Joel Cabrera

Advanced Cross-Sectional and Panel Econometrics (01:220:401)

Professor Piehl

October 6, 2019

**Introductory Econometrics: A Modern Approach 5th ed. – HW #2 Problems**

1a. μ is the error term that captures the effects of all other possible independent variables that are not included in the regression model. An example would be family income. A student who comes from an affluent background might be associated with a higher GPA (e.g. they can afford private tutors, can take more classes, etc.). Family income, then, could be correlated with PC ownership, in that wealthy students can afford a PC, as opposed to those who are poor, who cannot afford a PC.

1b. As similarly explained in the previous problem, annual parental income could be correlated with PC ownership in that: those who come from wealthy families are more likely to afford a PC than those who do not. In order for annual parental income to be an ideal instrumental variable (IV), it must satisfy two conditions: it must be correlated with the independent variable of interest, PC ownership (Corr(Z,X) != 0), and it must not be correlated with the error term (and thus, not correlated with the dependent variable, GPA), μ (Corr(Z, μ) = 0). While parental income is likely to be correlated with PC ownership, parental income is still likely to be correlated with the error term – and thus, also be correlated with GPA. Again, as explained in the previous problem, a higher parental income is likely to be associated with a higher GPA.

1c. Based on the information given, we can develop a binary IV called “Grant” – where 0 = did not receive university grant and 1 = did receive university grant. Grant is likely to be correlated with PC ownership in that receiving a grant allows one to obtain a PC, whereas not receiving a grant does not allow one to obtain a PC. Grant is also unlikely to be uncorrelated with μ, as grants were randomly assigned to students. Because of this, it can also be said that grant is likely to be uncorrelated with parental income.

7i. Choice school, despite having family income included in the regression model, might still be correlated with μ. This is because school choice can be correlated with other possible variables that can also explain score. An example of this is a student’s “ability” – where 0 = not capable and 1 = capable. A student who is capable (e.g. does all homework, gets good grades, etc.) are more likely to attend a choice school than those who are not capable, holding family income constant.

7ii. If the grant amounts were assigned randomly within each income class, grant would be uncorrelated with μ. Random assignment of the grants allows the “grant” variable to be uncorrelated with other possible explanatory variables (e.g. “ability”).

7iii. The reduced form equation for choice school – otherwise known as the first stage of two-stage least squares (2SLS) regression for choice school – would be as follows:

*choice* = γ + δ1 \* *faminc*+ δ2 \* *grant* + *v1*

In order for grant to be partially correlated with choice, we would have to find an IV for grant – that is, the IV must be correlated with grant but not correlated with *v*1 (and, thus, only correlated with *choice* through *grant*).

7iv. The reduced form equation for score, following the similar model constructed above, would be as follows: ρ

*score =* ρ + β \* *faminc* + π \* *grant* + *v*2

(Q: Shouldn’t *choice* from 1st equation be included here instead of *grant*?)

This equation for score is useful because it allows us to estimate the effects of given grant amounts on standardized test scores. Interpreting the coefficient on grant: an increase in one’s grant is associated with an increase in one’s score on a statewide test.

9. In equation 15.8, we have the following model:

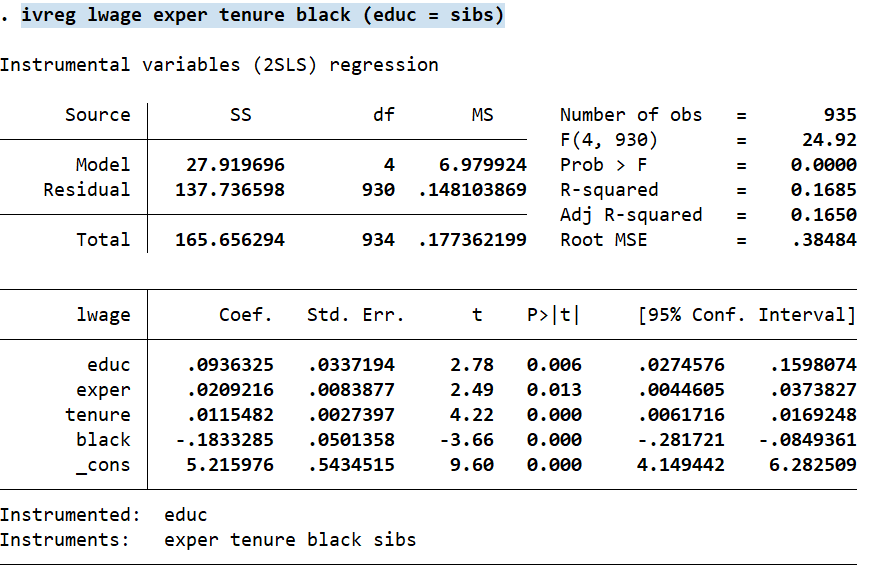
*score =* β0 + β1 \* *skipped* + μ

If we are unable to find a good instrumental variable for skipped, but possess information on combined SAT score and cumulative GPA prior to semester, then we reconstruct our model as a multivariate regression model. A model containing these additional independent variables would appear as the following:

*score =* β0 + β1 \* *SAT* + β2 \* *cumGPA* + β3 \* *skipped* + μ,

where *SAT* is combined SAT score and *cumGPA* is cumulative GPA prior to the semester. Adding these independent variables in our regression model would reduce μ, and thus, have a more accurate prediction of *score*. But, in order to find the causal effect of skipped on score, we would still need to find an instrument for the former.

C9i. Typing in the following code “ivreg lwage exper tenure black (educ = sibs)” in order to perform a 2SLS regression, we get the following output:



Assuming “usual form” means to translate the coefficient results into a readable regression model, we develop the following:

log(*wage*) = 5.22 + 0.02\**exper* + 0.01\**tenure* – 0.18\**black* + 0.09\**edu ­+* μ

(Note: all coefficients have been rounded to the nearest two decimals).

(Q: How to generate the error term in Stata output?).

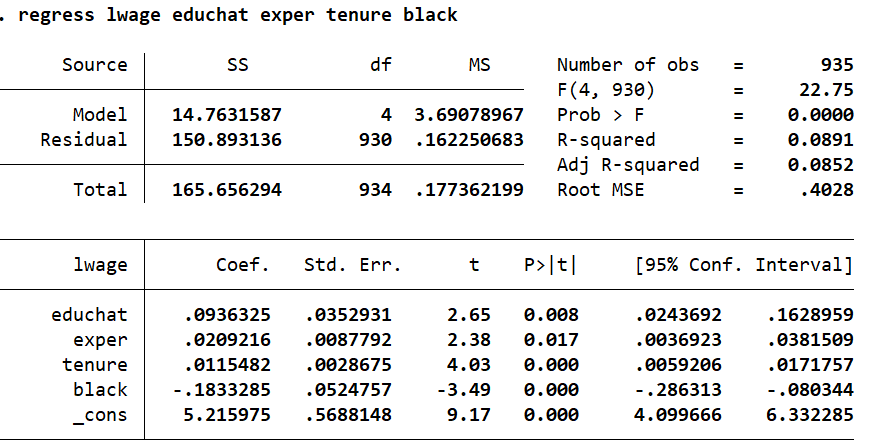
C9ii. Using the following code in Stata to manually perform 2SLS,

regress educ sibs exper tenure black

predict educhat

regress lwage educhat exper tenure black

we ultimately receive the following output:



Comparing the Stata output in C9i. and C9ii., we can see that the beta coefficients on independent variables are identical; however, their respective standard errors are somewhat different. Thus, in order to perform a successful 2SLS regression, we would have to do so in one go, rather than doing as such manually.

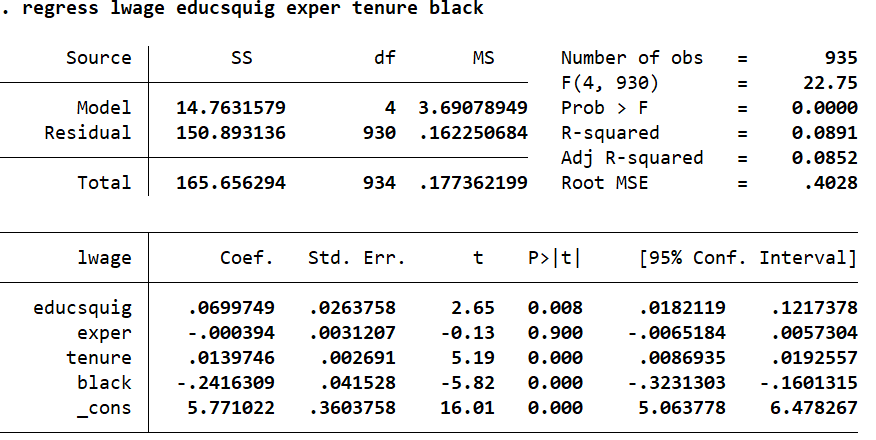
C9iii. Using the following code in Stata to manually perform 2SLS,

regress educ sibs

predict educsquig

regress lwage educsquig exper tenure black

we ultimately receive the following output:



We can see that the coefficient on and the standard error for education is approximately 0.07 and 0.03, respectively. Compared to the output in 9Ci., this not only reduces the estimates of the coefficient on and standard error of education, but also yields a different interpretation of the effect of education on wages, controlling for all other independent variables. The new interpretation would be: a one year increase in education is associated with a 7% (rather than 9%) increase in wages. In order to have accurate estimates and interpretations of our regression model, we would need to perform 2SLS all at once, rather than manually.