Instrumental Variables

Econ 671

Purdue University

Instrumental Variables

The failure of OLS methods stems from a violation of

-

or the weaker condition

-

If this condition were to hold, then

-

In practice, of course, the population expectation is unknown and thus we replace the expectation with its sample counterpart:

-

which leads to

-

Instrumental Variables

Once we have viewed this problem as a failure of a moment condition, it suggests a possible course for how we might approach solving it.

To this end, suppose that X is $n \times k$ and that there is another set of k variables, stacked into an $n \times k$ matrix Z, that satisfy

-

Again, the sample analog of this condition will yield an estimator of the form:

-

known as the *instrumental variables* or *IV* estimator.

Instrumental Variables

Under certain conditions, the IV estimator will be *consistent*. Specifically, we require:

1. The instruments Z are correlated with X in the sense that
   -
2. The instruments are uncorrelated with ϵ in the sense that
   -

Instrumental Variables

Under these conditions it is almost immediate that the IV estimator will be consistent for β in the linear regression model. To see this, note:

$$\begin{aligned} \hat{\beta}_{IV} &= (Z'X)^{-1}Z'y \\ &= \beta + (Z'X)^{-1}Z'\epsilon \end{aligned}$$

so that

-

The estimator is *not typically unbiased*, however, since:

-

Instrumental Variables

Similar to our results for the OLS estimator, we can also obtain an asymptotic distribution for the IV estimator:

$$y = X\beta + \epsilon, \quad E(\epsilon|X) \neq 0, \quad E(\epsilon\epsilon'|X, Z) = \sigma^2 I_n$$

$$\begin{aligned} \hat{\beta}_{IV} &= (Z'X)^{-1} Z'y \\ &= \beta + (Z'X)^{-1} Z'\epsilon \end{aligned}$$

so that

$$\sqrt{n}(\hat{\beta}_{IV} - \beta) = \left(\frac{Z'X}{n}\right)^{-1} \left(\frac{1}{\sqrt{n}} Z'\epsilon\right).$$

Again, a CLT can be applied to the second term to get

-

where

-

Instrumental Variables

Our assumption that Z is correlated with X also implies:

$$n^{-1}Z'X \xrightarrow{p} Q_{zx} \neq 0.$$

Thus,

-

Replacing Q_{zz} with $\frac{1}{n}Z'Z$ and Q_{zx} with $\frac{1}{n}Z'X$, we obtain

$$\widehat{\text{AsyVar}(\hat{\beta}_{IV}}|X,Z) = \hat{\sigma}^2 \frac{1}{n}\left(\frac{1}{n}Z'X\right)^{-1}\left(\frac{1}{n}Z'Z\right)\left[\left(\frac{1}{n}Z'X\right)^{-1}\right]'$$

or

$$\widehat{\text{AsyVar}(\hat{\beta}_{IV}}|X,Z) = \hat{\sigma}^2(Z'X)^{-1}Z'Z(X'Z)^{-1}.$$

Instrumental Variables

A consistent estimator of the variance parameter is obtained as:

$$\hat{\sigma}^2 = \frac{1}{n-k} \sum_{i=1}^{n} (y_i - x_i \hat{\beta}_{IV})^2.$$

Corrections for *heteroscedasticity* can also be made an a straightforward way.

- In the foregoing discussion we introduced the IV estimator. In that discussion, we supposed there were just as many instruments as variables that required instrumenting. This is called the *just identified* case.

- Here, we relax this requirement and show what to do when we have more instruments than are necessary (This is called the *overidentified* case). If we have too few instruments, then we are in trouble!

Too many instruments?

Consider the regression model:

$$y = X\beta + \epsilon, \quad E(X'\epsilon) \neq 0, \quad E(\epsilon\epsilon'|X) = \sigma^2 I_n, \quad E(Z'\epsilon) = 0.$$

In the above, X is $n \times k$ and Z is $n \times j, j > k$.

It is clear that direct application of the formula:

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'y$$

is not appropriate as the matrix product $Z'X$ is no longer square.

Any ideas what we should do, since we have *too much* information?

Too many instruments?

In some way, we must select k instruments among the set of j possible instruments. We restrict ourselves to linear combinations of the elements of Z. To this end, we can choose a $n \times k$ set of instruments W, by forming

-

and R is a $j \times k$ matrix with rank k. For example, setting:

$$R = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} I_k \\ 0_{j-k,k} \end{bmatrix}.$$

chooses only the first k elements of Z as the k elements of W.

Two Stage Least Squares

Based upon results in our previous lecture, the approximate (dropping expectations) asymptotic covariance matrix for this IV estimator is

$$\begin{aligned}\hat{\beta}_{IV} &= (W'X)^{-1}W'y. \\ \text{AsyVar}(\hat{\beta}_{IV}|W,X) &= \sigma^2(W'X)^{-1}W'W(X'W)^{-1}.\end{aligned}$$

Now consider a particular, (and at this point completely unjustified), choice of R.

Define the *two stage least squares* (2SLS) estimator of β as:

-

Note that this emerges upon choosing

$$W = ZR, \quad R = (Z'Z)^{-1}Z'X.$$

Two Stage Least Squares

Under this choice, the conditional asymptotic variance-covariance matrix of the 2SLS estimator is:

$$\text{Var}(\hat{\beta}_{2SLS}) = \sigma^2 \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1} \left[X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X\right] \left[X'Z(Z'Z)^{-1}Z'X\right]$$

which reduces to

$$\text{AsyVar}(\hat{\beta}_{2SLS}|Z, X) = \sigma^2 \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1}.$$

Two Stage Least Squares

Let

$$X^* \equiv Z(Z'Z)^{-1}Z'X.$$

With this definition, we can write

$$\text{AsyVar}(\hat{\beta}_{2SLS}|X, Z) = \sigma^2(X^{*\prime}X^*)^{-1}$$

since

$$X^{*\prime}X^* = X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X = X'Z(Z'Z)^{-1}Z'X.$$

Also write:

-

where

$$Z'\eta = Z'X - Z'Z(Z'Z)^{-1}Z'X = 0.$$

Two Stage Least Squares

Furthermore,

-

from this last result.

Where we are going with all of this may seem unclear. To this point, we have introduced a general (linear) IV estimator, derived its conditional asymptotic covariance matrix, and also introduced a *specific* IV estimator, known as two stage least squares, and derived its conditional asymptotic variance.

We claim that $\hat{\beta}_{2SLS}$ is *asymptotically efficient* among this class. That is, choosing

$$R = (Z'Z)^{-1}Z'X$$

is the best choice, as it will produce a linear estimator among the available instruments that has smallest (asymptotic) variance.

Two Stage Least Squares

To prove this, we will show that the difference between the asymptotic variances of $\hat{\beta}_{IV}$ and $\hat{\beta}_{2SLS}$ is a positive semidefinite matrix.

In order to do this, we first borrow a result from linear algebra which states:

$$A^{-1} - B^{-1} \text{ is p.s.d. } \Rightarrow B - A \text{ is p.s.d.}$$

Thus, we can work with the difference of the inverse covariance matrices (in reverse order) to establish the result.

Two Stage Least Squares

In the above, let:

$$B = \sigma^2 (W'X)^{-1} W'W(X'W)^{-1} \Rightarrow B^{-1} = \sigma^{-2} X'W(W'W)^{-1} W'X$$

and

$$A = \sigma^2 (X^{*\prime} X^*)^{-1} \Rightarrow A^{-1} = \sigma^{-2} X^{*\prime} X^*.$$

The theorem on the previous page implies it is sufficient to show

-

Two Stage Least Squares

To demonstrate this result, first note:

-

from our earlier derivation. Thus,

-

We can substitute this into the difference above to obtain:

-

Two Stage Least Squares

This last result shows that the difference between the inverse matrices is positive semidefinite, whence $B - A$ is also positive semidefinite.

Therefore, we have established that the 2SLS estimator is the best one, as it determines the "optimal" weighting of the instruments. Importantly, note that *all of the instruments are used* and ignoring valid instruments results in a loss of (asymptotic) efficiency. (A better proof, using expectations of the involved quantities follows similarly).

In what follows we investigate what the 2SLS estimator is doing, and why it is given this name.

Two Stage Least Squares

Recall the definition of X^* and the 2SLS estimator:

$$X^* = Z(Z'Z)^{-1}Z'X \qquad \hat{\beta}_{2SLS} = \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1} X'Z(Z'Z)^{-1}Z'y.$$

Note:

1. $(Z'Z)^{-1}Z'X$ denotes the matrix of regression coefficients obtained from regressing each column of X (i.e., each X variable) on Z.
2. $X^* = Z(Z'Z)^{-1}Z'X$ denotes the matrix of fitted values from these regressions.
3. Since
$$\hat{\beta}_{2SLS} = (X^{*\prime}X^*)^{-1}X^{*\prime}y,$$
the following interpretation can be given to this estimator:

Two Stage Least Squares

Step 1:

For each endogenous variable in the model, say x_l, obtain a vector of fitted values, $\hat{x}_l$ from a regression of x_l on a constant, all elements in Z and all *exogenous* elements of X.

Step 2: Replace the endogenous variables with their fitted values, and then regress y on a constant, the exogenous X variables and the collection of fitted values (which replace the endogenous variables in the model).

The estimates produced from this regression are the 2SLS estimates which are consistent and asymptotically efficient under the given assumptions. (Also note that the exogenous variables serve as IVs for themselves).

Two Stage Least Squares

Finally, note that the asymptotic covariance matrix for the 2SLS estimator can be calculated as:

$$\text{AsyVar}(\hat{\beta}_{2SLS}|X, Z) = \hat{\sigma}^2(X^{*\prime}X^*)^{-1}$$

with

$$\hat{\sigma^2} = \frac{1}{n-k}(y - X\hat{\beta}_{2SLS})'(y - X\hat{\beta}_{2SLS}),$$

which, importantly, uses the *observed* Xs and does not replace them with their fitted values.

IV With Measurement Error

Consider, again, the measurement error problem:

$$y_i = x_i^* \beta + u_i,$$

$$x_i = x_i^* + \omega_i.$$

We run a regression using data that we have:

$$y_i = x_i \beta + \eta_i, \quad \eta_i = u_i - \omega_i \beta.$$

Suppose $\exists Z$ (Z is $n \times 1$ and X is $n \times 1$) such that:

1. Z is correlated with X so that $Z'X/n \rightarrow Q \neq 0$,
2. Z is uncorrelated with the regression error u and the measurement error ω in the sense that $Z'u/n \xrightarrow{p} 0$ and $Z'\omega/n \xrightarrow{p} 0$

IV With Measurement Error

In this case it is clear that the IV estimator will be consistent since:

$$\hat{\beta}_{IV} = \beta + \left(\frac{Z'X}{n} \right)^{-1} \left(\frac{Z'(u - \omega\beta)}{n} \right),$$

and the last term clearly goes to zero in probability. Consider the following example:

-

where *Ed* denotes education (assumed exogenous) and A is an ability (test score) measure. further, R_i potentially includes lots of other things (which we assume are exogenous).

A common criticism of A is that it is an error-ridden measure of true ability; one observed test score is not an accurate measure of actual cognitive skills. Consequently, OLS estimation of the above produces biased and inconsistent estimates.

IV With Measurement Error

So, what can we do to fix this problem?

One possibility is to come up with an *instrument*. That is, we need to find a variable that is:

1. Correlated with Ability.
2. Uncorrelated with u (implying that it is uncorrelated with the error of measurement in A_i as well as the "true" regression disturbance.)

Some possibilities include:

1. Parental Education / Parental Income
2. Sibling's Test Score

IV With Measurement Error

Regardless of what you choose, suppose you settle on an $n \times 1$ vector, say F to use as an IV for A. In this case, you would construct:

$$Z = [1 \quad Ed \quad F \quad R], \qquad X = [1 \quad Ed \quad A \quad R]$$

and compute

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'y.$$

If F was, say $n \times l$, $l > 1$, we would first:

1. Regress A on 1, Ed, R and Z and get the fitted values, say $\hat{A}$.
2. Regress y on 1, Ed, $\hat{A}$ and R to obtain the 2SLS estimator. Note that this estimator is asymptotically more efficient than selecting any one of the elements of Z.

IV With Simultaneity

Consider, again, the SEM discussed before:

$$\begin{aligned} y_1 &= \alpha_1 y_2 + X_1\beta_1 + \epsilon_1 \\ y_2 &= \alpha_2 y_1 + X_2\beta_2 + \epsilon_2 \end{aligned}$$

Consider estimation of the first equation. The variable y_2 needs to be instrumented, since it is not uncorrelated with ϵ_1.

Any suggestions?

If x_2 is a scalar, it can be used as an IV for y_2. (If X_2 has several elements, 2SLS can be applied). This works since x_2 satisfies all the needed conditions as it is correlated with y_2 (see reduced form) and also uncorrelated with ϵ_1, by assumption.
Therefore, we can form:

-

and calculate

-

IV With Simultaneity

A TV show that I used to like, *the King of Queens*, illustrated a type of simultaneity problem:

Doug: "I'm always eating because you're always yelling at me!"

Carrie: "I'm always yelling at you because you're always eating!"

$$\begin{aligned} Eating &= \alpha_0 + \alpha_1 Yelling + \beta_1 X_1 + \epsilon_1 \\ Yelling &= \alpha_2 + \alpha_2 Eating + \beta_2 X_2 + \epsilon_2 \end{aligned}$$

Any thoughts on what might be legitimate IV's here?