Economics 522A, Spring 2007

Lecture Note 20: Linear Instrumental Variables Part II

1 Linear IV: Overidentified Case

Suppose that $m > k$ (so the dimension of z_i is greater than the dimension of x_i). Then, the equation

$$\frac{1}{n}\sum_{i=1}^{n} z_i(y_i - x_i'\hat{\gamma}) = 0$$

will not generally have a solution. We say that it is "overidentified."

If we cannot set the sample analog equation exactly equal to zero, might could instead try to set it "close" to zero. We will do this by solving the minimization problem

$$\min_{\gamma} \left(\frac{1}{n}\sum z_i(y_i - x_i'\gamma)\right)' A_n^{-1} \left(\frac{1}{n}\sum z_i(y_i - x_i'\gamma)\right),$$

where A_n^{-1} is a $m \times m$ symmetric and positive definite matrix. (The "n" subscript indicates that the weight matrix could be a function of the data.) Notice that we are minimizing a general "distance" between the vector of moment conditions $\frac{1}{n}\sum_i z_i(y_i - x_i'\delta)$ and 0.[1]

This problem has a fairly simple explicit solution (no need for numerical optimization techniques!). The matrix first order conditions associated with this minimization problem are

$$-2X'ZA_n^{-1}Z'y + 2X'ZA_n^{-1}Z'X\gamma = 0$$

which leads to the following equation:

$$X'ZA_n^{-1}Z'(y - X\gamma) = 0. \tag{1}$$

Define

$$W_n' = X'ZA_n^{-1}Z'.$$

Then equation (1) can be written as

$$W_n'(y - X\gamma) = 0$$

which leads to the solution

$$\hat{\gamma} = (W_n'X)^{-1}W_n'y.$$

So W_n acts like a matrix of instruments in the just-identified case.

[1]Note that if we are in the just-identified case, where $m = k$, then the solution will be the $\hat{\gamma}$ that sets the moment equation equal to 0, regardless of the choice of A_n.

2 Two-Stage Least Squares

Instead of minimizing the distance of the moment equation to 0, we could Suppose that $E[u_i|z_i] = 0$. (This is stronger than the orthogonality condition $E[z_i u_i] = 0$.) Then we could write

$$\begin{aligned} E[y_i|z_i] &= E[x_i'\gamma + u_i|z_i] \\ &= E[x_i|z_i]'\gamma + E[u_i|z_i] \\ &= E[x_i|z_i]'\gamma. \end{aligned}$$

So if we knew $E[x_i|z_i]$, we could regress y_i on $E[x_i|z_i]$ to get estimates of γ.

We don't know $E[x_i|z_i]$, but since we observe both vectors, we could use a preliminary regression to estimate it, and use the estimated values as the regressors.

This has a convenient representation in matrix form. Let X be the $n \times k$ matrix with typical row x_i', and let Z be the $n \times m$ matrix with typical row z_i'. We want to regress each column of Z on the entire matrix X. Rather than doing this one-by-one, we can form:

$$\hat{X} = Z(Z'Z)^{-1}Z'X = P_Z X.$$

Here, P_Z is the projection matrix which projects the columns of X onto the column space spanned by Z. So, if x_i and z_i have some elements in common (for example, if the variable age is in both x_i and z_i, then the projection will just return the observations on age).

Then, we can estimate γ by regressing y on $\hat{X}$:

$$\hat{\gamma}_{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y.$$

This is called the "two-stage least squares" estimator, or 2SLS (or TSLS).

2SLS is actually a special case of our general linear IV estimator. Suppose we set

$$A_n = (Z'Z).$$

Then

$$W_n = ZA_n^{-1}Z'X = Z(Z'Z)^{-1}Z'X = P_Z X.$$

Using the symmetry and idempotence of the projection matrix P_Z, we can write

$$\begin{aligned} \hat{\gamma} &= (W_n'X)^{-1}W_n'y \\ &= (X'P_Z'X)^{-1}X'P_Z'y \\ &= (X'P_Z'P_Z X)^{-1}X'P_Z'y \\ &= (\hat{X}'\hat{X})^{-1}\hat{X}'y. \end{aligned}$$

3 Asymptotic Theory

Asymptotic properties of the IV estimator in the overidentified case are similar to the just-identified case. It's easiest to write the estimator as

$$\begin{aligned}\hat{\gamma} &= (W_n'X)^{-1}W_n'y \\ &= (X'ZA_n^{-1}Z'X)^{-1}X'ZA_n^{-1}Z'y \\ &= \gamma + (X'ZA_n^{-1}Z'X)^{-1}X'ZA_n^{-1}Z'u.\end{aligned}$$

Assume that $A^{-1} = \text{plim}(A_n^{-1})$, is a finite positive definite matrix, and $Q = \text{plim}\left(\frac{Z'X}{n}\right)$ is a $m \times k$ matrix with rank k. As before, $\frac{Z'u}{n} \xrightarrow{\text{p}} 0$, so $\hat{\gamma} \xrightarrow{\text{p}} \gamma$. Likewise, it can be shown that

$$\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{\text{d}} N(0, \Delta\Omega\Delta'),$$

where

$$\Omega = E(u_i^2 z_i z_i'), \qquad \Delta = (Q'A^{-1}Q)^{-1}Q'A^{-1}$$

The variance can be shown to be minimized by setting $A = \Omega$, or any constant multiple of Ω. In practice, this means that it is optimal to choose A_n to be a consistent estimator of Ω. This leads to an asymptotic variance of

$$(Q'\Omega^{-1}Q)^{-1}.$$

In the homoskedastic case, where $\Omega = \sigma^2 E(z_i z_i')$, it is enough to have $A = E(z_i z_i')$, leading to the optimal asymptotic variance of

$$\sigma^2(Q'E(z_i z_i')^{-1}Q)^{-1}.$$

We can operationalize this by choosing the weight matrix to be

$$A_n^{-1} = \left(\frac{Z'Z}{n}\right)^{-1}$$

since $A_n \xrightarrow{\text{p}} E(z_i z_i')$. But this is just 2SLS (the $1/n$ term does not affect the estimator, because it cancels out in the expression for $\hat{\gamma}$.). This shows that 2SLS is optimal in the class of IV estimators under homoskedasticity. Under general heteroskedascity this optimality no longer holds.