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.
   BRIEF ANSWER: False, it is the best linear approximation of the population CEF

   (b) Heteroskedasticity implies that error terms are correlated.
   BRIEF ANSWER: False, the definition of heteroskedasticity is $E[u_i^2|x_i] = \sigma_i^2$ and not $E[u_i u_j|x_j] \neq 0$.

   (c) A regression with a very low R-squared cannot be trusted.
   BRIEF ANSWER: False. a regression with a low r2 will do a poo/noisy job in within sample prediction, but we can still have precisely estimated effects.

   (d) Failing to reject a null hypothesis is not the same as accepting it.
   BRIEF ANSWER: True, there is still a positive probability that the null hypothesis is false (Type I Error).

   (e) An explicit research design (RD, DID, IV, CIA) puts constraints on the econometric analysis, and that is a good thing.
   BRIEF ANSWER: true. it limits specification searches and resulting bias, and it reduces the scope for misspecification

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:

$$\log(Y_i) = \beta_0 + \beta_1 \log(L_i) + \beta_2 \log(K_i) + u_i, \qquad i = 1, ..., n. \tag{1}$$

   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 β_1 in model (1).
   BRIEF ANSWER: a 1% increase in L (K) increases Y with 0.91 (0.21) %.

   (b) Compute a test statistic for the joint null hypothesis $H_0 : \beta_1 = 0, \beta_2 = 0$ and give its distribution.
   BRIEF ANSWER: We use the F test, recognizing that the R-squared in the restricted model equals 0: F = ((0.8 - 0) / 2) / ((1 - 0.8) / (520 - 1 - 2) = 1034. This will follow and F(2,517) distribution and we reject if F>c, where c is the critical value (at the 5% percent level f.e.).

(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?

BRIEF ANSWER: We substitute $\theta = \beta_1 + \beta_2 - 1$ in the model. Rewriting gives

$$\log(Y_i/L_i) = \beta_0 + \theta \log(L_i) + \beta_2 \log(K_i/L_i)$$

and we can now use the t-statistic of the estimated coefficient of $\log(L_i)$.

(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 β_1 and β_2. In your answer list any model changes, data transformations or additional assumptions.

BRIEF ANSWER: Fixed effects estimation. We need to add management quality η_i to our model and add time subscripts:

$$\log(Y_{it}) = \beta_0 + \beta_1 \log(L_{it}) + \beta_2 \log(K_{it}) + \eta_i + u_{it}$$

and we can estimate β_1 and β_2 by

i. first-differencing (or the within transformation) which removes η_i,

ii. as long as regressors are time-varying.

iii. and when $E[\Delta u_{it} \Delta \log(L_{it})] = 0$ and $E[\Delta u_{it} \Delta \log(K_{it})] = 0$. so no feedback.

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
------------------------------------------------------------------------------
             |               Robust
         med |      Coef.   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
------------------------------------------------------------------------------
             |               Robust
          v1 |      Coef.   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
------------------------------------------------------------------------------
             |               Robust
          v2 |      Coef.   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)?

BRIEF ANSWER: in principle we need to worry about omitted variable bias. with perfect compliance med=lottery and lottery is random so OLS is fine.

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?

BRIEF ANSWER: the return to medical school for people who complete medical school when admitted, but not otherwise

(c) Discuss instrument i) exogeneity ii) excludability (can you provide a possible violation?), iii) monotonicity, and iv) relevance.

BRIEF ANSWER: exog: true, it's a lottery afterall. excludability: less clear; disappointment can affect counterfactual outcomes. monotonicity: a defier is someone who doesnt attend medical school when admitted, but does when not admitted. this is not possible. relevance: true, cf regression output.

(d) What are the fractions of compliers, always takers and never takers?

BRIEF ANSWER: compliers: .4746978, always-takers=.456730, nevertakers=1-pc-pa=.069

(e) What are the average potential outcomes for the compliers, and what is the IV estimate of the return to medical school?

BRIEF ANSWER:

y0 compliers: 1.46/.47=3.07

y1 compliers: 1.54/.47=3.25

return = 3.25-3.07=.18

(f) Explain (and calculate if possible) what average potential outcomes are identified for the never-takers and always-takers.

BRIEF ANSWER: y1 for always takers, and y0 for never takers.

first note that

E[y*d|z=0]=E[y|d=1,z=0]p(d=1|z=0)=E[y1|a]*pa

from the constant in -regress v1 adm- we can directly get E[y*d|z=0] = 1.41, and therefore E[y1|a] = 1.41/.46=3.08

similarly note

E[y*(1-d)|z=1]=E[y|d=0,z=1]p(d=0|z=1)=E[y0|n]*pn

from -regress v2 adm- we can get E[y*(1-d)|z=1]=-1.46+1.66=.20 so E[y0|n] = .20/.069=2.91.

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 Miama, 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):

Group	*1979*	*1980*	*1981*
Miami:			
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)
Comparison Cities:			
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.

BRIEF ANSWER:

before-after = 9.6 - 8.3 = 1.3

did = (9.6 - 8.3) - (10.3 - 12.6) = 3.6

(b) Outline how you would estimate the difference-in-differences estimate in (a) using OLS.

BRIEF ANSWER:

estimate on the black 79, 81 sample

define a 19881 dummy : post

define miami dummy : miami

and estimate y = b0 + b1 * miami*post + b2* miami + b3 *post + e

(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.

BRIEF ANSWER:

common trend: employment rates of blacks would have moved as in comparison cities without the boatlift

threats f.e.

migration: boatlift causes low skilled people to move away

it is a 2x2 design so we need to rule out city*year shocks

(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.

BRIEF ANSWER:

first note that ρ_x the within group correlation of the treatment is 1 this means that if ρ_u, the importance of the variance of city level random shocks relative to the total residual variance, is non-zero we would like to cluster at the city level

[but note that this is not possible in a 2x2 design]