Written Exam for the B.Sc. or M.Sc. in Economics Summer 2020

Applied Econometric Policy Evaluation

Take-home exam

June 18, 2020, from 10:00am to 10:00pm

This exam consists of 8 pages in total.

Answers only in English.

Be careful not to cheat at exams!
Exam cheating is for example if you:

- Copy other people's texts without making use of quotation marks and source referencing, so that it may appear to be your own text
- Use the ideas or thoughts of others without making use of source referencing, so it may appear to be your own idea or your thoughts
- Reuse parts of a written paper that you have previously submitted and for which you have received a pass grade without making use of quotation marks or source references (self-plagiarism)
- Receive help from others in contrary to the rules laid down in part 4.12 of the Faculty of Social Science's common part of the curriculum on cooperation/sparring

You can read more about the rules on exam cheating on your Study Site and in part 4.12 of the Faculty of Social Science's common part of the curriculum.

Exam cheating is always sanctioned by a written warning and expulsion from the exam in question. In most cases, the student will also be expelled from the University for one semester.

Practical instructions for the take-home exam

Read the entire exam before you respond. Answer every question in each problem. The exam consists of five problems in total.

The exam can be answered in groups of a **maximum of 2 students**. Hand-in a single report for the entire group **and specify each group member's contribution to the report**.

You must submit a comprehensive report with relevant tables and figures. The front page of the report must use the template available at `https://eksamen.ku.dk/`. Fill in the exam numbers of all group members on the front page. The second page of the template must specify which paragraphs and/or sections of the report is answered by which group member. This page may not contain other information.

Prepare one STATA do-file generating all tables and figures that appear in your report. The program must produce tables and figures in the same order as they appear in the report. Comments should clearly indicate which table or figure appearing in the report is being produced. Make sure that the do-file can be executed without any errors. The do-file must include the exam numbers of all group members.

The report must not exceed 12 (normal) pages. This includes the main text, tables and figures in the report, but not the front page and the list summarizing each group member's contribution to the report.

For the exam in Applied Econometric Policy Evaluation, a normal page is defined as a text document with the following attributes[1]:

- A4 format
- Font size set to 12
- Line spacing set to 1.5
- Margins (left/right/top/bottom) of at least 2.5 cm

The exam ends **June 18 at 22.00 (10:00pm)**. The report and the STATA do-file must be uploaded electronically no later than 22.00 (10:00pm).

Uploading your report

Each group must hand-in only one report in total. One student hands in the report by uploading it to University of Copenhagen's Digital Exam system and then adding the rest of the group members to the hand-in. Go to the website `https://eksamen.ku.dk/` and click on 'Log in as student'. Use your regular KU login and password to enter Digital Exam. Click on 'Applied Econometric Policy Evaluation' in your assignments. On the page 'Information about the hand-in', you must add the other group member to the handed-in answer (if you are in a group). Click on 'Add member' and follow the instructions on Digital Exam to invite your fellow group members. Group members will be added to the handed-in answer as soon as they **accept** your invitation.

[1]The Study Handbook for the Economics program defines a normal page as 2,400 characters, but for this exam, a normal page is instead defined in terms of format, font size, line spacing and margins.

Next, go to 'Upload hand-in' to upload your files. Each group must upload two files:

1. The report itself must be uploaded as a PDF file. The filename must start with the letter R followed by the exam numbers of all members of the group separated by _ ("underscore").
2. The STATA do-file must be uploaded as a file in plain text format (.txt). The filename must start with the letter P followed by the exam numbers of all members of the group separated by _ ("underscore").

Use the same combination of exam numbers for both files.

> Example: A group of two members with exam numbers 72 and 174 will submit the following files:
>
> 1. R_72_174.pdf
> 2. P_72_174.txt

If needed, a free PDF converter is available at `www.pdf995.com`.

If you have problems accessing the Digital Exam system at the deadline of the take-home exam or if you have difficulties with the upload function you must e-mail your answer to `samf-fak@samf.ku.dk` within 22:30 (10:30pm). Handing in your exam answer by e-mail requires that you describe the problems and provide screen dumps that document this.

Access to data

For the take-home exam, there are several data sets available on the Digital Exam website (`https://eksamen.ku.dk/`). Follow the instructions below to pick the correct data set for your group:

1. Determine the **lowest** number among the exam numbers of the group members. Use the **last** digit of the **lowest** exam number as your "group number".

   > Example: A group of two members with exam numbers 72 and 174 will have "2" as the last digit of the lowest exam number.

2. Download the STATA file groupdataX.dta from the Digital Exam website, where X is equal to the group number.

   > Example: The group from before downloads groupdata2.dta from the Digital Exam website.

3. Download the data to your computer.
4. Open the data in STATA and execute the **describe** command to ensure the data appears operative.

If you have trouble selecting or opening the data, you can contact Søren Leth-Petersen on telephone 3532 3084 or Daniel le Maire on telephone 3532 3063 during the period 10.00am to 12.00pm (noon) on June 18.

After this, no additional help will be provided for the exam.

Introduction to the assignment:

"Children and Women's Labor Supply"

The objective is to quantify the the impact of having children on labor supply. You have access to a randomly chosen subset of women who are all born in 1970. The data set includes information about the women's labor supply and children during the period 1990-2015, i.e. when the women were aged 20-45. The data set includes only observations for women who have had 1-3 children during this period. The variables available are summarized in Table 1 below.[2]

Table 2: List of variables

Variable name	*Description*
id	Social security number (anonymized)
age	Age of mother
numchild	The total number of children
child1_by	Birth year child 1
child2_by	Birth year child 2
child3_by	Birth year child 3
Boy1	Dummy variable=1 if child 1 is a boy
Boy2	Dummy variable=1 if child 2 is a boy
Boy3	Dummy variable=1 if child 3 is a boy
working	Dummy variable=1 if working

[2]The data used for this exam are simulated.

Problem 1 (10%): Summary and descriptive statistics

1. Provide a descriptive analysis of the variables in your data set using relevant summary statistics. Examples of relevant aspects to include are number of women in the data set, total number of observations, work status, the number of children. Note that these may not be the only aspects of the data that are relevant to describe. The descriptive analysis may include both table(s) and relevant graphical illustration(s) of the data.

2. Assume that fertility is completed by age 45. What is the distribution of number of children in the sample upon completed fertility?

3. At what age do women in the sample typically have their first child and what is the spacing between children for women who have more than one child.

Problem 2 (25%): Event study I

In this problem you are asked to estimate the causal effect of having the first child on labor supply using an event-study design

1. Illustrate how the event study design works by plotting three curves with the average propensity to work for women who have their first child in 1996, 2000, and 2004 respectively, and explain what comparisons the event study design makes in the data.

2. Consider the following regression:

$$working_{it} = \alpha_i + \gamma D_{it}^{Age} + \sum_{k=-B_1}^{A_1} D1_{it}^{k}\delta_k + u_{it} \tag{1}$$

where $working_{it}$ is the outcome variable indicating whether individual i works at time t, α_i is a fixed effect, D^{Age} is a set of age specific dummy variables and γ the associated parameters, $D1_{it}^{k} = 1$ if the first child was born k periods ago, δ_k measures the effect k periods after birth of the first child took place. $-B_1$ is the earliest period on the event axis for the first child, and A_1 is the latest period on the event axis for the first child.
Below you are asked to estimate different versions of equation (1). In all cases, report the estimates of $[\delta_{-10}; \delta_{10}]$ in a graph with the event time on the horizontal axis and the outcome on the vertical axis. In all cases you should give an account for the conditions under which the causal effect is identified, omit the appropriate number of pre-event dummies, and comment on your findings and assess whether the model well specified.

   (a) Estimate equation (1) including only event time dummies, i.e. leave out individual fixed effects and age dummies.

   (b) Estimate equation (1) including event time dummies and age dummies, i.e. leave out individual fixed.

(c) Estimate equation (1) including event time dummies, age dummies and individual fixed effects

(d) What is the preferred model?

Problem 3 (25%): Event study II

1. In this problem we will focus on the version of equation (1) without age dummies and without individual fixed effects and then augment this equation with a set of additional explanatory variables controlling for the timing of a possible second child.

$$working_{it} = \beta_0 + \sum_{k=-B_1}^{A_1} D1_{it}^{k}\delta_k + \sum_{j=-B_2}^{A_2} D2_{it}^{j}\theta_j + u_{it} \tag{2}$$

where β_0 is a constant term, $D2_{it}^{j} = 1$ if the second child was born j periods ago, and θ_j is the effect of the second child j periods after the birth took place. $-B_2$ is the earliest period on the event axis for the second child, and A_2 is the latest period on the event axis for the second child. For example, $D2^{-3} = 1$ when the second child is to be born three periods later. Given that you normalize the same number of pre-first-birth δ coefficients to zero for identification as in problem 2, question 2 (a), how many of the pre-second-birth θ parameters do you need to normalize for identification? What is the intuition?

2. Estimate equation (2), comment on the results and compare with your findings from estimating equation (1) without age dummies and without individual fixed effects. How does the labor supply profile change and how does the inclusion of the timing of the second child change the interpretation of the estimates of the effect of the first child on female labor supply?

3. Report the estimates of $[\delta_{-10}; \delta_{10}]$ and $[\theta_{-10}; \theta_{10}]$ in two separate graphs and compare the effect on the employment rate of having the first and second child? How should we interpret the differences?

Problem 4 (20%): Instrumental variable estimation

The effect of the third birth could also be estimated by extending the set-up in equation (2), but instead we will follow Angrist and Evans (1996) and estimate the effect of having an extra child on female labor supply by exploiting that there tend to be an increased probability of having an extra child, among parents who already have two children, if the first two children have the same sex. In the subsequent questions, you should only use observations (i) for women with at least two children, and (ii) for years after a woman gives birth to her second child.
Following Angrist and Evans (1996), we will consider the following equation

$$working_{it} = \alpha_0 + \alpha_1 ThirdChild_{it} + u_{it} \tag{3}$$

where $ThirdChild_{it}$ is a dummy for having three children at time t.

1. For women with at least two children construct a dummy variable, $Samesex_i$, indicating whether the two first children have the same sex. Estimate the first stage equation. Under what conditions does the 'same-sex' dummy variable work as a valid instrumental variable for having three children?

2. Explain why one cannot necessarily interpret the estimate of α_1 as the causal effect of having an additional child when estimating equation (3) by OLS.

3. Estimate equation (3) by OLS and by 2SLS where $Samesex_i$ is used as an instrument for $Thirdchild_{it}$. Enter the results into a table and describe your findings.

4. The event study design in Problem 2 and 3 and the 2SLS estimates based on using the same sex of the first two children as an instrument for the effect of the third child on labor supply represent different approaches to estimating the effect of an additional child on female labor supply. How should we interpret the estimate of α_1 obtained by 2SLS using the same sex instrument, and how does the interpretation differ from the interpretation of the event study estimates in Problem 2 and 3?

Problem 5 (20%): COVID-19 compensation of firms

With the spread of COVID-19 and the partial lock-down of the society in March 2019, the Danish government and all the parties of in the Danish Parliament signed an agreement on a comprehensive aid package for Danish workers and companies. A number of temporary measures were adopted to help Danish workers, Danish workplaces and Danish companies through the COVID-19 crisis.

Since many firms' sales were expected to be reduced substantially during the partial lock-down of society, but firms still needed to pay their fixed costs, a compensation scheme was introduced to compensate the firms for their fixed costs.[3]

The proportion of fixed costs that will be compensated depends on the percentage decline in sales. This decline is computed by comparing the sales in the *compensation period* to the sales in the *reference period*. The firms can apply for compensation for the period from March 9, 2020 to July 8, 2020. Subject to some exceptions, the *reference period* is from April 1 to June 30, 2019.[4]

The proportion of the firm's fixed costs that will be compensated is:

- 80 percent if the decline in sales is 80-100 percent.
- 50 percent if the decline in sales is 60-80 percent.
- 25 percent if the decline in sales is 35-60 percent.
- 0 percent if the decline in sales is less than 35 percent.

Suppose that you after the COVID-19 crisis are asked to study firm survival and the impact of this compensation scheme.

[3]We consider the updated rules from April 18, 2020. Furthermore, we only consider part of the compensation scheme for fixed costs. You are asked to base your answer solely on the description given in this question.

[4]Sales of the *reference period* is adjusted such that it can be compared to the *compensation period*, when the durations of the *reference period* and *compensation period* are different.

You will get access to the following variables for firms existing by March 1, 2020, but only if they were awarded some compensation for fixed costs.

- *id*: Firm id
- *estyear*: The year the firm was founded.
- *industry*: Primary industry by February 1, 2020
- *empl*: Number of employees by February 1, 2020
- $sales\_ref$: Adjusted sales from the *reference period*.
- *sales_comp*: Sales from the *compensation period*.
- $fixedcosts$: Fixed costs for the *compensation period*.
- *compensation*: The amount of fixed costs being compensated.
- *survival*: A dummy for firm survival by June 1, 2021.

You are asked to consider the following questions:

1. Which research design is appropriate for evaluating the impact of the compensation scheme for fixed costs on firm survival?
2. Discuss the practical implementation of your chosen research design.
3. Are there potential threats to your proposed research design? How can you examine whether these potential threats to your research design actually are a problem?