Instrumental Variables

Based on Greene’s Note 13

Instrumental Variables

- Framework: $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, K variables in **X**.
- There exists a set of K variables, **Z** such that

  $\text{plim}(\mathbf{Z'X/n}) \neq \mathbf{0}$ but $\text{plim}(\mathbf{Z'}\varepsilon/n) = \mathbf{0}$

  The variables in **Z** are called instrumental variables.
- An alternative (to least squares) estimator of β is

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

- We consider the following:
  - Why use this estimator?
  - What are its properties compared to least squares?
- We will also examine an important application

IV Estimators

Consistent

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

$= (\mathbf{Z'X}/n)^{-1} (\mathbf{Z'X}/n)\boldsymbol{\beta} + (\mathbf{Z'X}/n)^{-1}\mathbf{Z'}\boldsymbol{\varepsilon}/n$

$= \boldsymbol{\beta} + (\mathbf{Z'X}/n)^{-1}\mathbf{Z'}\boldsymbol{\varepsilon}/n \rightarrow \boldsymbol{\beta}$

Asymptotically normal (same approach to proof as for OLS)

Inefficient – to be shown.

LS as an IV Estimator

The least squares estimator is

$$(\mathbf{X}\ \mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}\ \mathbf{X})^{-1}\Sigma_i \mathbf{x}_i y_i$$

$$= \beta + (\mathbf{X}\ \mathbf{X})^{-1}\Sigma_i \mathbf{x}_i \varepsilon_i$$

If $\text{plim}(\mathbf{X}'\mathbf{X}/n) = \mathbf{Q}$ nonzero

$\text{plim}(\mathbf{X}'\boldsymbol{\varepsilon}/n) = \mathbf{0}$

Under the usual assumptions LS is an IV estimator **X** is its own instrument.

IV Estimation

Why use an IV estimator? Suppose that $\mathbf{X}$ and ε are *not* uncorrelated. Then least squares is neither unbiased nor consistent.

Recall the proof of consistency of least squares:

$$\mathbf{b} = \beta + (\mathbf{X'X}/n)^{-1}(\mathbf{X'}\varepsilon/n).$$

Plim $\mathbf{b} = \beta$ requires plim$(\mathbf{X'}\varepsilon/n) = \mathbf{0}$. If this does not hold, the estimator is inconsistent.

A Popular Misconception

A popular misconception. If only one variable in **X** is correlated with ε, the other coefficients are consistently estimated. False.

Suppose only the first variable is correlated with **ε**

Under the assumptions, $\text{plim}(\mathbf{X}'\boldsymbol{\varepsilon}/n) = \begin{pmatrix} \sigma_{1\varepsilon} \\ 0 \\ \cdots \\ . \end{pmatrix}$. Then

$$\text{plim } \mathbf{b} - \boldsymbol{\beta} = \text{plim}(\mathbf{X}'\mathbf{X}/n)^{-1} \begin{pmatrix} \sigma_{1\varepsilon} \\ 0 \\ \cdots \\ . \end{pmatrix} = \sigma_{1\varepsilon} \begin{pmatrix} q^{11} \\ q^{21} \\ \cdots \\ q^{K1} \end{pmatrix}$$

$= \sigma_{1\varepsilon}$ times the first column of $\mathbf{Q}^{-1}$

The problem is "smeared" over the other coefficients.

The General Result

By construction, the IV estimator is consistent. So, we have an estimator that is consistent when least squares is not.

Asymptotic Covariance Matrix of b_{IV}

$$\mathbf{b}_{\text{IV}} - \beta = (\mathbf{Z'X})^{-1}\mathbf{Z}'\varepsilon$$

$$(\mathbf{b}_{\text{IV}} - \beta)(\mathbf{b}_{\text{IV}} - \beta)' = (\mathbf{Z'X})^{-1}\mathbf{Z}'\varepsilon\varepsilon'\mathbf{Z}(\mathbf{X'Z})^{-1}$$

$$E[(\mathbf{b}_{\text{IV}} - \beta)(\mathbf{b}_{\text{IV}} - \beta)' \mid \mathbf{X}, \mathbf{Z}] = \sigma^2(\mathbf{Z'X})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{X'Z})^{-1}$$

Asymptotic Efficiency

Asymptotic efficiency of the IV estimator. The variance is larger than that of LS. (A large sample type of Gauss-Markov result is at work.)

(1) It's a moot point. LS is inconsistent.

(2) Mean squared error is uncertain:

MSE[estimator|$\boldsymbol{\beta}$]=Variance + square of bias.

IV may be better or worse. Depends on the data

Two Stage Least Squares

How to use an "excess" of instrumental variables

(1) **X** is K variables. Some (at least one) of the K variables in **X** are correlated with **ε**.

(2) **Z** is M > K variables. Some of the variables in **Z** are also in **X**, some are not. None of the variables in **Z** are correlated with **ε.**

(3) Which K variables to use to compute **Z'X** and **Z'y?**

Choosing the Instruments

- Choose K randomly?
- Choose the included Xs and the remainder randomly?
- Use all of them? How?
- A theorem: (Brundy and Jorgenson, ca. 1972) There is a most efficient way to construct the IV estimator from this subset:
  - (1) For each column (variable) in **X**, compute the predictions of that variable using all the columns of **Z**.
  - (2) Linearly regress **y** on these K predictions.
- This is two stage least squares

Algebraic Equivalence

- Two stage least squares is equivalent to
  - (1) each variable in **X** that is also in **Z** is replaced by itself.
  - (2) Variables in **X** that are not in **Z** are replaced by predictions of that **X** with all the variables in **Z** that are not in **X**.

2SLS Algebra

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

$$\mathbf{b_{2SLS}} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{y}$$

But, $\mathbf{Z(Z'Z)^{-1}Z'X} = (\mathbf{I} - \mathbf{M_Z})\mathbf{X}$ and $(\mathbf{I} - \mathbf{M_Z})$ is idempotent.

$\hat{\mathbf{X}}'\hat{\mathbf{X}} = \mathbf{X}'(\mathbf{I} - \mathbf{M_Z})(\mathbf{I} - \mathbf{M_Z})\mathbf{X} = \mathbf{X}'(\mathbf{I} - \mathbf{M_Z})\mathbf{X}$ so

$\mathbf{b_{2SLS}} = (\hat{\mathbf{X}}'\mathbf{X})^{-1}\hat{\mathbf{X}}'\mathbf{y}$ = a real IV estimator by the definition.

Note, plim$(\hat{\mathbf{X}}'\varepsilon/n)$ = $\mathbf{0}$ since columns of $\hat{\mathbf{X}}$ are linear combinations of the columns of $\mathbf{Z}$, all of which are uncorrelated with ε.

$$\mathbf{b_{2SLS}} = [\mathbf{X}'(\mathbf{I} - \mathbf{M_Z})\mathbf{X}]^{-1}\mathbf{X}'(\mathbf{I} - \mathbf{M_Z})\mathbf{y}$$

Asymptotic Covariance Matrix for 2SLS

General Result for Instrumental Variable Estimation

$$E[(\mathbf{b}_{IV}-\beta)(\mathbf{b}_{IV}-\beta)' \mid \mathbf{X},\mathbf{Z}] = \sigma^2(\mathbf{Z'X})^{-1}\mathbf{Z'Z}(\mathbf{X'Z})^{-1}$$

Specialize for 2SLS, using $\mathbf{Z} = \hat{\mathbf{X}} = (\mathbf{I}-\mathbf{M_Z})\mathbf{X}$

$$\begin{aligned}E[(\mathbf{b}_{2SLS}-\beta)(\mathbf{b}_{2SLS}-\beta)' \mid \mathbf{X},\mathbf{Z}] &= \sigma^2(\hat{\mathbf{X}}'\mathbf{X})^{-1}\hat{\mathbf{X}}'\hat{\mathbf{X}}(\mathbf{X}'\hat{\mathbf{X}})^{-1} \\ &= \sigma^2(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\hat{\mathbf{X}}(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1} \\ &= \sigma^2(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\end{aligned}$$

2SLS Has Larger Variance than LS

A comparison to OLS

$\text{Asy.Var[2SLS]}=\sigma^2(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}$

Neglecting the inconsistency,

$\text{Asy.Var[LS]} \quad =\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$

(This is the variance of LS around its mean, not $\boldsymbol{\beta}$)

Asy.Var[2SLS] $\geq$ Asy.Var[LS] in the matrix sense.

Compare inverses:

$\{\text{Asy.Var[LS]}\}^{-1} - \{\text{Asy.Var[2SLS]}\}^{-1} = (1/\sigma^2)[\mathbf{X}'\mathbf{X} - \hat{\mathbf{X}}'\hat{\mathbf{X}}]$

$= (1/\sigma^2)[\mathbf{X}'\mathbf{X} - \mathbf{X}'(\mathbf{I} - \mathbf{M}_Z)\mathbf{X}] = (1/\sigma^2)[\mathbf{X}'\mathbf{M}_Z\mathbf{X}]$

This matrix is nonnegative definite. (Not positive definite as it might have some rows and columns which are zero.)

Implication for "precision" of 2SLS.

The problem of "Weak Instruments"

Estimating σ^2

Estimating the asymptotic covariance matrix -

a caution about estimating σ^2.

Since the regression is computed by regressing y on $\hat{\mathbf{x}}$,

one might use

$$\hat{\sigma}^2 = \frac{1}{n}\Sigma_{i=1}^{n}(y_i - \hat{\mathbf{x}}'\mathbf{b}_{2sls})$$

This is inconsistent. Use

$$\hat{\sigma}^2 = \frac{1}{n}\Sigma_{i=1}^{n}(y_i - \mathbf{x}'\mathbf{b}_{2sls})$$

(Degrees of freedom correction is optional. Conventional, but not necessary.)

Measurement Error

$y = \beta x^* + \varepsilon$ all of the usual assumptions

$x = x^* + u$ the true x^* is not observed (education vs. years of school)

What happens when y is regressed on x? Least squares attenutation:

$$\text{plim } b = \frac{\text{cov}(x,y)}{\text{var}(x)} = \frac{\text{cov}(x^* + u, \beta x^* + \varepsilon)}{\text{var}(x^* + u)}$$

$$= \frac{\beta\,\text{var}(x^*)}{\text{var}(x^*) + \text{var}(u)} < \beta$$

Why Is Least Squares Attenuated?

$y = \beta x^* + \varepsilon$

$x = x^* + u$

$y = \beta x + (\varepsilon - \beta u)$

$y = \beta x + v$, $cov(x,v) = -\beta\, var(u)$

Some of the variation in x is not associated with variation in y. The effect of variation in x on y is dampened by the measurement error.

Measurement Error in Multiple Regression

Multiple regression: y = $\beta_1 x_1 * + \beta_2 x_2 * + \varepsilon$

$x_1 *$ is measured with error; $x_1 = x_1 * + u$

x_2 is measured without error.

The regression is estimated by least squares

Popular myth #1. b_1 is biased downward, b_2 consistent.

Popular myth #2. All coefficients are biased toward zero.

Result for the simplest case. Let

$\sigma_{ij} = \text{cov}(x_i *, x_j *), i, j = 1, 2$ (2x2 covariance matrix)

$\sigma^{ij} =$ ijth element of the inverse of the covariance matrix

$\theta^2 = \text{var}(u)$

For the least squares estimators:

$$\text{plim } b_1 = \beta_1 \left(\frac{1}{1 + \theta^2 \sigma^{11}} \right), \quad \text{plim } b_2 = \beta_2 - \beta_1 \left(\frac{\theta^2 \sigma^{12}}{1 + \theta^2 \sigma^{11}} \right)$$

The effect is called "smearing."

Twins

Application from the literature: Ashenfelter/Kreuger: A wage equation that includes “schooling.”

Orthodoxy

- A proxy is not an instrumental variable
- Instrument is a noun, not a verb