Computer Exercise 2: Instrument Variable Estimation

Use the data `wage2.csv` and script `iv.r` to study the following model

$$\begin{aligned} log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 married \\ + \beta_5 south + \beta_6 urban + \epsilon, \end{aligned} \tag{1}$$

1. The variable *brthord* is birth order (*brthord* is one for a first-born child, *two* for a second-born child, and so on). The variable *sibs* is the number of siblings the individual has. Explain why *educ* might be negatively correlated with *brthord* and *sibs*.

2. Regress *educ* on *brthord*, *sibs* and the other explanatory variables in (1) to determine whether *brthord* and *sibs* are jointly significant.

3. Do you expect there will be any bias while estimating (1) by OLS? Do a Hausman specification test under homoskedasticity.

4. Use *brthord* and *sibs* as IVs for *educ* to estimate (1). Carry out an overidentification test to check whether *brthord* and *sibs* are valid instruments under homoskedasticity.

5. Study how `tsres2` is obtained. Are there any differences in the results of `tsres2`, `ivres` and `tsres`?

6. Study how `btsvc` is calculated. Divide it by `vcov(tsres2)` and `vcov(ivres)`. Can you explain why the standard errors in `tsres2` and `ivres` are different from those in `tsres` now?

7. How `btsvc2` is calculated? If you divide it by `vcov(tsres)`, what can you notice?

8. Can you see any difference between `btsls` and `btsls2`? Which estimator do you think would be more efficient asymptotically?