Exam 4137 – Spring 2021

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. Each subquestion carries equal weight.

---

1. Economists have long been interested to estimate the return to schooling. The Mincer equation is often used to explain the employment earnings as a function of schooling and labor market experience:

$$\log w_i = \beta_0 + \beta_1 s_i + \beta_2 x_i + \beta_3 x_i^2 + \varepsilon_i \tag{1}$$

where, $\log w_i$ is the log of hourly earnings, s is the years of schooling, and x is the years of potential labor market experience. Suppose that you have access to a cross-section data of 3010 individuals, which contains weekly earnings, weekly number of hours of worked, years of schooling and potential labor market experience, and some other variables for each individual.

   (a) Suppose that you have estimated eq.(1) using OLS on the available data, and obtained $\hat{\beta}_1 = 0.07$ with estimated standard error 0.003. How would you interpret this estimate?

   (b) One of your fellow students estimated a similar model, but she used weekly earnings instead of hourly earnings. However, her OLS estimate of β_1 is much higher than yours, what could be the reason?

   (c) You are worried that unobserved factors, such as individual ability, might impact both the hourly earnings and years of schooling. In particular, those with high ability tend to have higher earnings and obtain also more years of schooling. If this is case, what can you say about the estimate from eq.(1)?

   (d) To cope with the potential problem in (c), some authors propose an instrumental variables (IV) strategy whereby they instrument the variable schooling with a college-in-the-county dummy variable. Namely, for individual i

$$D_i = \begin{cases} 1 & \text{if between age 14 and 24 lived in a county with local colleges} \\ 0 & \text{otherwise} \end{cases}$$

   Specify and discuss the assumptions that need to hold for D_i to be a valid instrument. Are these assumptions likely to hold?

   (e) The estimates of both the reduced form regression and first stage regressions are given in the following table (standard error in parentheses). Calculate the IV estimates of β_1.

Dependent Variable	Regressor	Estimate
log (hourly earnings)		
	College in the county	0.042
		(0.018)
Years of schooling		
	College in the county	0.320
		(0.088)
N		3,010

2. In the following, we look at a different strategy to identify the causal effect of additional years schooling. In one study, the author utilized a policy change in the minimal school leaving age in Britain in 1947, in which the minimal school leaving age is changed from 14 to 15. The following figure shows the fraction of British born adults that left full-time education by age 14 and by age 15.

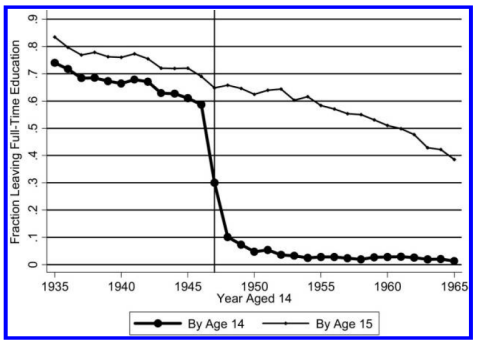


FIGURE 1. FRACTION LEFT FULL-TIME EDUCATION BY YEAR AGED 14 AND 15 (*Great Britain*)

Note: The lower line shows the proportion of British-born adults aged 32 to 64 from the 1983 to 1998 General Household Surveys who report leaving full-time education at or before age 14 from 1935 to 1965. The upper line shows the same, but for age 15. The minimum school-leaving age in Great Britain changed in 1947 from 14 to 15.

(a) The author used a Regression Discontinuity Design (RDD) to identify the causal effect he wants to study. Explain briefly the main idea of this identification strategy based on the setting presented above. What is the key assumption of this identification strategy? What type of causal effect is estimated in this case?

(b) For simplicity, assume that we only look at the individuals who became 14 years old at the year 1946 and 1948. Let S_i be an indicator for whether individual i leaves at age 14 ($S_i = 1$) or still attends school at age 15 ($S_i = 0$). Let Z_i be a binary variable with $Z_i = 0$ if the minimum school-leaving age for individual i equals 15, and $Z = 1$ if it equals 14. Let y_i be the log hourly earning when they were 45. Suppose that you have the following data:

		Obs	Average log hourly earnings at 45
$Z_i = 0$	$S_i = 0$	4500	1.65
$Z_i = 0$	$S_i = 1$	500	1.45
$Z_i = 1$	$S_i = 0$	2000	1.7
$Z_i = 1$	$S_i = 1$	3000	1.5

Compute an estimate of the RDD estimate of the causal effect of dropping out from school at age *14* and interpret the parameter.

(c) Define "compliers", "always takers", and "never takers" in this case. Then compute estimates of the fractions of the population that are "compliers", "always takers", and "never takers".

(d) Compute estimates of the mean log hourly earnings:

i. for always takers, if they drop out at 14

ii. for never takers, if they don't drop out at 14

iii. for compliers, if they drop out at 14

iv. for compliers, if they don't drop out at 14

3. A group of unemployed individuals are assigned to an online training program which lasts for 60 hours. Suppose that the time that each individual actually spends watching the program x is given by

$$f_X(x|\theta) = \begin{cases} \theta x^{\theta-1}/60^{\theta} & 0 < x \leq 60 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $\theta > 1$ is an unknown parameter. It is easy to verify that $E(X) = 60\theta/(\theta+1)$. We have collected information on 100 unemployed individuals and their time spent watching the program. See attached Stata output for detailed information of the data. (You might find the natural logarithm table attached useful.)

(a) Write down the log-likelihood functions, and outline how one can estimate the model using MLE.

(b) Suppose that we know that the true value of θ is in the set $\{2,3\}$. That is, θ takes either value 2 or 3. What is the $\hat{\theta}_{MLE}$ in this case?

(c) Suppose instead that θ can take all real values that is greater than 1, what will be the $\hat{\theta}_{MLE}$ and corresponding log likelihood value $lnL(x|\hat{\theta}_{MLE})$?

Suppose that we want study whether θ varies across gender, i.e., ask whether $\theta_f = \theta_m$.

(d) Specify a model where you can test the hypothesis that $\theta_f = \theta_m$. Based on the Stata output, will you reject the null hypothesis? (Hint, use the LR test!)

Stata Output

```
* data for problem 1;

. gen lnx=ln(x)
. su lnx x gender

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         lnx |        100    3.625334    .4754063   1.830871   4.089396
           x |        100    40.99326    14.26466   6.239318   59.70381
      gender |        100          .5    .5025189          0          1

 . tabstat lnx x, by(gender)

Summary statistics: mean
  by categories of: gender
  gender |       lnx         x
---------+--------------------
       0 |  3.775763  45.48103
       1 |  3.474904  36.5055
---------+--------------------
   Total |  3.625334  40.99326
------------------------------


Critical Values of Chi-square
df       .50      .25      .10      .05      .025     .01      .001
1        0.45     1.32     2.71     3.84     5.02     6.63     10.83
2        1.39     2.77     4.61     5.99     7.38     9.21     13.82
3        2.37     4.11     6.25     7.81     9.35     11.34    16.27
4        3.36     5.39     7.78     9.49     11.14    13.28    18.47
5        4.35     6.63     9.24    11.07     12.83    15.09    20.52
6        5.35     7.84    10.64    12.59     14.45    16.81    22.46
7        6.35     9.04    12.02    14.07     16.01    18.48    24.32
8        7.34    10.22    13.36    15.51     17.53    20.09    26.12
9        8.34    11.39    14.68    16.92     19.02    21.67    27.88
10       9.34    12.55    15.99    18.31     20.48    23.21    29.59
```

The natural logarithm table (Equal to or more than 1.0)

n	$\log_e n$	n	$\log_e n$	n	$\log_e n$	n	$\log_e n$
1.0	0.00000	3.0	1.09861	5.0	1.60944	25.0	3.21887
1.1	0.09531	3.1	1.13140	6.0	1.79176	26.0	3.25809
1.2	0.18232	3.2	1.16315	7.0	1.94591	27.0	3.29583
1.3	0.26236	3.3	1.19392	8.0	2.07944	28.0	3.33220
1.4	0.33647	3.4	1.22377	9.0	2.19722	29.0	3.36729
1.5	0.40546	3.5	1.25276	10.0	2.30258	30.0	3.40119
1.6	0.47000	3.6	1.28093	11.0	2.39789	40.0	3.68888
1.7	0.53063	3.7	1.30833	12.0	2.48491	50.0	3.91202
1.8	0.58779	3.8	1.33500	13.0	2.56495	60.0	4.09434
1.9	0.64185	3.9	1.36097	14.0	2.63905	70.0	4.24849
2.0	0.69314	4.0	1.38629	15.0	2.70805	80.0	4.38202
2.1	0.74193	4.1	1.41099	16.0	2.77259	90.0	4.49981
2.2	0.78845	4.2	1.43508	17.0	2.83321	100.0	4.60517
2.3	0.83291	4.3	1.45861	18.0	2.89037	200.0	5.29832
2.4	0.87547	4.4	1.48160	19.0	2.94444	300.0	5.70378
2.5	0.91629	4.5	1.50408	20.0	2.99573	400.0	5.99146
2.6	0.95551	4.6	1.52605	21,0	3.04452	500.0	6.21461
2.7	0.99325	4.7	1.54756	22,0	3.09104	600.0	6.39693
2.8	1.02962	4.8	1.56861	23.0	3.13549	700.0	6.55108
2.9	1.06471	4.9	1.58923	24.0	3.17805	800.0	6.68461