Problem Set 4

Applied Statistics and Econometrics II
Spring 2018, NYU
Ercan Karadas

(Due: March 1, in class)

[1] Consider the structural equation

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i \tag{1}$$

with x_i treated as endogenous so that $E[x_i\varepsilon_i] \neq 0$. Assume y_i and x_i are scalar. Suppose we have a scalar instrument z_i which satisfies

$$E[x_i\varepsilon_i] \neq 0$$

so in particular $E[\varepsilon_i] = 0$, $E[z_i\varepsilon_i] = 0$ and $E[z_i^2\varepsilon_i] = 0$.

a) Should x_i^2 be treated as endogenous or exogenous?

b) Suppose we have a scalar instrument z_i which satisfies

$$x_i = \gamma_0 + \gamma_1 z_i + u_i \tag{2}$$

with u_i independent of z_i and mean zero.
Consider using $[1\ z_i\ z_i^2]$ as instruments. Is this a sufficient number of instruments? (Would this be just-identified, over-identified, or under-identified)?

c) Write out the reduced form equation for x_i^2. Under what condition on the reduced form parameters (2) are the parameters in (1) identified?

[2] Consider the model

$$\begin{aligned} y_i &= \mathbf{x}_i'\boldsymbol{\beta} + e_i \\ E(e_i|\mathbf{z}_i) &= 0 \end{aligned}$$

where y_i is scalar and $\mathbf{x}_i$ and $\mathbf{z}_i$ each a $k \times 1$ vector. You have a random sample $\{y_i, \mathbf{x}_i, \mathbf{z}_i\}_{i=1}^n$

a) Write the IV estimator $\widehat{\boldsymbol{\beta}}_{IV}$ for $\boldsymbol{\beta}$

b) Suppose that $\mathbf{x}_i$ is exogenous in the sense that $E(e_i|\mathbf{z}_i, \mathbf{x}_i) = 0$. Is $\widehat{\boldsymbol{\beta}}_{IV}$ unbiased for $\boldsymbol{\beta}$?

c) Continuing to assume that $\mathbf{x}_i$ is exogenous, find the variance matrix for $\widehat{\boldsymbol{\beta}}_{IV}$, var $\left(\widehat{\boldsymbol{\beta}}_{IV}|\mathbf{X}, \mathbf{Z}\right)$.

[3] Use the data in 401KSUBS for this exercise. The equation of interest is a linear probability model:

$$pira = \beta_0 + \beta_1\, p401k + \beta_2\, inc + \beta_2\, inc^2 + \beta_4\, age + \beta_4\, age^2 + \varepsilon$$

The goal is to test whether there is a tradeoff between participating in a 401(k) plan and having an individual retirement account (IRA). Therefore, we want to estimate β_1.

a) Estimate the equation by OLS and discuss the estimated effect of `p401k`.

b) For the purposes of estimating the ceteris paribus tradeoff between participation in two different types of retirement savings plans, what might be a problem with ordinary least squares?

c) The variable `e401k` is a binary variable equal to one if a worker is eligible to participate in a 401(k) plan. Explain what is required for `e401k` to be a valid IV for `p401k`. Do these assumptions seem reasonable?

d) Estimate the reduced form for `p401k` and verify that `e401k` has significant partial correlation with `p401k`. Since the reduced form is also a linear probability model, use a heteroskedasticity-robust standard error.

e) Now, estimate the structural equation by IV and compare the estimate of β_1 with the OLS estimate. Again, you should obtain heteroskedasticity-robust standard errors.

[4] The data in `FERTIL2` include, for women in Botswana during 1988, information on number of children, years of education, age, and religious and economic status variables.

a) Estimate the model

$$children = \beta_0 + \beta_1\, educ + \beta_2\, age + \beta_3\, age^2 + \varepsilon$$

by OLS and interpret the estimates. In particular, holding `age` fixed, what is the estimated effect of another year of education on fertility? If 100 women receive another year of education, how many fewer children are they expected to have?

b) The variable `frsthalf` is a dummy variable equal to one if the woman was born during the first six months of the year. Assuming that `frsthalf` is uncorrelated with the error term from part (a), show that `frsthalf` is a reasonable IV candidate for `educ`. (You need to do a regression.)

c) Estimate the model from part (a) by using `frsthalf` as an IV for `educ`. Compare the estimated effect of education with the OLS estimate from part (a).

d) Add the binary variables `electric`, `tv`, and `bicycle` to the model and assume these are exogenous. Estimate the equation by OLS and 2SLS and compare the estimated coefficients on `educ`. Interpret the coefficient on `tv` and explain why television ownership has a negative effect on fertility.

[5] During the 1880s, a cartel known as the Joint Executive Committee (JEC) controlled the rail transport of grain from the Midwest to eastern cities in the United States. The cartel preceded the Sherman Antitrust Act of 1890, and it legally operated to increase the price of grain above what would have been the competitive price. From time to time, cheating by members of the cartel brought about a temporary collapse of the collusive price- setting agreement. In this exercise, you will use variations in supply associated with the cartels collapses to estimate the elasticity of demand for rail transport of grain. The data file `JEC` contains weekly observations on the rail shipping price and other factors from 1880 to 1886. A detailed description of the data is contained in `JEC_DataDescription`.

Suppose that the demand curve for rail transport of grain is specified as

$$\ln(Q_i) = \beta_0 + \beta_1 \ln P_i + \beta_2 Ice_i + \sum_{j=1}^{12} \beta_{2+j} Seas_{j,i} + \varepsilon_i$$

where Q_i is the total tonnage of grain shipped in week i, P_i is the price of shipping a ton of grain by rail, Ice_i is a binary variable that is equal to 1 if the Great Lakes are not navigable because of ice, and $Seas_j$ is a binary variable that captures seasonal variation in demand. Ice is included because grain could also be transported by ship when the Great Lakes were navigable.

a) Estimate the demand equation by OLS. What is the estimated value of the demand elasticity and its standard error?

b) Explain why the interaction of supply and demand could make the OLS estimator of the elasticity biased.

c) Consider using the variable `cartel` as instrumental variable for `ln(P)`. Use economic reasoning to argue whether `cartel` plausibly satisfies the two conditions for a valid instrument.

d) Estimate the first-stage regression. Is `cartel` a weak instrument?

e) Estimate the demand equation by instrumental variable regression. What is the estimated demand elasticity and its standard error?

f) Does the evidence suggest that the cartel was charging the profit-maximizing monopoly price? Explain. (Hint: What should a monopolist do if the price elasticity is less than 1?)