Exercise Class - Econometrics Class 5

Instructor: Irene Iodice
Email: irene.iodice@malix.univ-paris1.fr

Part 1: Conclude exercises from last class

Part 2: Exercise from last year exam on IV

Consider the following regression model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \tag{1}$$

where $E[u_i|X_i] \neq 0$, that is with X is endogenous. Consider that in the database you have another variable Z.

1. Under which conditions is Z a valid instrument for X to be used in a TSLS regression?
2. Why having a valid instrument is important?
3. How would you test if Z is a weak instrument?
4. Imagine that the R^2 associated with $X_i = \pi_0 + \pi_1 Z_i + v_i$, with $E[v_i|Z_i] = 0$ is $R^2 = 0.1$ with $n = 50$. Assume that v is conditionally homoskedastic and discuss if Z is a weak instrument in this case.
5. If n=1000 would your answer to the previous question change? Comment.
6. Describe how would you test in this case the exogeneity of Z.

Part 3: We look to OVB exercise we started in the review class (those of you acquainted with the topic may leave!)

Pampilio Piratta is deciding if it is worthwhile to work an extra year at his enterprise or to go back to study for a master. With this aim he is exploring the relation between wage, education and tenure. Sadly Pampilio Piratta's research efforts are limited by the fact that he knows how to estimate linear regression models only if they contain one single regressor. Then he estimates the following three models:

i. $log(WAGE_i) = b_0 + b_1 EDUC_i + u_i$

```
> summary(lm(lwage ~ educ, data=wage1))

Call:
lm(formula = lwage ~ educ, data = wage1)

Residuals:
     Min       1Q   Median       3Q      Max
-2.21158 -0.36393 -0.07263  0.29712  1.52339

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.583773   0.097336   5.998 3.74e-09 ***
educ        0.082744   0.007567  10.935  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4801 on 524 degrees of freedom
Multiple R-squared:  0.1858,    Adjusted R-squared:  0.1843
F-statistic: 119.6 on 1 and 524 DF,  p-value: < 2.2e-16
```

ii. $log(WAGE_i) = a_0 + a_1 TENURE_i + w_i$

```
> summary(lm(lwage ~ tenure, data=wage1))

Call:
lm(formula = lwage ~ tenure, data = wage1)

Residuals:
     Min       1Q   Median       3Q      Max
-2.15984 -0.38530 -0.04478  0.32696  1.46072

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.501007   0.026866  55.870  < 2e-16 ***
tenure      0.023951   0.003039   7.881 1.89e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5031 on 524 degrees of freedom
Multiple R-squared:  0.106,     Adjusted R-squared:  0.1043
F-statistic: 62.11 on 1 and 524 DF,  p-value: 1.89e-14
```

iii. $TENURE_I = c_0 + c_1 EDUC_i + v_i$

```
> summary(lm(tenure ~ educ, data=wage1))

Call:
lm(formula = tenure ~ educ, data = wage1)

Residuals:
   Min     1Q Median     3Q    Max
-6.946 -4.894 -2.601  1.520 38.813

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.9457     1.4638   4.745 2.69e-06 ***
educ         -0.1466     0.1138  -1.288    0.198
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.22 on 524 degrees of freedom
Multiple R-squared:  0.003155,  Adjusted R-squared:  0.001253
F-statistic: 1.659 on 1 and 524 DF,  p-value: 0.1984
```

1. Explain why the OLS estimator of b_1 fails to produce an unbiased estimation of the true value of this parameter. Which of the OLS assumption is likely to be violated? Explain the meaning of this assumption.

2. Based on the results obtained above discuss, using the OVB formula asses whether b_1 is likely to be upward or downward biased. Then compute the value of the bias for the data at hand through the same formula and the results below from estimating the long model:
$$log(WAGE_i) = \beta_0 + \beta_1 EDUC_i + \beta_2 TENURE_i + \epsilon_i \quad (2)$$

3. Would it be possible for Pampilios Piratta to obtain an estimate of b_1 not affected by this OVB but without estimating a linear model with both EDUC and TENURE? Check your answer with the results provided.

```
> summary(lm(lwage ~ educ+tenure, data=wage1))

Call:
lm(formula = lwage ~ educ + tenure, data = wage1)

Residuals:
     Min       1Q   Median       3Q      Max
-2.10350 -0.29287 -0.04081  0.28672  1.44967

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.404474   0.091696   4.411 1.25e-05 ***
educ        0.086528   0.006991  12.377  < 2e-16 ***
tenure      0.025814   0.002680   9.634  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4428 on 523 degrees of freedom
Multiple R-squared:  0.3085,    Adjusted R-squared:  0.3059
F-statistic: 116.7 on 2 and 523 DF,  p-value: < 2.2e-16
```

```
> summary(lm(lwage ~ u_eductenure.hat, data=wage1))

Call:
lm(formula = lwage ~ u_eductenure.hat, data = wage1)

Residuals:
     Min       1Q   Median       3Q      Max
-2.20181 -0.35600 -0.06182  0.30338  1.50708

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.62327    0.02072   78.36   <2e-16 ***
u_eductenure.hat 0.08653    0.00750   11.54   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4751 on 524 degrees of freedom
Multiple R-squared:  0.2025,    Adjusted R-squared:  0.201
F-statistic: 133.1 on 1 and 524 DF,  p-value: < 2.2e-16
```