TD4 : Endogeneity and instrumental variables (computer session)

Exercise 1 (from Wooldridge (2009))**:**
Use the data set WAGE2_1314, which gives individual characteristics for 935 men, to estimate return to education.

1. Estimate the model M0: $Ln(wage) = a + b.educ + u$. Comment.
2. Estimate the previous model using *siblings* as an instrument for *educ*.
3. The variable *brthord* is birth order (=1 for a first-born child, 2 for a second-born child, …). Explain why *educ* and *brthord* might be negatively correlated. Regress *educ* on *brthord* to determine whether there is a statistically significant negative correlation.
4. Use *brthord* as an IV for *educ* in model M0. Comment.
5. We include now *siblings* as an explanatory variable in model M0, and we assume it is exogenous. We want to use *brthord* as an IV for educ. State and test the identification assumption.
6. Estimate the model specified in question 5. Comment.
7. Using fitted values, $\widehat{educ}$, from question 5, compute the correlation between $\widehat{educ}$ and *sibs*. Use this result to explain your findings in question 6.
8. The purpose of this question is to compare the estimates and standard errors obtained by correctly using 2SLS with those obtained using inappropriate procedure.
   (a) Use the TSLS estimation method to estimate the model:
   $$Ln(wage) = a_0 + a_1\, educ + a_2\, exper + a_3\, tenure + a_4\, black + u$$
   with *sibs* as an IV for *educ*.
   (b) Now, manually carry out 2SLS. That is perform the 2 steps using OLS separately. Comment.
   (c) Now, use the following two-step procedure. In step 1, regress educ on sibs only and obtain the fitted values, denoted $\widetilde{educ}$ (note this is an incorrect first stage regression). In a second step, run the regression of Ln(wage) on $\widetilde{educ}, exper, tenure\ and\ black$. Comment.

Exercise 2 (from Wooldridge (2009))**:**
Use the data set MROZ_1314, which gives individual characteristics for 428 worked women, to estimate return to education.
We consider the model:
$$M0:\ Ln(wage) = a_0 + a_1\, educ + a_2\, exper + a_3\, exper^2 + u$$

1. Estimate the model *M0* where we suspect *educ* to be endogenous, using *fatheduc* and *motheduc* as instruments.
2. State and test the identification assumption.
3. Test for endogeneity of *educ*.
4. Test overidentification restrictions.