UCL - Department of Economics *G003 - Microeconometrics*
Mónica Costa Dias *Spring 2008*

Problem set 1: Instrumental Variables

1. Suppose you wish to measure the impact of smoking on the weight of newborns. You are planning to use the following model,

$$\log(bw_i) \quad = \quad \beta_0 + \beta_1 male_i + \beta_2 order_i + \beta_3 y_i + \beta_4 cig_i + \epsilon_i$$

where bw is the birth weight, $male$ is a dummy variable assuming the value 1 if the baby is a boy or 0 otherwise, $order$ is the birth order of the child, y is the log income of the family, cig is the amount of cigarettes per day smoked during pregnancy, i indexes the observation and the β's are the unknown parameters.

   (a) What could be the problem in using OLS to estimate the above model?

   (b) Suppose you have data on the average price of cigarettes in the state of residence. Would this information help to identify the true parameters of the model?

   (c) Use data on BirthWeight.raw to estimate the model above. Use OLS and 2SLS. Discuss the results.

   (d) Estimate the reduced form for cig. Discuss.

2. Consider the model of earnings,

$$\ln y_i \quad = \quad \beta_0 + \beta_1 exp_i + \beta_2 educ_i + \epsilon_i$$

where y is hourly earnings, exp is accumulated experience, $educ$ is highest qualification obtained, ϵ is the unobservable component of the model, i indexes the observation and the β's are the unknown parameters.

   (a) What could be the problem in using OLS to estimate the above model?

   (b) Suppose you have data on two additional variable, distance to nearest school ($dschool$) and distance to nearest college ($dcollege$) at 16 years of age. Discuss whether these are likely to be good instruments.

   (c) Suppose you use $dschool$ and $dcollege$ as instruments in the estimation of the above model. Write down the reduced form for $educ$ and state the conditions under which the parameters of the model above are identified.

(d) To test that *dschool* and *dcollege* are in fact uncorrelated with ϵ it was suggested to use OLS on the equation,

$$\ln y_i = \beta_0 + \beta_1 exp_i + \beta_2 educ_i + \alpha_1 dschool_i + \alpha_2 dcollege + \mu_i$$

and test $\alpha_1 = \alpha_2 = 0$. Would this method work? Why?

(e) How can the assumption $\text{rank}\,(Z'X) = k$ be tested?

3. Consider the following model,

$$y = z_1\beta + w\alpha + \epsilon$$

where $E(z\epsilon) = 0$ and $z = (z_1, z_2)$ is vector of exogenous variables. The variable w is endogenous: $E(w\epsilon) \neq 0$.

Suppose we use the following procedure to estimate (β, α):

Step 1: Regress w on z_2 and obtain the fitted values, $\widehat{w}$.

Step 2: Regress y on $(z_1, \widehat{w})$ and obtain $(\widehat{\beta}, \widehat{\alpha})$.

(a) Will $(\widehat{\beta}, \widehat{\alpha})$ be generally consistent? Show.

(b) When will $(\widehat{\beta}, \widehat{\alpha})$ be consistent?

4. Consider the regression model

$$y = \beta_0 + \beta_1 x + \epsilon$$

where x is endogenous and z is a binary instrument for x.

(a) Show that the IV estimator for β_1 is,

$$\beta_1^{IV} = \frac{\overline{y}_1 - \overline{y}_0}{\overline{x}_1 - \overline{x}_0}$$

where $\overline{y}_d$ and $\overline{x}_d$ and the averages of y and x, respectively, when $z = d$.

(b) What is the interpretation of β_1^{IV} if x is also binary (say, if it represents participation in treatment)?

5. Suppose you wish to estimate β in

$$y_i = \alpha + x_i\beta + u_i$$

(a) Derive the consequences for the OLS estimator of y being measured with error which is independent of x.

(b) Instead of measuring x you measure x^* where $x_i^* = x_i + \epsilon_i$ and ϵ_i is a measurement error which is independent across individuals and independent of x. Show that the OLS estimator converges asymptotically to $\delta\beta$ where $0 \leq \delta \leq 1$. Explain the implication of this result for estimating the elasticity of hours worked with respect to wages when wages are measured with iid errors.

6. Show that $W = E\left[u_i^2 z_i' z_i\right]^{-1}$ is the optimal weighting matrix for the GMM estimator.