Submission of Econometrics Assessment: Instrumental Variable Regression by 2406461242

In this exercise, we will continue to examine the*"return on education"* which consists as a simple function of wage and education. The difference here is that we will implement instrumental variable regression. We will use IFLS 5 data, and follow the specification of paper*"Purnastuti, L. (2013). Instrumenting Education and Returns to Schooling in Indonesia. Jurnal Economia, 9(2), 166-174."*

The model specification is as follows:

$$Income = \beta_0 + \beta_1 SchYrs + \beta_2 Age + \beta_3 Age^2 + \beta_4 Tenure + \beta_5 Tenure^2 + \beta_6 Urban + \beta_7 Female + u$$

Variable	Definition
Income	Monthly income for the last job
SchYrs	Years of schooling
Age	Individual age in years
Month of Birth	IV for schyrs
Tenure	Time spent for the primary job in years
Urban	0=non-urban ; 1=urban
Female	0= male; 1= female

For the instrument, we will use the month of birth, replicating *"Angrist, J. D., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings?. The Quarterly Journal of Economics, 106(4), 979-1014."* Due to compulsory education, different children starts school at different ages, and it affects their schooling. We define the following specification:

$$SchYrs = \beta_0 + \beta_1 BirthMonth + u$$

Where the BirthMonth is the birth month of the individual: 1 to 12.

(1) First, identify which *book* and *dta* files you need from the codebooks. You can also directly see the questionnaires.
⇒Only use book 3A (b3a_tk2.dta ; b3a_dl1.dta) and book K (bk_ar1.dta ; bk_sc1.dta)

(2) Next, do some data cleaning. You need to **recode the urban** and **female data into dummies**, and **birthmonth** according to the number of the respective months. The SchYrs also need to be created from categorical variables. Assume that Pesantren has 12 years of education (similar with Education C, SLA, SMA, SMK, etc.)
⇒Attached to the do-file.

(3) Run the regression using OLS and interpret the results. What did you find?
⇒ In testing using OLS, several things can be analyzed:
1. Schyrs is worth 239873.4426, which means that every 1-year increase in education will increase income by 239873 units.

2. Age is worth -35393, which means that every increase in age will decrease income by 35393 units
3. Age2, to see the linearity of the model, is worth 636.6921. However, in both of these age variables, it is not significant to income. So, it can be indicated that there is no relationship between income and changes in age.
4. Tenure is worth 159844.1400, meaning that every tenure increase will increase income by 159844 units. This also applies to Tenure2, which is significant to income but has a different sign. This can indicate an endogeneity problem.
5. Location of region (Urban or rural) and gender are not significant to income. This means there is no relationship between these two variables and changes in income.

(4) Run the IV regression and compare it with the previous one. Is the results indicate endogeneity for SchYrs? Why?
⇒ After using the instrumental variables (month of birth), all independent variables aren't significant. In several variables, signs of causality tend to change. It indicates an endogeneity problem in this model, and IV is considered inappropriate to fix this endogeneity problem.

Table 1. Regression Results

	(1) OLS	(2) IV
Schyrs	239873.4426***	16629693.1239
	(43925.7746)	(39428366.9499)
Age	-35393.2357	-33824.3911
	(30819.9001)	(129933.1878)
Age2	636.6921	5110.8816
	(444.7207)	(10925.2102)
Tenure	159844.1400**	-1335298.3635
	(50441.2702)	(3603043.6154)
Tenure2	-3642.9828*	45216.1893
	(1546.6175)	(117717.9265)
Urban	-639016.7656	26713804.3152
	(373133.2161)	(65819703.1885)
Female	25527.5641	164449.0631
	(344143.4101)	(1488264.7732)
_cons	-392878.5578	-1.6995e+08
	(725173.4669)	(4.0791e+08)
N	8307	8307
R-sq	0.007	.

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001