Applied Economics
IV in gretl

It is necessary to generate a script (.inp file) that allows you to answer each of the questions.

1. We want to analyze the effect of cigarette smoking on child birth weight. We use the file bwght.gdt. Without other explanatory variables, the model is:

$$log(bwght) = \beta_0 + \beta_1 packs + u$$

where *bwght* is the child's birth weight and *packs* is the number of packs smoked by the mother per day.

   (a) Estimate this model and give an interpretation for $\hat{\beta}_1$. Do you think that $E(u|packs) = 0$ is a reasonable assumption? Mention two factors that might be included in u.

   (b) A more interesting model would include more variables:

$$log(bwght) = \beta_0 + \beta_1 packs + \beta_2 male + \beta_3 parity + \beta_4 lfaminc + u$$

   where *male* is a dummy variable taking the value 1 when the baby is male, *parity* represents the birth order of the child and $lfaminc$ represents the logarithm of the family income (measured in thousands of dollars).

   Estimate this new model and give an interpretation for $\hat{\beta}_1$. Does this estimation differ from the previous one? Do you think that $E(u|packs, male, parity, lfaminc) = 0$ is a reasonable assumption?

   (c) If we believe that *packs* might be correlated with u (for example we might worry that cigarette consumption is correlated with other health factors or good prenatal care), using an instrumental variable (IV) is probably a good idea. A potential IV for *packs* is the average price of cigarettes in the state of residence, *cigprice*. If cigarettes are a typical consumption good, basic economic theory suggests that *packs* and *cigprice* are negatively correlated, so that *cigprice* can be considered a relevant IV for packs.

   Estimate the previous model using *cigprice* as an IV for *packs*. What is your conclusion about the effect of cigarette smoking on birth weight?

   (d) But, is *cigprice* a good instrument? What are the conditions that a good instrument should verify? Test the conditions (if possible) and decide if *cigprice* is a good instrument or not.

   (e) Which estimation do you prefer, OLS or IV? Explain.

2. We are going to use the file wage2.gdt to investigate the usual wage equation that characterizes the return to education:

$$lwage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 cap + \epsilon, \quad (1)$$

where $E(\epsilon|educ, exper, tenure, cap) = 0$.

Variables we need:

$lwage$ = logarithm of the monthly wage (in thousand dollars);

$educ$ = completed years of education;

$exper$ = labor experience (in years);

$tenure$ = tenure in last job (in years);

cap = ability or capacity.

The expected effects of these four variables on wages are positive. We assume that education and ability are positively correlated: $C(educ, cap) > 0$, and that both experience and tenure are not correlated with education and ability. However, ability is unobservable, so the model that can actually be estimated is:

$$lwage = \gamma_0 + \gamma_1 educ + \gamma_2 exper + \gamma_3 tenure + \varepsilon \tag{2}$$

Nevertheless, there is the possibility of using an imperfect measure of ability. An available measure in our sample is the individual's intelligence quotient (IQ).

In this dataset we also have information about the following variables (assumed to be uncorrelated with ϵ): $meduc$ (mother's years of education) and KWW (score attained by the individual in a general knowledge test).

(a) Although we are interested in the coefficients of equation (1), given that ability is unobservable, suppose that we consider the OLS estimation of equation (2). What would you expect about the estimation of the effect of education on wages? Estimate the model in equation (2) and give an interpretation of the coefficients.

(b) Assume that the intelligence quotient (IQ) is exactly equal to ability ($IQ = cap$), and estimate equation (1). For a given experience, tenure and ability, compute the effect of an additional year of education on wages.

(c) Evaluate if IQ is an endogenous variable. Explain your answer.

(d) Evaluate if $meduc$ and KWW are valid instruments for IQ. Explain your answer.

(e) Decide which method is better in this case, OLS or IV. Compute the appropriate return to education.

3. How fertility affects the work supply? That is, how much the work supply of a woman is reduced when she has one more child? In this exercise, we will estimate this effect using data of married women in the US census of 1980. The dataset fertility.gdt contains information on married women between 21 and 35 years old with two or more children. The variable $weeksm1$ represents the weeks worked (work supply), $morekids$ is a dummy variable taking the value one if the woman has more than two kids.

(a) Run the regression of the variable $weeksm1$ on the indicator variable $morekids$ by OLS. On average, do women with more than two children work less than women with two children? How much less?

(b) Explain why the OLS regression estimated before is not appropriate to estimate the causal effect of fertility on the work supply.

(c) The database contains the variable *samesex*, which is equal to 1 if the first two children are the same gender (boy-boy or girl-girl) and is equal to 0 otherwise. Have the couples whose first two children are of the same sex a larger probability of having a third child? Is this effect large? Is statistically significant?

(d) Explain why the variable *samesex* is a valid instrument for the variable *morekids*.

(e) Estimate the regression of the variable *weeksm*1 on the variable *morekids* using the variable *samesex* as an instrument. Which is the effect of fertility on the work supply?

(f) Does the result change when the variables *black*, *hispan*, and *othrace* are included in the regression of the work supply (considering these variables as exogenous)? Explain.