

***UNIVERSITY OF OSLO***  
***DEPARTMENT OF ECONOMICS***

Exam: **ECON4137 – Applied Micro Econometrics**

Date of exam: Wednesday, June 7, 2017    **Grades are given:** June 20, 2017

Time for exam: 2.30 p.m. – 5.30 p.m.

The problem set covers 4 pages (incl. cover sheet)

Resources allowed:

- Open-book exam, where all written and printed resources, as well as calculator are allowed

The grades given: A-F, with A as the best and E as the weakest passing grade. F is fail.

# Exam ECON4137 – Spring 2017

---

IMPORTANT: Always motivate your answers. Show knowledge and understanding of the concepts taught in the course. Your answers should be as short as possible and as long as necessary. Every subquestion receives equal points unless indicated otherwise.

---

1. (25 points) Discuss whether each of the following statements is correct or not.
  - (a) A misspecified OLS regression is uninterpretable.
  - (b) Heteroskedasticity implies that error terms are correlated.
  - (c) A regression with a very low R-squared cannot be trusted.
  - (d) Failing to reject a null hypothesis is not the same as accepting it.
  - (e) An explicit research design (RD, DID, IV, CIA) puts constraints on the econometric analysis, and that is a good thing.
  
2. (25 points) A researcher wants to analyze output using a production function with labor and capital as input factors. Cross-section data are available for  $n = 520$  firms on output ( $Y$ ), labor ( $L$ ) and capital ( $K$ ). The researcher decides to estimate the following regression model for output:

$$\log(Y_i) = \beta_0 + \beta_1 \log(L_i) + \beta_2 \log(K_i) + u_i, \quad i = 1, \dots, n. \quad (1)$$

Assume that error terms are homoskedastic. The Ordinary Least Squares (OLS) estimation results are:

$$\widehat{\log(Y_i)} = \underset{(0.13)}{2.57} + \underset{(0.04)}{0.91} \log(L_i) + \underset{(0.03)}{0.21} \log(K_i),$$

where the numbers between parentheses are estimated standard errors,  $SSR = 314$  and  $R^2 = 0.80$ .

- (a) Give an interpretation, in words, of the coefficient  $\beta_1$  in model (1).
- (b) Compute a test statistic for the joint null hypothesis  $H_0 : \beta_1 = 0, \beta_2 = 0$  and give its distribution.
- (c) You want to test for constant returns to scale,  $H_0 : \beta_1 + \beta_2 = 1$ , in equation (1). What regression would you estimate (give the specification) that allows you to test this hypothesis using a t-test on a single estimated coefficient?
- (d) (10 points) The researcher suspects that management quality also matters for output and is correlated with the inputs. Although information on management quality is not available, the researcher observes output and inputs over two periods. Describe in detail and under what assumptions you would still be able to obtain consistent estimates of  $\beta_1$  and  $\beta_2$ . In your answer list any model changes, data transformations or additional assumptions.

3. (30 points) In the Netherlands medical school admission was based on a lottery. The Stata output below shows summary statistics on the outcomes of the lottery based admission for a single cohort. The variable -adm- equals 1 if the applicant was admitted and is 0 otherwise. The variable -med- equals 1 if the admitted applicant completed medical school and is 0 otherwise. The variable -lnw- is log earnings 15 years after admission. You want to estimate the causal effect of completing medical school on earnings 15 years after admission.

```
. sum adm med lnw
```

Variable	Obs	Mean	Std. Dev.	Min	Max
adm	733	.7162347	.4511322	0	1
med	733	.7967258	.4027096	0	1
lnw	733	3.124762	.5523705	-2.747271	6.491241

```
. g v1 = lnw
. replace v1 = 0 if med==0
. g v2 = lnw
. replace v2 = 0 if med==1

. regress med adm , robust noheader
```

	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
adm	.4746978	.0363067	13.07	0.000	.40342 .5459756
_cons	.4567308	.0345859	13.21	0.000	.3888312 .5246303

```
. regress v1 adm , robust noheader
```

	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
adm	1.540733	.1164305	13.23	0.000	1.312155 1.769311
_cons	1.408397	.1090269	12.92	0.000	1.194353 1.62244

```
. regress v2 adm , robust noheader
```

	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
adm	-1.457028	.1143711	-12.74	0.000	-1.681563 -1.232493
_cons	1.656413	.1090769	15.19	0.000	1.442272 1.870554

- (a) Someone suggests to regress -lnw- on -med- using OLS. Does this give a consistent estimate of your effect of interest? Does this depend on the compliance rate (the rate at which admitted applicants complete medical school)?

An alternative approach is IV.

- (b) What would be the LATE interpretation of the IV estimate of the return to completing medical school in the current application?
- (c) Discuss instrument i) exogeneity ii) excludability (can you provide a possible violation?), iii) monotonicity, and iv) relevance.
- (d) What are the fractions of compliers, always takers and never takers?
- (e) What are the average potential outcomes for the compliers, and what is the IV estimate of the return to medical school?

- (f) Explain (and calculate if possible) what average potential outcomes are identified for the never-takers and always-takers.
4. (20 points) From May to September 1980 some 125,000 Cuban immigrants arrived in Miami on a flotilla of privately chartered boats. This increased the Miami labor force by 7%, and the percentage increase in labor supply to less-skilled occupations and industries was even greater because most of the immigrants were relatively unskilled.

Card (1990) studies the effects of the Boatlift on the Miami labor market, focusing on wages and unemployment rates of less-skilled workers.

Card assembled labor force survey data for whites, blacks, and Hispanics in Miami, as well as for four other cities: Atlanta, Los Angeles, Houston, and Tampa-St. Petersburg. These four cities had relatively large populations of blacks and Hispanics and exhibited a pattern of economic growth similar to that in Miami over the late 1970s and early 1980s, and economic conditions were very similar in Miami and the average of the four comparison cities between 1976 and 1984.

The following table (Table 4 from Card, 1990) shows unemployment rates of individuals age 16-61 in Miami and Four Comparison Cities for 1979 (before the boatlift), 1980 (the year of the boatlift), and 1980 (after the boatlift):

<i>Group</i>	<i>1979</i>	<i>1980</i>	<i>1981</i>
<i>Miami:</i>			
Whites	5.1 (1.1)	2.5 (0.8)	3.9 (0.9)
Blacks	8.3 (1.7)	5.6 (1.3)	9.6 (1.8)
Cubans	5.3 (1.2)	7.2 (1.3)	10.1 (1.5)
Hispanics	6.5 (2.3)	7.7 (2.2)	11.8 (3.0)
<i>Comparison Cities:</i>			
Whites	4.4 (0.3)	4.4 (0.3)	4.3 (0.3)
Blacks	10.3 (0.8)	12.6 (0.9)	12.6 (0.9)
Hispanics	6.3 (0.6)	8.7 (0.6)	8.3 (0.6)

- (a) Calculate the before-after, and difference-in-differences estimates of the boatlift on the unemployment rates of blacks.
- (b) Outline how you would estimate the difference-in-differences estimate in (a) using OLS.
- (c) Discuss the counterfactual assumptions you need to make in (b) in order to consistently estimate the effect of the boatlift, and outline at least 2 potential threats to the design.
- (d) You worry about city level random shocks to potential outcomes. Discuss how would you calculate your standard errors in (b) using the Moulton factor.