Economics 522A, Spring 2007

Lecture Note 22: IV Part III – Practical Issues; Simultaneous Equations

1 Practical Issue: Exogenous Variation

As we have seen, in order to interpret OLS or IV estimates causally, we need that $E[x_i u_i] = 0$ or $E[z_i u_i] = 0$; in other words, we need to be sure that some variables are orthogonal to any unobserved factors that may influence the outcome.

This is a difficult requirement to meet in many settings. One of the most important things to understand is the following:

Just because you have data, does not mean you can answer a causal question!!!

When can we be guaranteed of orthogonality? If the "treatment" of interest is randomly assigned, then it must be independent of any unobserved factors. For example, if we could somehow take a group of individuals and randomly assign half to go to college, and half to stop at high school (and make sure they *comply* with their assignment), then by construction, the distribution of pretreatment "motivation" and "ability" would be the same in both the college and high school groups.

So a randomized experiment represents an ideal situation, where orthogonality holds by construction. This is why medical therapies must be evaluated by randomized clinical trials in the U.S., and why many economists have turned to running experiments to try to estimate causal effects. There are many different types of social experiments: small-scale "laboratory" experiments run in classrooms or special computer labs, where subjects are asked to play games against each other or participant in a simulated market; and experiments "in the field" where different interventions or social policies are randomly assigned to individuals.

If a true randomized experiment cannot be run, one might look for "natural" sources of randomness to use as instrumental variables.[1] However, one must be careful to make sure that such exogenous variation constitutes a valid instrumental variable — it must affect the treatment of interest, and it must be excluded from the main equation, so that any influence of the IV is mediated through the treatment of interest. The exclusion restriction, that the IV does not directly affect the outcome (only affecting outcome through the treatment of interest) is not generally testable and must be argued on a case-by-case basis. Finding good instruments is very difficult!

Another possibility is to try to locate a rich set of covariates that will adequately proxy for the omitted factors. Again, one must argue on a case-by-case basis that the orthogonality condition is valid — it is not guaranteed just because you have a lot of regressor variables. In settings without some "exogenous source of variation," it can be difficult to persuade others that your regression results should be interpreted causally.

[1] In addition, if a randomized experiment is run, but individuals do not comply with their assigned treatment, sometimes the assignment can be viewed as an instrumental variable.

2 Practical Issue: Weak Instruments

In addition to the orthogonality requirement that $E[z_i u_i] = 0$, and the exclusion restriction that z_i affects the outcome y_i only through x_i, we also need that $E[z_i x_i']$ has full rank. This essentially means that z_i is sufficiently strongly correlated with the endogenous regressors in x_i.

If the correlation is instead very small, we say that the instruments are "weak." Intuitively, the instruments do not induce enough variation in the treatments of interest to be able to learn much about the effect of the treatment. This leads to two problems. First, the asymptotic variance of the IV estimator will become very large, because the inverse of $E[z_i x_i']$ blows up. Second, the asymptotic normality approximation becomes a poor approximation to the finite-sample distribution of the IV estimator.

Since the correlation between z_i and x_i can be estimated, one can examine directly whether the instruments are strongly correlated with the regressors. For example, in 2SLS, one should report the results from the first-stage regression of x_i on z_i. It is common to report the F-statistic, for the test that all of the slope coefficients in a regression of the endogenous components in x_i on z_i are equal to 0. A low value for the F-statistic indicates that the correlation between the endogenous variable and the instruments is weak, which in turn would indicate that the 2SLS estimates may be very unreliable.

3 Practical Issue: Many Instruments

Suppose that we have many instrumental variables, so that $m = dim(z_i)$ is very large compared to $k = dim(x_i)$.

Recall that in 2SLS, the first stage involves regressing x_i on z_i and taking the fitted values as the regressors for the second stage. If you have many instruments, you will be able to fit x_i very well (recall that adding regressors to an OLS regression always increases the R^2). At the extreme, with enough instrumental variables, you will get a perfect fit to x_i, so that $\hat{x}_i = x_i$. But then the 2SLS estimate will be equal to the OLS estimate.

This illustrates a bias problem with 2SLS. When there are many instruments, 2SLS will be biased towards the OLS estimates, which by assumption are biased for the causal parameter γ. (That is why we are doing IV in the first place!) The problem turns out to be especially pronounced when the instruments are weakly correlated with the endogenous regressors. Again, reporting the F-statistic from the first-stage regression can be useful as a diagnostic tool; one would want to be even more cautious when there are many instruments and they collectively appear to be weak.

The weak instruments and the many weak instruments problems are areas of active research, and researchers have recently developed tests and confidence intervals that are more robust to these problems.

4 A Model of Supply and Demand

Here is another, slightly different application of the instrumental variables approach. We consider a *simultaneous equation* model of supply and demand.

Suppose that we are interested in estimating a demand function. We have observations on prices and quantities p_i, q_i from a collection of markets $i = 1, \ldots, n$. (Or, we might have observations for price and quantity from the same market, but in different time periods.)

Each market i is assumed to follow a standard supply-demand model, with demand function $Q_i^d(p)$ and supply function $Q_i^s(p)$. So $Q_i^d(p)$ gives the quantity demanded by consumers in market i at price p, and $Q_i^s(p)$ gives the quantity supplied by producers in market i at price p.[2]

We assume that quantities and prices are determined by equating supply and demand: q_i^* and p_i^* satisfy

$$q_i^* = Q_i^d(p_i^*) = Q_i^s(p_i^*).$$

To keep things simple, let us put some more structure on the demand and supply equations. Suppose we can write

$$\begin{aligned} Q_i^d(p) &= \alpha_1 + \alpha_2 p + u_i^d \\ Q_i^s(p) &= \beta_1 + \beta_2 p + u_i^s. \end{aligned}$$

This says that all the markets $i = 1, \ldots, n$ have demand functions with the same slope α_2. (Presumably, $\alpha_2 < 0$.) Likewise, all markets have supply functions with the same slope β_2. There is some variation across markets, however, due to u_i^d and u_i^s. There are "shocks" which shift supply up or down, and shocks which shift demand up or down, but we are restricting these shocks to cause parallel shifts in the demand and supply functions.

We will assume that

$$E[u_i^d] = E[u_i^s] = 0,$$

and, for simplicity, we assume the supply and demand shocks are uncorrelated: $E[u_i^s u_i^d] = 0$.

First, let's solve for the equilibrium price in market i: we set

$$\alpha_1 + \alpha_2 p + u_i^d = \beta_1 + \beta_2 p + u_i^s.$$

The equilibrium price is:

$$p_i^* = \frac{\beta_1 - \alpha_1}{\alpha_2 - \beta_2} + \frac{u_i^s - u_i^d}{\alpha_2 - \beta_2}.$$

Notice that price depends on both u_i^s and u_i^d.

Suppose we try to estimate the demand equation by regressing q_i^* on p_i^*. The demand equation implies

$$q_i^* = \alpha_1 + \alpha_2 p_i^* + u_i^d.$$

[2] Often we work with the logarithms of price and quantity, but I will just refer to these as "price" and "quantity" to keep the terminology simple.

But a regression will not consistently estimate α_2, because p_i^* is correlated with u_i^d. The equilibrium price is *endogenous*, because it is determined "inside the system of equations" and is therefore correlated with u_i^d and u_i^s.

5 Supply Shifters as Instrumental Variables

In the simple model of the previous section, there is **no way** to consistently estimate α_2, the demand elasticity parameter. We say that α_2 is **not identified.** The problem is that demand shocks affect price, and also affect quantity. If we could somehow isolate shifts in supply, then it is plausible that we could estimate the demand curve.

Suppose we have an additional variable w_i, which shifts supply but not demand. For example, Graddy (1995) collected data on a market for whiting (a kind of fish) at the Fulton Fish Market in New York City. For each of 111 days, she collected observations on the price and quantity of whiting sold, along with some other variables. (So here, i indexes the market on different days.) A key additional variable she collected was a measure of offshore weather. The idea is that variation in offshore weather affects the supply of fish, but not demand.

Extending the model of the previous section, we could write the supply and demand curves as:

$$Q_i^d(p) = \alpha_1 + \alpha_2 p + u_i^d;$$
$$Q_i^s(p) = \beta_1 + \beta_2 p + \beta_3 w_i + u_i^s.$$

As before, we assume that $E[u_i^d] = E[u_i^s] = E[u_i^d u_i^s] = 0$, and now we assume that w_i satisfies

$$E[w_i u_i^d] = 0, \quad E[w_i u_i^s] = 0.$$

At this point, we can consider the demand equation

$$q_i^* = \alpha_1 + \alpha_2 p_i^* + u_i^d.$$

Let $x_i = (1, p_i^*)'$, and let the instrument vector be $z_i = (1, w_i)'$. By assumption,

$$E[z_i u_i^d] = 0.$$

So z_i is a valid instrumental variable, and we can estimate α_1 and α_2 by linear IV.

There is another way to get to the same place. It's useful to work out explicitly the forms for equilibrium price and quantity and work from there. The equilibrium price is

$$p_i^* = \frac{\beta_1 - \alpha_1}{\alpha_2 - \beta_2} + \frac{\beta_3}{\alpha_2 - \beta_2} w_i + \frac{u_i^s - u_i^d}{\alpha_2 - \beta_2}. \tag{1}$$

Substituting this in to the demand equation we get the equilibrium quantity as:

$$\begin{aligned} q_i^* &= \alpha_1 + \alpha_2 p_i + u_i^d & (2) \\ &= \alpha_1 + \frac{\alpha_2(\beta_1 - \alpha_1)}{\alpha_2 - \beta_2} + \frac{\alpha_2 \beta_3}{\alpha_2 - \beta_2} w_i + \frac{\alpha_2}{\alpha_2 - \beta_2}(u_i^s - u_i^d) + u_i^d. & (3) \end{aligned}$$

We can write (1) and (3) simply as:

$$p_i^* = \pi_{11} + \pi_{12} w_i + v_{i1} \quad (4)$$
$$q_i^* = \pi_{21} + \pi_{22} w_i + v_{i2} \quad (5)$$

where

$$v_{i1} = \frac{u_i^s - u_i^d}{\alpha_2 - \beta_2},$$

and

$$v_{i2} = \frac{\alpha_2}{\alpha_2 - \beta_2}(u_i^s - u_i^d) + u_i^d$$

Since both v_{i1} and v_{i2} are linear combinations of u_i^d and u_i^s, it follows that $E(z_i v_{i1}) = 0$ and $E(z_i v_{i2}) = 0$. This latter representation is called the **reduced form**, because *only endogenous variables appear on the left hand side, and only exogenous variables appear on the right hand side.*

The reduced form is useful, because by construction, the reduced form parameters $(\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$ can be estimated simply by regressing p_i^* on z_i, and regressing q_i^* on z_i.

Let $(\hat{\pi}_{11}, \hat{\pi}_{12}, \hat{\pi}_{21}, \hat{\pi}_{22})$ denote the OLS estimates. Then by standard arguments,

$$\hat{\pi}_{11} \xrightarrow{p} \pi_{11} = \frac{\beta_1 - \alpha_1}{\alpha_2 - \beta_2},$$
$$\hat{\pi}_{12} \xrightarrow{p} \pi_{12} = \frac{\beta_3}{\alpha_2 - \beta_2},$$
$$\hat{\pi}_{21} \xrightarrow{p} \pi_{21} = \alpha_1 + \frac{\alpha_2(\beta_1 - \alpha_1)}{\alpha_2 - \beta_2},$$
$$\hat{\pi}_{22} \xrightarrow{p} \pi_{22} = \frac{\alpha_2 \beta_3}{\alpha_2 - \beta_2}.$$

Notice that

$$\frac{\pi_{22}}{\pi_{12}} = \alpha_2.$$

So we can estimate α_2 by the ratio

$$\frac{\hat{\pi}_{22}}{\hat{\pi}_{12}} \xrightarrow{p} \frac{\pi_{22}}{\pi_{12}} = \alpha_2.$$

So α_2 is identified (consistently estimable). Likewise, we can write

$$\begin{aligned} \pi_{21} &= \alpha_1 + \frac{\alpha_2(\beta_1 - \alpha_1)}{\alpha_2 - \beta_2} \\ &= \alpha_1 + \alpha_2 \pi_{11} \\ &= \alpha_1 + \frac{\pi_{22}}{\pi_{12}} \pi_{11}, \end{aligned}$$

giving us the relation

$$\alpha_1 = \pi_{21} - \frac{\pi_{22}}{\pi_{12}} \pi_{11}.$$

So an estimate of α_1 can be formed as

$$\hat{\alpha}_1 = \hat{\pi}_{21} - \frac{\hat{\pi}_{22}}{\hat{\pi}_{12}}\hat{\pi}_{11}.$$

We can rewrite $\hat{\alpha}_2$ to see the connection to the linear IV estimator. By standard OLS results, we can write

$$\hat{\pi}_{22} = \frac{\sum_i (q_i^* - \bar{q})(w_i - \bar{w})}{\sum_i (w_i - \bar{w})^2},$$

and

$$\hat{\pi}_{12} = \frac{\sum_i (p_i^* - \bar{p})(w_i - \bar{w})}{\sum_i (w_i - \bar{w})^2}.$$

Therefore,

$$\hat{\alpha}_2 = \frac{\hat{\pi}_{22}}{\hat{\pi}_{12}} = \frac{\sum_i (q_i^* - \bar{q})(w_i - \bar{w})}{\sum_i (p_i^* - \bar{p})(w_i - \bar{w})}.$$

This is equivalent to the IV estimate of α_2 in the equation

$$q_i^* = \alpha_1 + \alpha_2 p_i^* + u_i^d,$$

using w_i as an *instrument* for p_i^*.

6 References

Graddy, K., (1995), "Testing for Imperfect Competition in the Fulton Fish Market," *RAND Journal of Economics*.