \left(\sum_{i=1}^{n} \Delta x_{i2}' \Delta x_{i2}\right)^{-1} \sum_{i=1}^{n} \Delta x_{i2}' \Delta y_{i2},$$

a) Calculate $\mathbb{E}[\Delta w_{i2}\Delta u_{i2}]$ and $\mathbb{E}[\Delta y_{i1}\Delta u_{i2}]$. [5 credits]

b) Is $\hat{\beta}_{FD-OLS}$ a consistent estimator for β? Show this using your results
from (a). [10 credits]

_____

[15 credits]

4.) The Department of Education of Watopia wants to evaluate the impact of a scholarship program for bachelor students and asks you for advice. More specifically, the aim of the analysis is to determine the impact of a scholarship (treatment) on the starting salary (outcome).

a) In a first step, the Department of Education wants you to compare the raw means starting salary between scholarship recipients and non-recipients. Why do you refrain from such a comparison? Do you over or underestimate the impact of a scholarship if you follow the Department's recommendation? Explain. [5 credits]

b) In Watopia the probability of being offered a scholarship jumps at a certain SAT-threshold (Scholastic Assessment Test). This threshold is a priori unknown and changes every year. But if you score above the threshold the probability of receiving a scholarship increases significantly since these applications are given special consideration. However, it does not guarantee that you are offered a scholarship.
Draw a simplified graph with SAT-score (ranging from 0 to 100 points) on the x-axis and the probability of receiving a scholarship on the y-axis. The SAT-threshold for „special consideration" can be assumed to be 75. Is scholarship assignment a deterministic function of the SAT-score? [5 credits]

c) Using the toolkit of an applied micro-econometrician you want to identify the causal impact of scholarship receipt on the starting salary. Given the previous setting of scholarship provision in Watopia, discuss in detail one empirical approach (bullet points are sufficient). Name the approach of your choice and discuss your chosen methodology critically:

- What is the identifying assumption? Can you test it? How would you convince a potential referee that is holds?
- Write down the underlying regressions that you would want to estimate. How would you want to estimate them: parametrically or non-parametrically? What additional decisions do you need to take?

[12 credits]

d) Which effect do you identify with your choosen methodology? Discuss the external validity of your potential results for all scholarship recipients within Watopia and the effect of scholarships in other countries. [8 credits]

_____

[30 credits]