MOOC Econometrics: Text Exercise 4

Mario Marchetti

April 26, 2020

Questions

A challenging and very relevant economic problem is the measurement of the returns to schooling. In this question we will use the following variables on 3010 US men:

- `logw`: log wage
- `educ`: number of years of schooling
- `age`: age of the individual in years
- `exper`: working experience in years
- `smsa`: dummy indicating whether the individual lived in a metropolitan area
- `south`: dummy indicating whether the individual lived in the south
- `nearc`: dummy indicating whether the individual lived near a 4-year college
- `dadeduc`: education of the individual's father (in years)
- `momeduc`: education of the individual's mother (in years)

This data is a selection of the data used by D. Card (1995)[1]

(a) Use OLS to estimate the parameters of the model:

$$\texttt{logw} = \beta_1 + \beta_2 \texttt{educ} + \beta_3 \texttt{exper} + \beta_4 \texttt{exper}^2 + \beta_5 \texttt{smsa} + \beta_6 \texttt{south} \quad (1)$$

Next page.

[1] "Using Geographic Variation in College Proximity to Estimate the Return to Schooling". In L.N. Christofides, E.K. Grant, and R. Swidinsky, editors, Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp. Toronto: University of Toronto Press, 1995.

The result of the regression are:

```
OLS, using observations 1-3010
Dependent variable: logw

            coefficient   std. error    t-ratio    p-value
  ------------------------------------------------------------
  const       4.61101      0.0678950     67.91    0.0000    ***
  educ        0.0815797    0.00349904    23.31    1.19e-110 ***
  exper       0.0838357    0.00677352    12.38    2.36e-34  ***
  exper2     -0.00220211   0.000323833   -6.800   1.26e-11  ***
  smsa        0.150801     0.0158360      9.523   3.35e-21  ***
  south      -0.175176     0.0146486    -11.96    3.12e-32  ***

Mean dependent var   6.261832   S.D. dependent var   0.443798
Sum squared resid    436.6819   S.E. of regression   0.381270
R-squared            0.263160   Adjusted R-squared   0.261934
F(5, 3004)           214.5742   P-value(F)           3.7e-196
Log-likelihood      -1365.617   Akaike criterion     2743.234
Schwarz criterion    2779.292   Hannan-Quinn         2756.202
```

The β_2 coefficient can be interpreted as: *the effect of one more year of education on the logarithm of wage*. Therefore, with each additional year of schooling the wage increases by about:

$$\mathrm{e}^{0.0816} \simeq 1.085 \simeq 8.5\%$$

.

(b) OLS may be inconsistent in this case as educ and exper may be endogenous. Give a reason why this may be the case. Also indicate whether the estimate in part (a) is still useful.
It is possible the wage, experience and education variables to be affected by some other variable (individual ability, intelligence, family income, etc.) in a way, such as, a higher ability lead to a higher wage, longer education ect. In this case, these variables would be endogenous and the OLS estimates would be biased and inconsistent.

(c) Give a motivation why `age` and $\texttt{age}^2$ can be used as instruments for `exper` and $\texttt{exper}^2$.
Because `age` and $\texttt{age}^2$ are positively correlated with `exper` and $\texttt{exper}^2$, and not effect directly wage. Obviously, the older a person, the higher the probability that he or she has more work experience. Age is also exogenous as it cannot be influenced by the people.

(d) Run the first-stage regression for educ for the two-stage least squares estimation of the parameters in the model above when age, age square , nearc, dadeduc, and momeduc are used as additional instruments. What do you conclude about the suitability of these instruments for schooling?
The 1SLS regression provides the following results:

```
First-Stage LS,OLS, using observations 1-3010
Dependent variable: educ

            coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const      -5.65235      3.97634     -1.421    0.1553
  age         0.989610     0.278714     3.551    0.0004   ***
  age2       -0.0170190    0.00483782  -3.518    0.0004   ***
  nearc       0.264554     0.0990846    2.670    0.0076   ***
  dadedec     0.190443     0.0156115   12.20     1.92e-33 ***
  momedec     0.234515     0.0170276   13.77     6.81e-42 ***
  smsa        0.529566     0.101504     5.217    1.94e-07 ***
  south      -0.424851     0.0910370   -4.667    3.19e-06 ***

Mean dependent var   13.26346   S.D. dependent var   2.676913
Sum squared resid    16245.17   S.E. of regression   2.326252
R-squared            0.246586   Adjusted R-squared   0.244829
F(7, 3002)           140.3614   P-value(F)           2.1e-179
Log-likelihood      -6808.218   Akaike criterion     13632.44
Schwarz criterion    13680.51   Hannan-Quinn         13649.73
```

The regression highlights that all the instrument are correlated with the endogenous variable `educ`. Especially `dadeduc, momed` has the highest t-ratio, so the hypothesis that parents highly educated encourage sons to achieve more education, seems confirmed.
Next page.

(e) Estimate the parameters of the model for log wage using two-stage least squares where you correct for the endogeneity of education and experience. Compare your result to the estimate in part (a).

```
2SLS, using observations 1-3010
Dependent variable: logw

               coefficient   std. error    t-ratio    p-value
  ------------------------------------------------------------
  const         4.41690      0.117861      37.48     1.59e-252 ***
  hat_educ      0.0998429    0.00671279    14.87     2.46e-48  ***
  hat_exper     0.0728669    0.0170667      4.270    2.02e-05  ***
  hat_exper^2  -0.00163929   0.000855864   -1.915    0.0555    *
  smsa          0.134937     0.0171240      7.880    4.54e-15  ***
  south        -0.158987     0.0160170     -9.926    7.18e-23  ***

Mean dependent var   6.261832   S.D. dependent var   0.443798
Sum squared resid    462.7457   S.E. of regression   0.392483
R-squared            0.219181   Adjusted R-squared   0.217882
F(5, 3004)           168.6487   P-value(F)           1.8e-158
Log-likelihood      -1452.866   Akaike criterion     2917.732
Schwarz criterion    2953.790   Hannan-Quinn         2930.700
```

We can see that both models are similar, and that both education and experience still have a positive effect while the squared experience still has a negative effect to `logw`. But with 2SLS `exper`^{2} lost his significance at 5% level of confidence (pvalue = 0.0555). The impact of education is now about 10% on log wage instead of 8% in the classical OLS. The impact of both expertise variables is slightly different.

(f) Perform the Sargan test for validity of the instruments. What is your conclusion?

```
Sargan over-identification test
  Null hypothesis: all instruments are valid
  Test statistic: LM = 3.70239
  with p-value = P(Chi-square(2) > 3.70239) = 0.157049
```

All instrument are valid!