EXAM 1 Convocatoria Extraordinaria Miguel A. Delgado 02/07/20

Use the data in FERTIL2 in the attached file to answer this question. The data contains the number of children (*children*) and different socioeconomic variables for a sample of Boswana's women. Variables are described below.

- Question 1 (30%): Compute the average number of children for those without electricity and do the same for those with electricity (10%). Test whether the population means are the same using linear regression (10%). Can you infer that having electricity "causes" women to have fewer children? (10%).
- **SOLUTION:** We can get the average directly, but we prefer to use the regression model in order to answer both questions at once,

$$children = \gamma_0 + \gamma_1 electric + error.$$

The population mean for those with electricity is $\mu_1 = \gamma_0 + \gamma_1$ and the mean for those without electricity is $\mu_0 = \gamma_0$. Therefore, $\gamma_1 = \mu_1 - \mu_0$ is the difference between the means. The estimated model (heteroskedasticity robust SE in parenthesis) is

$$\widehat{children_i} = 2.32773 - 0.429202 \cdot electric_i \cdot (0.0362206) - (0.0967339)$$

Therefore, the average children for those with electricity is $\hat{\mu}_1 = 2.32773 - 0.42920 = 1.8985$, and the average for those without electricity is $\hat{\mu}_0 = 2.32773$. In order to test that the population means are the same, it suffices to test

$$H_0: \gamma_1 = 0 \text{ vs } H_1: \gamma_1 \neq 0,$$

The t ration is t = -0.429202/0.0967339 = -4.4369. Therefore, we reject H_0 at any significance level. We cannot infer that electricity "causes" that women have more children. The absence of electricity reflects a degree of deprivation, highly correlated with a lack of education, as well as cultural and behavioural issues. Therefore, there is an ommitted variable problem, and the significance of *electric* may be *spurious*. Therefore, we cannot conclude that there is a causal relation between *children* and electricity.

Question 2 (40%): Consider the model

$$children = \beta_0 + \beta_1 age + \beta_2 educ + \beta_3 electric + U, \tag{1}$$

where the error U is possibly heteroskedastic. How does the effect of having electricity compare with that in question 1? Is it still statistically significant? (10%). Now add age^2 , urban, and the three religious affiliation dummies. Is still statistical significant? (10%) Test that the effect of being catholic is identical to the effect of being protestant (20%)

SOLUTION: The OLS estimate of β_3 is -0.361758 with a robust SE equal to 0.0637644. Therefore, the *electric* effect is still significant at any significance level. The augmented model is

$$\begin{array}{lll} children &=& \beta_0 + \beta_1 age + \beta_2 educ + \beta_3 electric + \beta_4 age^2 + \beta_5 urban \\ &+ \beta_6 catholic + \beta_7 protest + \beta_8 spirit + error, \end{array}$$

Now the OLS β_3 estimates is -0.305719 with SE 0.0640662. Therefore, *electric* is still significant at any significance level. When testing

$$H_0: \beta_6 = \beta_7 \text{ vs } H_1: \beta_6 \neq \beta_7,$$

we get a p-value = 0.593982. Therefore, we cannot reject H_0 at any reasonable significance level.

Question 3 (30%): Add the interaction between *electric* and *educ*. Is the coefficient statistically significant? (15%) Is still the effect of having electricity statistically significant? (15%). Justify your answers.

SOLUTION: Now the model is

$$\begin{array}{lll} children &=& \beta_0 + \beta_1 age + \beta_2 educ + \beta_3 electric + \beta_4 age^2 + \beta_5 urban \\ &+ \beta_6 catholic + \beta_7 protest + \beta_8 spirit + \beta_9 educ \cdot electric + error, \end{array}$$

The β_3 estimate is -0.129092 with SE equal to 0.183415. Therefore, the coefficient is not significant at any reasonable significance level. However, if we test the significance of *electric* partial effect, we should consider the hypothesis

$$H_0: \beta_3 = \beta_9 = 0 \text{ vs } H_1: \beta_3 \neq 0 \text{ or/and } \beta_9 \neq 0,$$

the robust F statistic is F = 16.4238 with $p - value = 7.83508 \cdot 10^{-8}$. Therefore, we reject H_0 at any significance level.

Variables in FERTIL2

- 1. mnthborn : month woman born
- 2. yearborn : year woman born
- 3. *age* : age in years
- 4. electric = 1 if has electricity
- 5. radio = 1 if has radio

6.
$$tv = 1$$
 if has tv

- 7. bicycle = 1 if has bicycle
- 8. educ: years of education
- 9. ceb: children ever born
- 10. agefbrth: age at first birth

- 11. children: number of living children
- 12. knowmeth = 1 if know about birth control
- 13. usemeth = 1 if ever use birth control
- 14. monthfm: month of first marriage
- 15. yearfm: year of first marriage
- 16. agefm: age at first marriage
- 17. *idlnchld*: 'ideal' number of children
- 18. heduc: husband's years of education
- 19. $agesq: age^2$
- 20. urban=1 if live in urban area
- 21. *urbeduc*: urban*educ
- 22. spirit=1 if religion == spirit
- 23. protest=1 if religion == protestant
- 24. catholic = 1 if religion == catholic
- 25. frsthalf=1 if mnthborn ≤ 6
- 26. educ0 = 1 if educ = 0
- 27. evermarr = 1 if ever married