Unit 3 Wooldridge exercise C152

November 13, 2020

Econometrics II, Bachelor degree in Economics

Universitat Autònoma de Barcelona

Student, Roylan Martinez Vargas

Professor, Michael Creel

NIU: 1539069

C2 The data in FERTIL2 include, for women in Botswana during 1988, information on number of children, years of education, age, and religious and economic status variables.

(i) Estimate the model

$$children = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 age^2 + u$$

by OLS and interpret the estimates. In particular, holding *age* fixed, what is the estimated effect of another year of education on fertility? If 100 women receive another year of education, how many fewer children are they expected to have?

(ii) The variable *frsthalf* is a dummy variable equal to one if the woman was born during the first six months of the year. Assuming that *frsthalf* is uncorrelated with the error term from part (i), show that *frsthalf* is a reasonable IV candidate for *educ*. (*Hint*: You need to do a regression.)

(iii) Estimate the model from part (i) by using *frsthalf* as an IV for *educ*. Compare the estimated effect of education with the OLS estimate from part (i).

(iv) Add the binary variables *electric*, *tv*, and *bicycle* to the model and assume these are exogenous. Estimate the equation by OLS and 2SLS and compare the estimated coefficients on *educ*. Interpret the coefficient on *tv* and explain why television ownership has a negative effect on fertility.

[17]:
```
# TOOLS
import numpy as np
import pandas as pd
from statsmodels.formula.api import ols
from scipy.stats import t
```

[18]:
```
# Data without null values
dataExcel = pd.read_csv("fertility2.csv").dropna()
dataExcel
```

[18]:

	mnthborn	yearborn	age	electric	radio	tv	bicycle	educ	ceb	\
1	1	56	32	1.0	1.0	1.0	1.0	13	3	
2	7	58	30	1.0	0.0	0.0	0.0	5	1	
3	11	45	42	1.0	0.0	1.0	0.0	4	3	

```
4           5      45  43      1.0    1.0  1.0      1.0    11    2
5           8      52  36      1.0    0.0  0.0      0.0     7    1
…        …      …   …       …     …   …      …    …   …
4353       11      38  49      0.0    1.0  0.0      0.0     0    9
4354        8      57  31      0.0    1.0  0.0      0.0     2    3
4355        4      61  27      0.0    0.0  0.0      0.0     6    4
4359        1      62  26      0.0    1.0  0.0      0.0     0    1
4360       12      51  36      0.0    1.0  0.0      0.0     3    6

      agefbrth  …  heduc  agesq  urban  urbeduc  spirit  protest  catholic  \
1         25.0  …   12.0   1024      1       13       0        0         0
2         27.0  …    7.0    900      1        5       1        0         0
3         17.0  …   11.0   1764      1        4       0        0         0
4         24.0  …   14.0   1849      1       11       0        1         0
5         26.0  …    9.0   1296      1        7       0        0         0
…          …  …    …      …      …      …        …       …
4353      16.0  …    0.0   2401      0        0       0        1         0
4354      22.0  …    0.0    961      0        0       0        1         0
4355      18.0  …    7.0    729      0        0       0        1         0
4359      25.0  …    7.0    676      0        0       1        0         0
4360      22.0  …    0.0   1296      0        0       0        1         0

      frsthalf  educ0  evermarr
1            1      0         1
2            0      0         1
3            0      0         1
4            1      0         1
5            0      0         1
…          …     …        …
4353         0      1         1
4354         0      0         1
4355         1      0         1
4359         1      1         1
4360         0      0         1

[1719 rows x 27 columns]
```

i)

[19]:

```
# Initial model
modelA = ols("children ~ educ + age + agesq", data = dataExcel).fit()
print(modelA.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               children   R-squared:                       0.402
Model:                            OLS   Adj. R-squared:                  0.401
```

```
Method:                 Least Squares   F-statistic:                     383.7
Date:                Fri, 13 Nov 2020   Prob (F-statistic):          1.24e-190
Time:                        20:29:04   Log-Likelihood:                -3306.5
No. Observations:                1719   AIC:                             6621.
Df Residuals:                    1715   BIC:                             6643.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -6.1621      0.716     -8.608      0.000      -7.566      -4.758
educ          -0.1217      0.010    -12.729      0.000      -0.140      -0.103
age            0.4985      0.044     11.420      0.000       0.413       0.584
agesq         -0.0052      0.001     -8.005      0.000      -0.006      -0.004
==============================================================================
Omnibus:                       29.415   Durbin-Watson:                   1.807
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               55.751
Skew:                           0.026   Prob(JB):                     7.83e-13
Kurtosis:                       3.881   Cond. No.                     2.17e+04
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.17e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

The OLS model would be:
children = - 6.1621 - 0.1217 * **educ** + 0.4985 * **age** - 0.0052 * **agesq**
An additional year of education, keeping age fixed, would result in an expected reduction of 0.1217 in the number of children. In other words, if 100 women decide to receive another year of education then it would result in an overall expected reduction of approximately 12 children.

ii)

[20]:
```python
# Instrumental test model
modelB = ols("educ ~ frsthalf", data = dataExcel).fit()
print(modelB.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   educ   R-squared:                       0.018
Model:                            OLS   Adj. R-squared:                  0.017
Method:                 Least Squares   F-statistic:                     30.88
Date:                Fri, 13 Nov 2020   Prob (F-statistic):           3.17e-08
Time:                        20:29:04   Log-Likelihood:                -4904.9
No. Observations:                1719   AIC:                             9814.
Df Residuals:                    1717   BIC:                             9825.
```

```
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      5.8110      0.154     37.793      0.000       5.509       6.113
frsthalf      -1.1358      0.204     -5.557      0.000      -1.537      -0.735
==============================================================================
Omnibus:                       60.191   Durbin-Watson:                   1.257
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               58.300
Skew:                           0.409   Prob(JB):                     2.19e-13
Kurtosis:                       2.621   Cond. No.                         2.80
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

The t statistic of the **frsthalf** coefficient regressor of the OLS model of **educ** on **frsthalf** is -5.557 which means that them both have a statistically significant negative correlation. In fact, **frsthalf** explains arround a 2% of the variation in the sample.

iii)

[21]:
```python
# Initial model using
dataExcel['educIv'] = modelB.fittedvalues

modelC = ols("children ~ educIv + age + agesq", data = dataExcel).fit()
print(modelC.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               children   R-squared:                       0.348
Model:                            OLS   Adj. R-squared:                  0.347
Method:                 Least Squares   F-statistic:                     305.1
Date:                Fri, 13 Nov 2020   Prob (F-statistic):          1.08e-158
Time:                        20:29:05   Log-Likelihood:                -3380.2
No. Observations:                1719   AIC:                             6768.
Df Residuals:                    1715   BIC:                             6790.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -6.0302      0.834     -7.233      0.000      -7.665      -4.395
educIv        -0.2046      0.074     -2.759      0.006      -0.350      -0.059
age            0.5064      0.046     11.116      0.000       0.417       0.596
agesq         -0.0051      0.001     -7.616      0.000      -0.006      -0.004
```

```
==============================================================================
Omnibus:                       29.722   Durbin-Watson:                   1.697
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               43.523
Skew:                           0.180   Prob(JB):                     3.54e-10
Kurtosis:                       3.692   Cond. No.                     2.42e+04
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.42e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

The OLS model would be:
children = - 6.0302 - 0.2046 * **educ** + 0.5064 * **age** - 0.0051 * **agesq**
An additional year of education, keeping age fixed, would result in an expected reduction of 0.2046 in the number of children. In other words, if 100 women decide to receive another year of education then it would result in an overall expected reduction of approximately 20 children. As wee see now the expected reduction has become bigger, what might mean the OLS estimate of the initial model is too small.
It is important to note that as expected there is also a bigger standard error for **educ** and where if we perform this t test:
h0: b1 = -0.1217
h1: b1 != -0.1217
t test = | (- 0.2046 - 0.1217) / 0.074 | = 4.4094

[22]:
```
p_value = (1 - t.cdf(x=4.4094, df=1715)) * 2
p_value
```

[22]:
```
1.1009215059365474e-05
```

As we see the null hypothesis cannot be rejected and therefore the 99% confidence interval contains the original OLS estimate.

iv)

[23]:
```
modelD = ols("children ~ educ + age + agesq + electric + tv + bicycle", data =␣
 ↪dataExcel).fit()
print(modelD.summary())
modelE= ols("educ ~ electric + tv + bicycle", data = dataExcel).fit()
print(modelE.summary())
dataExcel['instEduc'] = modelE.fittedvalues
modelF = ols("children ~ instEduc + age + agesq", data = dataExcel).fit()
print(modelF.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:               children   R-squared:                       0.419
```

```
Model:                            OLS   Adj. R-squared:                  0.417
Method:                 Least Squares   F-statistic:                     205.5
Date:                Fri, 13 Nov 2020   Prob (F-statistic):          1.51e-197
Time:                        20:29:06   Log-Likelihood:                -3281.6
No. Observations:                1719   AIC:                             6577.
Df Residuals:                    1712   BIC:                             6615.
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -6.6770      0.713     -9.371      0.000      -8.074      -5.280
educ          -0.0882      0.011     -7.796      0.000      -0.110      -0.066
age            0.5158      0.043     11.914      0.000       0.431       0.601
agesq         -0.0053      0.001     -8.367      0.000      -0.007      -0.004
electric      -0.4323      0.136     -3.170      0.002      -0.700      -0.165
tv            -0.4672      0.160     -2.929      0.003      -0.780      -0.154
bicycle        0.3647      0.086      4.226      0.000       0.195       0.534
==============================================================================
Omnibus:                       30.035   Durbin-Watson:                   1.858
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               57.532
Skew:                           0.021   Prob(JB):                     3.21e-13
Kurtosis:                       3.895   Cond. No.                     2.19e+04
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.19e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
                            OLS Regression Results
==============================================================================
Dep. Variable:                   educ   R-squared:                       0.283
Model:                            OLS   Adj. R-squared:                  0.282
Method:                 Least Squares   F-statistic:                     225.9
Date:                Fri, 13 Nov 2020   Prob (F-statistic):          1.65e-123
Time:                        20:29:06   Log-Likelihood:                -4634.0
No. Observations:                1719   AIC:                             9276.
Df Residuals:                    1715   BIC:                             9298.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      4.0563      0.110     36.988      0.000       3.841       4.271
electric       2.9248      0.290     10.092      0.000       2.356       3.493
tv             4.1408      0.333     12.444      0.000       3.488       4.793
bicycle        0.2631      0.189      1.391      0.164      -0.108       0.634
```

```
==============================================================================
Omnibus:                       84.577   Durbin-Watson:                   1.608
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               43.857
Skew:                           0.213   Prob(JB):                     3.00e-10
Kurtosis:                       2.344   Cond. No.                         4.98
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
                            OLS Regression Results
==============================================================================
Dep. Variable:               children   R-squared:                       0.391
Model:                            OLS   Adj. R-squared:                  0.390
Method:                 Least Squares   F-statistic:                     366.7
Date:                Fri, 13 Nov 2020   Prob (F-statistic):          5.50e-184
Time:                        20:29:06   Log-Likelihood:                -3321.8
No. Observations:                1719   AIC:                             6652.
Df Residuals:                    1715   BIC:                             6673.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -6.7045      0.719     -9.320      0.000      -8.116      -5.294
instEduc      -0.2042      0.018    -11.348      0.000      -0.240      -0.169
age            0.5438      0.044     12.313      0.000       0.457       0.630
agesq         -0.0056      0.001     -8.607      0.000      -0.007      -0.004
==============================================================================
Omnibus:                       32.686   Durbin-Watson:                   1.814
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               56.247
Skew:                           0.134   Prob(JB):                     6.11e-13
Kurtosis:                       3.844   Cond. No.                     2.16e+04
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.16e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

The OLS model would be:
children = - 6.6770 - 0.0882 * **educ** + 0.5158 * **age** - 0.0053 * **agesq** - 0.4323 * **electric** - 0.467 * **tv** + 0.3647 * **bicycle**

The Auxiliar instrumental model would be:
instEduc = 4.0563 + 2.9248 * **electric** + 4.1408 * **tv** + 0.2631 * **bicycle**

The 2SLS model would be:

children = - 6.7045 - 0.2042 * **instEduc** + 0.5438 * **age** - 0.0056 * **agesq**

The coefficients **electric**, **tv** and **bicycle** seem to be correlated with **educ**, specially **electric** and **tv** that have a significance level of 99% and between them all explain arround 28% of the variation in the sample of instEduc. So they can be considered as rational instrumental variables.
Using the 2SLS model, tv has an effect on **educ** and consequently **educ** has an effect on the fertility variable **children**
tv seems to be related with **educ**, in other words, having a tv seems to be related on average with 4 additional years of education and this is intuitive since the poor people, the ones who do not have a tv, will probably not have the same resources to study as the richer ones.
Since having a **tv** is related on average with more education an a higher **educ** seems to be related with smaller amount of **children**, then we could consider the statement that having a tv is on average related with less children. This observation is reflected in the normal OLS model and in the 2SLS model.