



Stockholm  
University

Department of Economics

Course name: Empirical Methods in Economics 2  
Course code: EC2404  
Type of exam: Main  
Examiner: Ferenc Szucs  
Number of credits: 7.5  
Date of exam: Tuesday March 16th, 2020  
Examination time: 3 hours (9:00-12:00)  
Aids: Books and notes are allowed.

---

Write your identification number on each answer sheet (the number stated in the upper right hand corner on your exam cover). If you do not have an exam cover, please write your personal identity number.

Start each new question on a new answer sheet.

---

Explain notions/concepts and symbols. If you think that a question is vaguely formulated, specify the conditions used for solving it. Only legible exams will be marked.

---

The exam consists of 5 questions. If you submitted 4 accepted home assignments you do not need to solve question 5 (you will get full credit anyway).

The maximum total point is 100. For the grade E 45 points are required, for D 50 points, C 60 points, B 75 points and A 90 points.

---

Your results will be made available on your Ladok account ([www.student.ladok.se](http://www.student.ladok.se)) within 15 working days from the date of the examination.

---

**Good luck!**

Question 1 – Potential outcomes framework (20 points)

A local government that wants to tackle unemployment in its village has started a program by which it guarantees 100 days per year of employment to those that enroll. With this program, the local government seeks to increase the employability of participants, and thus its objective is to increase the total number of employment days per year (both within and outside the program) for those that participate. Enrollment to the program is open to the whole village, and 50% of the village decides to join.

1. Explain what the outcome variable ( $Y_i$ ) and the treatment variable ( $D_i$ ) are in this example. (3 points)
2. What would the counterfactual be in this example? Explain in words and using the notation of the potential outcomes framework. (6 points)
3. Explain why selection bias could be a problem in this example. Explain in words and mathematically (by linking selection bias with the observed differences and the average treatment effect on the treated. (7 points)
4. What is your prediction about the sign of the selection bias in this case? Would a naive assessment of the observed differences under- or overestimate the causal effect of treatment in this example? (4 points)

Question 2 – Control variables (20 points)

You are asked to evaluate a scholarship program. The program provides financial support to pay university tuition. Eligibility is determined based on both merit and need, so students with a high enough SAT scores and a low enough socioeconomic backgrounds can get the scholarship. We are interested in the effect of the program on the subjects' income after college. You have access to a survey dataset taken in 2010 of 10,000 people graduated from high school at 2000. The dataset contains their income, whether they were awarded by the scholarship, variables on parental background, their SAT score if they applied to a university, and the census block they live in 2010. This latter information enables you to construct a new variable on the average income of the neighborhood they live.

1. Consider the following regression equation,

$$y_i = \beta_0 + \beta_1 T_i + u_i,$$

where  $y_i$  is wage in 2010 and  $T_i$  is participation in the scholarship program. Does  $\beta_1$  give you the average treatment effect of the scholarship program on latter income? Explain why or why not. (6 points)

2. Now, let's have a look at the following equation,

$$y_i = \alpha_0 + \alpha_1 T_i + \alpha_2 SES_i + \alpha_3 SAT_i + v_i,$$

where  $SES_i$  is the socioeconomic status of the parents in 2000 and  $SAT_i$  is the entrance exam score. Do you think the inclusion of these controls have changed the coefficient of the treatment? Do you think we need to include these controls? Explain why or why not. (7 points)

3. Finally, consider the equation,

$$y_i = \gamma_0 + \gamma_1 T_i + \gamma_2 SES_i + \gamma_3 SAT_i + \gamma_4 N_i + w_i$$

where  $N_i$  is the average income of  $i$ 's neighborhood in 2010. Do you think we need to include this control? Explain yourself. (7 points)

### Question 3 – Randomized Control Trial (20 points)

An NGO wants to evaluate the impact of free bednets on malaria in a village in Tanzania. The researchers at the NGO use an RCT to evaluate the impact. Before the experiment they run a baseline survey collecting basic demographic information about the households. 20% of households in the village already own bednets from last year. Among the remaining 80% of households, they give bednets to a random half of them (i.e. to 40% of the village), and the remaining 40% of the village does not get a bednet. Among the households that received bednets: some households use the bed nets as intended, some households use their bednets to catch fish in the river, and other households sell their bednets to households in the village that did not receive bednets. In addition, among households that already owned bednets from last year, some of them decide to throw away their bednets.

1. First, explain why the non-experimental correlation of the use of mosquito nets and death rate from malaria is misleading. (3 points)
2. What regression would you run to estimate the treatment effects of this intervention? Explain all terms, including how each variable is defined and the sample used in the analysis. (6 points)
3. How would you check whether the randomization was executed properly? (4 points)
4. Who are the compliers, always takers, and never takers in this analysis? (3 points)
5. Suppose you want to add a control for whether the individual lives close to a river. Will the inclusion of this control affect the expected estimate of your treatment effect in the above regression? Based on this, is it a good idea to include this as a control? (4 points)



#### Question 4 – Fixed-effects model (20 points)

Many economists and policymakers believe that job displacement has a permanent effect on workers' income. Assume you have access to the registry data of all employees in Sweden for years 2000-2015. You want to estimate the effect of losing a job on household income by comparing workers of firms having mass layoffs to workers of firms not experiencing such employment shocks (hint.: use a fixed-effects specification).

1. What specification would you use? Please spell out the exact equation you are about to estimate and answer the following questions: (12 points)
  - a. What is the unit of observation?
  - b. What are the outcome and the explanatory variable of interest?
  - c. How do you control for unobserved individual characteristics of workers?
  - d. How do you control for the aggregate income shocks (such as the effect of a recession)?
2. What is the main identifying assumption behind this specification? (4 points)
3. How could you check if this assumption holds? (4 points)



Question 5 – Neal and Johnson (1996) paper (20 points)

In class we covered the paper titled “The Role of Premarket Factors in Black-White Wage Differences”. Please answer the following questions about the paper:

1. What is the main question of the paper? (2 points)
2. Why doesn't the simple correlation between race and wage capture the causal effect of labor market discrimination? (3 points)
3. How does the paper mitigate the problem you highlighted above? (4 points)
4. Why do they prefer AFQT score to more traditional measures of productivity such as schooling, work experience or occupation? (4 points)
5. Does AFQT score measure the true effect of labor market productivity on wages? (3 points)
6. Why is it important that AFQT is a racially unbiased measure of skills and abilities? (4 points)