Econ 448
Instrumental variables - Human Capital

Pedro Hancevic

1 Instrumental Variables

Recall from previous discussion the linear regression model

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{1}$$

with its main assumption: $E(\varepsilon|x) = E(u) = 0$. This assumption is telling us that for any given x, the average of the unobservables is the same, and by assumption (normalization) it is equal to zero. The Ordinary Least Squares Estimators (OLS) presented in our previous meeting were crucially based on this assumption. It turns out that in many cases this assumption is not satisfied and then we have serious *inconsistent* and *biased* estimates which in turn might cause erroneous conclusions in our empirical research. For example, suppose that the true model is

$$Growth = \beta_0 + \beta_1 schooling + \beta_2 schooling^2 + \beta_3 ability + \varepsilon$$

where we expect that in reality ability affects growth positively ($\beta_3 > 0$). Whenever the zero conditional mean assumption holds for an explanatory variable, we say it is exogenous. When this assumption does not hold, we say that the explanatory variable is endogenous. For example, the latter case occurs if ability is not fully observable and may be part of the error term ε or may be correlated with education. This first cause of endogeneity is referred as the *omitted variable bias*. How could we assess the bias? Well, if we suppose education and ability are positively correlated , then if education is increased , so is ability (e.g. selection by schools). Therefore, by not considering the effect of ability on growth we assign an additional effect of education on growth that should have been attributed to ability instead.

Another potential source of endogeneity is the *reverse causality problem*: the left-hand-side variable may be affecting the right-hand-side variable. For example countries that growth faster can afford higher spending in education. In this case we could end up overestimating the causal effect of education on growth since the effect is positive in both directions.

One way to deal with the endogeneity problem is to use *instrumental variables* (IV). Consider an instrumental variable z, the conditions for a good instrumental variable are:

1. Instrument relevance condition: the instrumental variable z is correlated with the endogenous variable x.

2. Instrument exogeneity: the instrumental variable has no correlation with the error term, i.e. $E(\varepsilon|z) = 0$.

Intuitively, it is as if the endogenous variable x had two components: one component that is correlated with ε and one component that is uncorrelated with it. Hence, using IV method, amounts to only considering the exogenous component of x. Example: an instrument for schooling could be "the number of siblings per individual" which affects the years of education negatively but is uncorrelated with the error term (and ability in particular).

Two-stage least squares estimation method (2SLS):

1. Regress $x = \gamma_0 + \gamma_1 z + \eta$ by OLS method. We then obtain the part of x that is predicted by z, i.e. the exogenous part of x. Specifically, we get $\hat{x} = \hat{\gamma}_0 + \hat{\gamma}_1 z$
2. Regress $y = \beta_0 + \beta_1 \hat{x} + \varepsilon$ again by OLS.

2 Human Capital and Education

Highest Level of Education	**Years of schooling**	**Wage Relative to No Schooling**	**Percentage of the Population** Developing Countries	Advanced Countries
No Schooling	0	1.00	34.4	3.7
Incomplete Primary	4	1.65	22.6	11.7
Complete Primary	8	2.43	11.9	13.4
Incomplete Secondary	10	2.77	16.3	26.5
Complete Secondary	12	3.16	8.3	16.6
Incomplete Higher	14	3.61	3.5	15.1
Complete Higher	16	4.11	3.0	13.0

Source: Barro and Lee (2000).