Additional Notes on 2SLS and Simultaneous Equations

14.32, Spring 2007

Recitation 5/11/07

Consider the generic model

$$\begin{aligned} q_d &= \alpha_0 + \alpha_1 p + \alpha_2 x + \varepsilon \\ q_s &= \beta_0 + \beta_1 p + \beta_2 z + v \\ q_s &= q_d \end{aligned}$$

In this system, both the supply and demand equations are exactly identified. We can estimate α_1 and β_1 using indirect least squares or two-stage least squares.

To obtain the 2SLS estimate of α_1 :

1. Regress p on x *and* z

2. Collect the predicted values $\hat{p}$

3. Regress $q_d = \alpha_0 + \alpha_1 \hat{p} + \alpha_2 x + \varepsilon + (p - \hat{p})$

The important thing to notice is that in the first stage of 2SLS, we regress p on *all exogenous variables*, not just the instrument x. The only explanation for this is mechanical, and it's quite tedious to show why. But here's a sketch of how the mechanics work in our simple example.

Going back to the derivation of instrumental variables, the IV estimator for α_1 is

$$\alpha_1 = \frac{Cov(\tilde{q}, \tilde{z})}{Cov(\tilde{p}, \tilde{z})}$$

Note that we have to partial out the effect of x in the demand equation before we can derive the IV estimator.

In order for the 2SLS estimator to equal this IV estimator, we must also parital out the effects of x from all variables.

$$
\begin{aligned}
\hat{\alpha}_{1,2SLS} &= \frac{Cov(\tilde{q}, \widehat{\tilde{p}})}{Var(\widehat{\tilde{p}})} \\
&= \frac{Cov(\tilde{q}, \frac{Cov(\tilde{p},\tilde{z})}{var(\tilde{z})}\tilde{z} + \kappa)}{Var(\frac{Cov(\tilde{p},\tilde{z})}{var(\tilde{z})}\tilde{z} + \kappa)} \\
&= \frac{\frac{Cov(\tilde{p},\tilde{z})}{var(\tilde{z})} Cov(\tilde{q}, \tilde{z})}{\left(\frac{Cov(\tilde{p},\tilde{z})}{var(\tilde{z})} z\right)^2 Var(\tilde{z})} \\
&= \frac{Cov(\tilde{q}, \tilde{z})}{Cov(\tilde{q}, \tilde{z})} = \alpha_{1,IV}
\end{aligned}
$$

Even though x is not technically an instrument, we still must include it in the first stage. To make matters more confusing, we have to define x as an instrument PROC SYSLIN:

```
proc syslin data=one 2sls;
  endogenous p;
  instruments x z;
  model q=p+x;
```

Including only z as an "instrument" will give you the wrong answer! As a rule of thumb, don’t forget to include all exogenous variables in your first stage. The instrument is the *excluded* variable in the 2nd stage.