Topic 2: Asymptotic Properties of Various Regression Estimators

- Our results to date apply for any *finite* sample size (n).
- In more general models we often can't obtain *exact* results for estimators' properties.
- In this case, we might consider their properties as $n \to \infty$.
- A way of "approximating" results.
- Also of interest in own right – inferential procedures should "work well" when we have lots of data
- Previous example – hypothesis tests that are "consistent".

Definition: An estimator, $\hat{\boldsymbol{\theta}}$, for $\boldsymbol{\theta}$, is said to be (weakly) *consistent* if

$$\lim_{n\to\infty}\{Pr.[|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}| < \boldsymbol{\epsilon}]\} = 1.$$

Note: A *sufficient* condition for this to hold is that *both*

(i) $Bias(\hat{\boldsymbol{\theta}}_n) \to \mathbf{0}$; as $n \to \infty$.

(ii) $V(\hat{\boldsymbol{\theta}}_n) \to 0$; as $n \to \infty$.

We call this "Mean Square Consistency". (Often useful for checking.)

If $\hat{\boldsymbol{\theta}}$ is weakly consistent for $\boldsymbol{\theta}$, we say that "the probability limit of $\hat{\boldsymbol{\theta}}$ equals $\boldsymbol{\theta}$.

We denote this by using "*plim*" operator, and we write

$$plim(\hat{\boldsymbol{\theta}}_n) = \boldsymbol{\theta} \quad \text{or,} \quad \hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}$$

Example $x_i \sim [\mu, \sigma^2]$ (*i.i.d*)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$E[\bar{x}] = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}(n\mu) = \mu \qquad \text{(unbiased, for all } n\text{)}$$

$$var.[\bar{x}] = \frac{1}{n^2}var.[\sum_{i=1}^{n} x_i] = \frac{1}{n^2}\sum_{i=1}^{n} var.(x_i)$$

$$= \frac{1}{n^2}(n\sigma^2) = \sigma^2/n$$

So, $\bar{x}$ is an unbiased estimator of μ, and $\lim_{n\to\infty}\{var.[\bar{x}]\} = 0$.

This *implies* that $\bar{x}$ is both a **mean-square consistent**, and **weakly consistent** estimator of μ .

Note:

- If an estimator is *inconsistent*, then it is a pretty useless estimator!
- There are many situations in which our LS estimator is *inconsistent*!
- For example –

(i) $y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + \varepsilon_t$

and $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$

(ii) $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \varepsilon_{1t}$

and $x_{2t} = \gamma_1 y_t + \gamma_3 x_{3t} + \gamma_4 x_{4t} + \varepsilon_{2t}$

Slutsky's Theorem

Let $plim(\widehat{\boldsymbol{\theta}}_n) = \boldsymbol{c}$, and let $f(\,.\,)$ be any *continuous* function.

Then, $plim[f(\widehat{\boldsymbol{\theta}}_n)] = f(\boldsymbol{c})$.

For example –

$$plim\left(\frac{1}{\hat{\theta}}\right) = \frac{1}{c} \quad ; \quad scalars$$

$$plim\left(e^{\widehat{\boldsymbol{\theta}}}\right) = e^{\boldsymbol{c}} \quad ; \quad vectors$$

$$plim\left(\hat{\Theta}^{-1}\right) = C^{-1} \quad ; \quad matrices$$

A very useful result – the "**plim**" operator can be used very flexibly.

Asymptotic Properties of LS Estimator(s)

- Consider LS estimator of β under our standard assumptions, in the "large *n*" *asymptotic* case.
- Can relax some assumptions:
  (i) Don't need Normality assumption for the error term of our model
  (ii) Columns of X can be random – just assume that $\{\boldsymbol{x}_i' ,\ \varepsilon_i\}$ is a random and *independent* sequence; $i = 1, 2, 3, \ldots\ldots$
  (iii) Last assumption implies $plim[n^{-1}X'\boldsymbol{\varepsilon}] = \mathbf{0}$. (Greene, pp. 64-65.)
- *Amend* (extend) our assumption about X having full column rank – assume instead that $plim[n^{-1}X'X] = Q$; positive-definite & finite
- Note that Q is $(k \times k)$, symmetric, and *unobservable*.
- What are we assuming about the elements of X, which is $(n \times k)$, as n increases without limit?

Theorem: The LS estimator of $\boldsymbol{\beta}$ is *weakly consistent*.

Proof: $$\boldsymbol{b} = (X'X)^{-1}X'\boldsymbol{y} = (X'X)^{-1}X'(X\boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

$$= \boldsymbol{\beta} + (X'X)X'\boldsymbol{\varepsilon}$$

$$= \boldsymbol{\beta} + \left[\frac{1}{n}(X'X)\right]^{-1}\left[\frac{1}{n}X'\boldsymbol{\varepsilon}\right].$$

If we now apply Slutsky's Theorem repeatedly, we have:

$$plim(\boldsymbol{b}) = \boldsymbol{\beta} + Q^{-1}.\mathbf{0} = \boldsymbol{\beta}\ .$$

- We can also show that s^2 is a consistent estimator for σ^2.
- Do this in two ways (different assumptions).
- First, assume the errors are Normally distributed – get a strong result.
- Then, relax this assumption and get a weaker result.

Theorem: If the regression model errors are Normally distributed, then s^2 is a *mean-square consistent* estimator for σ^2.

Proof:

If the errors are Normal, then we know that

$$\frac{(n-k)s^2}{\sigma^2} \sim \chi^2_{(n-k)}$$

Now, (1) $E\left[\chi^2_{(n-k)}\right] = (n-k)$

(2) $var.\left[\chi^2_{(n-k)}\right] = 2(n-k)$

So, $E(s^2) = \frac{\sigma^2 E\left[\chi^2_{(n-k)}\right]}{n-k} = \sigma^2$; *unbiased*

$$var.\left[\frac{(n-k)s^2}{\sigma^2}\right] = 2(n-k)$$

$\Rightarrow$ $\left[\frac{(n-k)^2}{\sigma^4}\right] var.(s^2) = 2(n-k)$

$\Rightarrow$ $var.(s^2) = 2\sigma^4/(n-k)$

So, $var.(s^2) \to 0$, as $n \to \infty$ (and *unbiased*)

This implies that s^2 is a *mean-square consistent* estimator for σ^2.

(Implies, in turn, that it is also a *weakly consistent estimator*.)

- With the addition of the (relatively) strong assumption of Normally distributed errors, we get the (relatively) strong result.
- Note that $\hat{\sigma}^2 = (e'e)/n$ is also a *consistent* estimator, even though it is *biased*.
- What other assumptions did we use in the above proof?
- What can we say if we relax the assumption of Normality?
- We need a preliminary result to help us.

Theorem (Khintchine ; WLLN):

Suppose that $\{x_i\}_{i=1}^{n}$ is a sequence of random variables that are *uncorrelated*, and all drawn from the same distribution with a *finite* mean, μ, and a f*inite* variance, σ^2.

Then, $plim(\bar{x}) = \mu$.

Theorem: In our regression model, s^2 is a w*eakly consistent* estimator for σ^2.

(Notice that this also means that $\hat{\sigma}^2$ is also a weakly consistent estimator, so start with the latter estimator.)

Proof:

$$\hat{\sigma}^2 = \left(\frac{e'e}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

$$= \frac{1}{n}(M\boldsymbol{\varepsilon})'(M\boldsymbol{\varepsilon}) = \frac{1}{n}\boldsymbol{\varepsilon}'M\boldsymbol{\varepsilon}$$

$$= \frac{1}{n}[\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}'X(X'X)^{-1}X'\boldsymbol{\varepsilon}]$$

$$= \left[\left(\frac{1}{n}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right) - \left(\frac{1}{n}\boldsymbol{\varepsilon}'X\right)\left(\frac{1}{n}X'X\right)^{-1}\left(\frac{1}{n}X'\boldsymbol{\varepsilon}\right)\right].$$

So, $$plim(\hat{\sigma}^2) = plim\left(\frac{1}{n}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right) - \mathbf{0}'Q^{-1}\mathbf{0} = plim\left[\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2\right].$$

Now, if the errors are pair-wise uncorrelated, so are their squared values.

Also, $E[\varepsilon_i^2] = var.(\varepsilon_i) = \sigma^2$.

By Khintchine's Theorem, we immediately have the result:

$$plim(\hat{\sigma}^2) = \sigma^2,$$

and so $$plim(s^2) = \sigma^2.$$

- Relaxing the assumption of Normally distributed errors led to a *weaker result* for the consistent estimation of the error variance.
- What other assumptions were used, and where?

An Issue

- Suppose we want to compare the (large *n*) asymptotic behaviour of our LS estimators with those of other potential estimators.
- These other estimators will presumably also be *consistent*.
- This means that *in each case* the sampling distributions of the estimators collapse to a "spike", located exactly at the true parameter values.
- So, how can we compare such estimators when *n* is very large – aren't they *indistinguishable*?
- If the limiting density of any consistent estimator is a degenerate "spike", it will have zero variance, in the limit.
- Can we still compare large-sample variances of consistent estimators?
  In other words, is it meaningful to think about the concept of asymptotic efficiency?

Asymptotic Efficiency

- **The key to asymptotic efficiency is to "control" for the fact that the distribution of any *consistent* estimator is "collapsing", as $n \to \infty$.**
- The *rate* at which the distribution collapses is crucially important.
- This is probably best understood by considering an example.
- $\{x_i\}_{i=1}^{n}$; *random sampling* from $[\mu, \sigma^2]$.
- $E[\bar{x}] = \mu$; $var.[\bar{x}] = \sigma^2/n$
- Now construct: $y = \sqrt{n}(\bar{x} - \mu)$.
- Note that $E(y) = \sqrt{n}(E(\bar{x}) - \mu) = 0$.
- Also, $var.[y] = \left(\sqrt{n}\right)^2 var.(\bar{x} - \mu) = n\, var.(\bar{x}) = \sigma^2$.
- The scaling we've used results in a finite, non-zero, variance.
- $E(y) = 0$, and $var.[y] = \sigma^2$; *unchanged* as $n \to \infty$.
- So, $y = \sqrt{n}(\bar{x} - \mu)$ has a well-defined "limiting" (asymptotic) distribution.
- The *asymptotic mean* of *y* is zero, and the *asymptotic variance* of *y* is σ^2.
- Question – Why did we scale by $\sqrt{n}$, and not (say), by *n* itself ?

- In fact, because we had *independent* x_i's (*random sampling*), we have the additional result that $y = \sqrt{n}(\bar{x} - \mu) \xrightarrow{d} \boldsymbol{N}[0\,, \sigma^2]$, the *Lindeberg-Lévy Central Limit Theorem*.
- Now we can define "Asymptotic Efficiency" in a meaningful way.

Definition: Let $\hat{\theta}$ and $\tilde{\theta}$ be two *consistent* estimator of θ ; and suppose that

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} [0\,, \sigma^2] \text{ , and } \sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{d} [0\,, \varphi^2] \text{ .}$$

Then $\hat{\theta}$ is "*asymptotically efficient*" relative to $\tilde{\theta}$ if $\sigma^2 < \varphi^2$.

In the case where $\boldsymbol{\theta}$ is a *vector*, $\hat{\boldsymbol{\theta}}$ is "*asymptotically efficient*" relative to $\tilde{\boldsymbol{\theta}}$ if

$\Delta = asy.V(\tilde{\boldsymbol{\theta}}) - asy.V(\hat{\boldsymbol{\theta}})$ is positive definite.

Asymptotic Distribution of the LS Estimator:

Let's consider the full asymptotic distribution of the LS estimator, $\boldsymbol{b}$, for $\boldsymbol{\beta}$ in our linear regression model.

We'll actually have to consider the behaviour of $\sqrt{n}(\boldsymbol{b} - \boldsymbol{\beta})$:

$$\sqrt{n}(\boldsymbol{b} - \boldsymbol{\beta}) = \sqrt{n}[(X'X)^{-1}X'\boldsymbol{\varepsilon}]$$

$$= \left[\frac{1}{n}(X'X)\right]^{-1}\left(\frac{1}{\sqrt{n}}X'\boldsymbol{\varepsilon}\right).$$

It can be shown, by the Lindeberg-Feller Central Limit Theorem, that

$$\left(\frac{1}{\sqrt{n}}X'\boldsymbol{\varepsilon}\right) \xrightarrow{d} N[0\,, \sigma^2 Q],$$

where $$Q = plim\left[\frac{1}{n}(X'X)\right].$$

So, the asymptotic covariance matrix of $\sqrt{n}(\boldsymbol{b} - \boldsymbol{\beta})$ is

$$plim\left[\frac{1}{n}(X'X)\right]^{-1}(\sigma^2 Q)plim\left[\frac{1}{n}(X'X)\right]^{-1} = \sigma^2 Q^{-1}.$$

In full, the asymptotic distribution of $\boldsymbol{b}$ is correctly stated by saying that:

$$\sqrt{n}(\boldsymbol{b} - \boldsymbol{\beta}) \xrightarrow{d} N[\mathbf{0}, \sigma^2 Q^{-1}]$$

The asymptotic covariance matrix is *unobservable*, for two reasons:

1. σ^2 is typically *unknown*.
2. Q is *unobservable*.

- We can estimate σ^2 *consistently*, using s^2.
- To estimate $\sigma^2 Q^{-1}$ consistently, we can use $ns^2(X'X)^{-1}$:

$$plim[ns^2(X'X)^{-1}] = plim(s^2)plim\left[\frac{1}{n}(X'X)\right]^{-1} = \sigma^2 Q^{-1}.$$

The square roots of the diagonal elements of $ns^2(X'X)^{-1}$ are the *asymptotic std. errors* for the elements of $\sqrt{n}(\boldsymbol{b} - \boldsymbol{\beta})$.

Loosely speaking, the asymptotic covariance matrix for $\boldsymbol{b}$ itself is $s^2(X'X)^{-1}$; and the square roots of the diagonal elements of this matrix are the *asymptotic std. errors* for the b_i's themselves.

Instrumental Variables

- We have been assuming *either* that the columns of X are *non-random*; *or* that the sequence $\{\boldsymbol{x}_i', \varepsilon_i\}$ is *independent.* Often, neither of these assumptions is tenable.
- This implies that $plim\left(\frac{1}{n}X'\boldsymbol{\varepsilon}\right) \neq \mathbf{0}$, and then the LS estimator is *inconsistent* (prove this).
- In order to motivate a situation where $\{\boldsymbol{x}_i', \varepsilon_i\}$ are *dependent*, consider an omitted, or unobservable variable.

We will consider a situation where the unobservable variable is correlated with one of the regressors, and correlated with the dependent variable.

Consider the population model:

$$\boldsymbol{y} = X_1\beta_1 + X_2\beta_2 + \boldsymbol{\varepsilon}_1. \qquad [1]$$

Consider that $cov(X_1, X_2) \neq 0$. For example, X_2 causes X_1:

$$X_1 = X_2\gamma + \boldsymbol{\varepsilon}_2. \qquad [2]$$

Now consider that X_2 is unobservable, so that the observable model is:

$$\boldsymbol{y} = X_1\beta_1 + \boldsymbol{\varepsilon}_3. \qquad [3]$$

- Notice that in [3], $\boldsymbol{\varepsilon}_3$ contains $\beta_2 X_2$, so that X_1 and $\boldsymbol{\varepsilon}_3$ are not independent (X_1 is *endogenous*)
- OLS will be biased, since $E[\boldsymbol{\varepsilon}_3|X_1] \neq \mathbf{0}$
- Note that when estimating from [3], $E[b_1] = \beta_1 + \gamma^{-1}\beta_2$
- OLS will be inconsistent, since $plim\left(\frac{1}{n}X_1{'}\boldsymbol{\varepsilon}_3\right) \neq \mathbf{0}$
- In such cases we want a safe way of estimating β_1.
- We just want to ensure that we have an estimator that is (at least) *consistent*.
- One general family of such estimators is the family of **Instrumental Variables (I.V.) Estimators**.

An instrumental variable, Z, must be:

1. Correlated with the endogenous variable(s) X_1
    - Sometimes called the "relevance" of an I.V.
    - This condition can be tested
2. Uncorrelated with the error term, or equivalently, uncorrelated with the dependent variable other than through its correlation with X_1
    - Sometimes called the "exclusion" restriction
    - This restriction cannot be tested directly

Suppose now that we have a variable Z which is

- Relevant: $cov(Z, X_1) \neq 0$
- Satisfies exclusion restriction: $cov(Z, \boldsymbol{\varepsilon}) = 0$. In the above D.G.P.s ([1]- [3]), it is sufficient for the instrument to be uncorrelated with the unobservable variable: $cov(Z, X_2) = 0$.

Validity means that [2] becomes:

$$X_1 = Z\delta + X_2\gamma + \boldsymbol{\varepsilon}_4 \qquad [4]$$

Substituting [4] into [1]:

$$\boldsymbol{y} = X_2\gamma\beta_1 + Z\delta\beta_1 + X_2\beta_2 + \boldsymbol{\varepsilon}_5. \qquad [5]$$

X_2 is still unobservable, but is uncorrelated with Z! The observable population model is now:

$$\boldsymbol{y} = Z\delta\beta_1 + \boldsymbol{\varepsilon}_6. \qquad [6]$$

Now, we have a population model involving β_1, and where $cov(Z, \boldsymbol{\varepsilon_6}) = 0$. So, $(\delta\beta_1)$ can be estimated by OLS. But we need β_1!

By Slutsky's Theorem, if $plim\left(\widehat{\delta\beta_1}\right) = \delta\beta_1$, and if $plim\left(\hat{\delta}\right) = \delta$, then $plim\left(\hat{\delta}^{-1}\widehat{\delta\beta_1}\right) = \beta_1$. So if we can find a consistent estimator for δ, we can find one for β_1. How to estimate δ?

Recall [4]. Since X_2 and Z are uncorrelated, we can estimate δ by an OLS regression of X_1 on Z:

$$\hat{\delta} = (Z'Z)^{-1}Z'X_1$$

Now solve for $\widehat{\beta_1}$:

$$\widehat{\beta_1} = \hat{\delta}^{-1}\widehat{\delta\beta_1} = [(Z'Z)^{-1}Z'X_1]^{-1}(Z'Z)^{-1}Z'\boldsymbol{y}$$

If Z and X_1 have the same number of columns, then:

$$\widehat{\beta_1} = (Z'X_1)^{-1}Z'Z(Z'Z)^{-1}Z'\boldsymbol{y} = (Z'X_1)^{-1}Z'\boldsymbol{y}$$

In this example we had one endogenous variable (X_1) and one instrument (Z). In this case, the I.V. estimate may be found by the OLS estimate from a regression of $\boldsymbol{y}$ on Z by the OLS estimates of a regression of X_1 on Z.

In more general models, we will have more explanatory variables. As long as there is one instrument per endogenous variable, I.V. is possible and the simple I.V. estimator is:

$$\boxed{b_{IV} = (Z'X)^{-1}Z'\boldsymbol{y}}$$

In general, this estimator is biased. We can show it's consistent, however:

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$plim\left(\frac{1}{n}X'X\right) = Q \quad ; \quad \text{p.d. and } \textit{finite}$$

$$plim\left(\frac{1}{n}X'\boldsymbol{\varepsilon}\right) = \boldsymbol{\gamma} \neq \boldsymbol{0}$$

Find a (random) $(n \times k)$ matrix, Z, such that:

1. $plim\left(\frac{1}{n}Z'Z\right) = Q_{ZZ} \quad ; \quad$ p.d. and *finite*.
2. $plim\left(\frac{1}{n}Z'X\right) = Q_{ZX} \quad ; \quad$ p.d. and *finite*.
3. $plim\left(\frac{1}{n}Z'\boldsymbol{\varepsilon}\right) = \boldsymbol{0}$.

Then, consider the estimator: $\boldsymbol{b_{IV}} = (Z'X)^{-1}Z'\boldsymbol{y}$. This is a *consistent* estimator of $\boldsymbol{\beta}$.

$$\boldsymbol{b_{IV}} = (Z'X)^{-1}Z'\boldsymbol{y} = (Z'X)^{-1}Z'(X\boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

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

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

$$= \boldsymbol{\beta} + \left(\frac{1}{n}Z'X\right)^{-1}\left(\frac{1}{n}Z'\boldsymbol{\varepsilon}\right) .$$

So, $$plim(\boldsymbol{b_{IV}}) = \boldsymbol{\beta} + \left[plim\left(\frac{1}{n}Z'X\right)\right]^{-1} plim\left(\frac{1}{n}Z'\boldsymbol{\varepsilon}\right)$$

$$= \boldsymbol{\beta} + Q_{ZX}^{-1}\boldsymbol{0} = \boldsymbol{\beta} \qquad \textit{(consistent)}$$

Choosing different Z matrices generates different members of I.V. family.

Although we won't *derive* the full asymptotic distribution of the I.V. estimator, note that it can be expressed as:

$$\boxed{\sqrt{n}(\boldsymbol{b_{IV}} - \boldsymbol{\beta}) \xrightarrow{d} N[\mathbf{0}, \sigma^2 Q_{ZX}{}^{-1} Q_{ZZ} Q_{XZ}{}^{-1}]}$$

where $Q_{XZ} = Q_{ZX}{}'$. [How would you estimate Asy. Covar. Matrix?]

Interpreting I.V. as two-stage least squares (2SLS)

1st stage: Regress X on Z, get $\hat{X}$.

- $\hat{X}$ contains the variation in X due to Z *only*
- $\hat{X}$ is not correlated with $\boldsymbol{\varepsilon}$

2nd stage: Estimate the model $\boldsymbol{y} = \hat{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

From 1st stage: $\hat{X} = Z(Z'Z)^{-1}Z'X$

From 2nd stage: $\boldsymbol{b_{IV}} = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'\boldsymbol{y}$

In fact, this is the *Generalized* I.V. estimator of $\boldsymbol{\beta}$. We can actually use more instruments than regressors (the "Over-Identified" case).

Note that if X and Z have the same dimensions, then the generalized estimator collapses to the simple one.

Let's check the consistency of the I.V. estimator. Let $M_z = Z(Z'Z)^{-1}Z'$. Then the generalized I.V. estimator is:

$$\boxed{\boldsymbol{b_{IV}} = [X'M_ZX]^{-1}X'M_Z\boldsymbol{y}}$$

$$\begin{aligned}\boldsymbol{b_{IV}} &= [X'M_ZX]^{-1}X'M_Z\boldsymbol{y} = [X'M_ZX]^{-1}X'M_Z(X\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= [X'M_ZX]^{-1}X'M_ZX\boldsymbol{\beta} + [X'M_ZX]^{-1}X'M_Z\boldsymbol{\varepsilon} \\ &= \boldsymbol{\beta} + [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'\boldsymbol{\varepsilon}\end{aligned}$$

So,

$$\boldsymbol{b_{IV}} = \boldsymbol{\beta} + \left[\left(\frac{1}{n}X'Z\right)\left(\frac{1}{n}Z'Z\right)^{-1}\left(\frac{1}{n}Z'X\right)\right]^{-1}\left(\frac{1}{n}X'Z\right)\left(\frac{1}{n}Z'Z\right)^{-1}\left(\frac{1}{n}Z'\boldsymbol{\varepsilon}\right).$$

Modify our assumptions:

We have a (random) $(n \times L)$ matrix, Z, such that:

1. $plim\left(\frac{1}{n}Z'Z\right) = Q_{ZZ}$; $(L \times L)$, p.d.s. and *finite*.

2. $plim\left(\frac{1}{n}Z'X\right) = Q_{ZX}$; $(L \times k)$, rank = k, and *finite*.

3. $plim\left(\frac{1}{n}Z'\boldsymbol{\varepsilon}\right) = \mathbf{0}$; $(L \times 1)$

So,

$$plim(\boldsymbol{b_{IV}}) = \boldsymbol{\beta} + [Q_{XZ}Q_{ZZ}^{-1}Q_{ZX}]^{-1}Q_{XZ}Q_{ZZ}^{-1}\mathbf{0} = \boldsymbol{\beta} \quad ; \quad \textit{consistent}$$

Similarly, a *consistent estimator* of σ^2 is

$$s_{IV}^2 = (\boldsymbol{y} - X\boldsymbol{b_{IV}})'(\boldsymbol{y} - X\boldsymbol{b_{IV}})/n$$

residual vector

- Recall that each choice of Z leads to a *different* I.V. estimator.
- Z must be chosen in way that ensures consistency of the I.V. estimator.
- How might we choose a suitable set of instruments, *in practice*?
- If we have several "valid" sets of instruments, how might we choose between them?

For the "simple" regression model, recall that:

$$\sqrt{n}(\boldsymbol{b_{IV}} - \boldsymbol{\beta}) \xrightarrow{d} N[\mathbf{0}, \sigma^2 Q_{ZX}{}^{-1}Q_{ZZ}Q_{XZ}{}^{-1}]$$

so if $k = 1$,

$$Q_{ZZ} = plim(n^{-1}\sum_{i=1}^{n} z_i^2)$$

$$Q_{ZX} = plim\left(n^{-1}\sum_{i=1}^{n} z_i x_i\right) = Q_{XZ}$$

The *asymptotic efficiency* of $\boldsymbol{b_{IV}}$ will be higher, the more highly correlated are Z and X, *asymptotically*.

We need to find instruments that are uncorrelated with the errors, but highly correlated with the regressors – *asymptotically*.

This is not easy to do!

- **Time –series data** -
    1. Often, we can use lagged values of the regressors as suitable instruments.
    2. This will be fine as long as the errors are serially uncorrelated.
- **Cross-section data** –
    1. Geography, weather, biology.
    2. Various "old" tricks – *e.g.*, using "ranks" of the data as instruments.

Testing if I.V. estimation is needed

- Why does LS fail, and when do we need I.V.?
- If $plim\left(\frac{1}{n}X'\boldsymbol{\varepsilon}\right) \neq \mathbf{0}$.
- We can ***test*** to see if this is a problem, & decide if we should use LS or I.V.

The Hausman Test

We want to test $H_0 : plim\left(\frac{1}{n}X'\boldsymbol{\varepsilon}\right) = \mathbf{0}$ *vs.* $H_A : plim\left(\frac{1}{n}X'\boldsymbol{\varepsilon}\right) \neq \mathbf{0}$

- If we reject H_0, we will use I.V. estimation.
- If we cannot reject H_0, we'll use LS estimation.
- Hausman test is a general "testing strategy" that can be applied in many situations – not just for this particular situation!
- Basic idea – construct 2 estimators of $\boldsymbol{\beta}$:

1. $\boldsymbol{b_E}$: estimator is both *consistent and asymptotically efficient* if H_0 true.
2. $\boldsymbol{b_I}$: estimator is at least *consistent*, even if H_0 false.

- In our case here, $\boldsymbol{b_E}$ is the LS estimator; and $\boldsymbol{b_I}$ is the I.V. estimator.
- If H_0 is true, we'd expect ($\boldsymbol{b_I}$ - $\boldsymbol{b_E}$) to be "small", at least for large *n*, as both estimators are consistent in that case.
- Hausman shows that $\hat{V}(\boldsymbol{b_I} - \boldsymbol{b_E}) = \hat{V}(\boldsymbol{b_I}) - \hat{V}(\boldsymbol{b_E})$, if H_0 is true.
- So, the test statistic is, $H = (\boldsymbol{b_I} - \boldsymbol{b_E})'\left[\hat{V}(\boldsymbol{b_I}) - \hat{V}(\boldsymbol{b_E})\right]^{-1}(\boldsymbol{b_I} - \boldsymbol{b_E})$.
- $H \xrightarrow{\boldsymbol{d}} \boldsymbol{\chi^2_J}$, if H_0 is true.
- Here, *J* is the number of columns in *X* which *may* be correlated with the errors, & for which we need instruments.
- Problem – often, $\left[\hat{V}(\boldsymbol{b_I}) - \hat{V}(\boldsymbol{b_E})\right]$ is *singular*, so *H* is *not defined*.
- <u>One option</u> is to replace the "regular inverse" with a "generalized inverse".
- <u>Another option</u> is to modify *H* so that it becomes:

$$H^* = (\boldsymbol{b_I}^* - \boldsymbol{b_E}^*)'\left[\hat{V}(\boldsymbol{b_I}^*) - \hat{V}(\boldsymbol{b_E}^*)\right]^{-1}(\boldsymbol{b_I}^* - \boldsymbol{b_E}^*) \xrightarrow{\boldsymbol{d}} \boldsymbol{\chi^2_J} \text{ ; if } H_0 \text{ true.}$$

- Here, $\boldsymbol{b_I}^*$ and $\boldsymbol{b_E}^*$ are the $(J \times 1)$ vectors formed by using only the elements of $\boldsymbol{b_I}$ and $\boldsymbol{b_E}$ that correspond to the "problematic" regressors.
- Constructing *H** is not very convenient unless $J = 1$.

The Durbin-Wu Test

This test is *specific* to testing

$$H_0 : plim\left(\frac{1}{n}X'\boldsymbol{\varepsilon}\right) = \mathbf{0} \qquad vs. \qquad H_A : plim\left(\frac{1}{n}X'\boldsymbol{\varepsilon}\right) \neq \mathbf{0}$$

Again, an asymptotic test.

Testing the exogeneity of Instruments

The key assumption that ensures the consistency of I.V. estimators is that

$$plim\left(\frac{1}{n}Z'\boldsymbol{\varepsilon}\right) = \mathbf{0} \,.$$

This condition involves the *unobservable* $\boldsymbol{\varepsilon}$. In general, it cannot be tested.

"Weak Instruments" – Problems arise if the instruments are ***not*** well correlated with the regressors (not relevant).

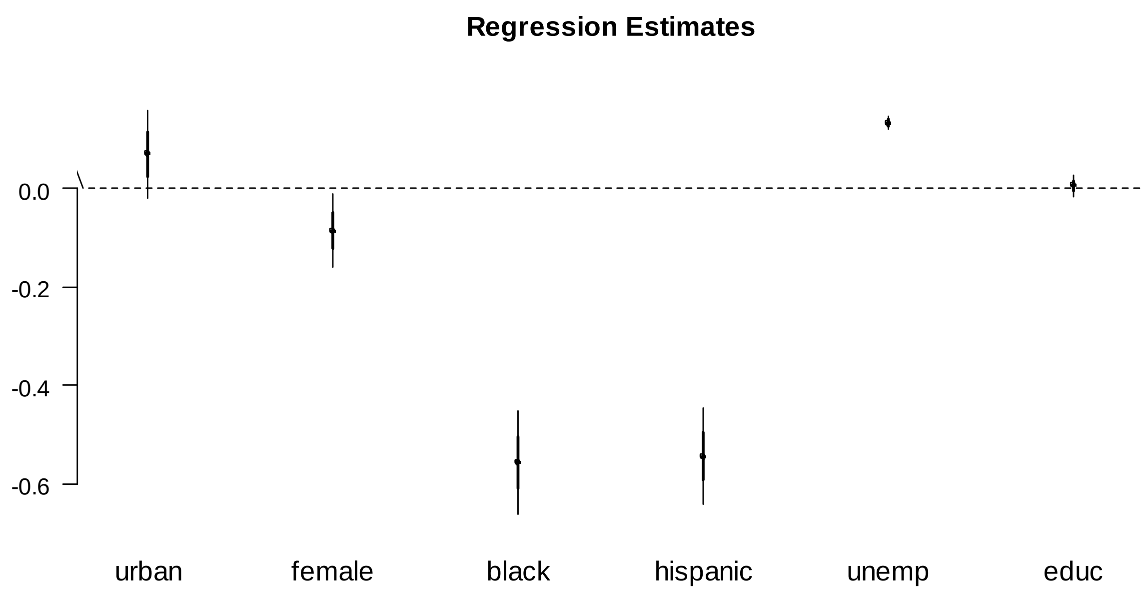
- These problems go beyond loss of asymptotic efficiency.
- Small-sample bias of I.V. estimator can be greater than that of LS!
- Sampling distribution of I.V. estimator can be bi-modal!
- Fortunately, we can again *test* to see if we have these problems.

Empirical Example: Using geographic variation in college proximity to estimate the return to schooling[1]

- Have data on wage, years of education, and demographic variables
- Want to estimate the return to education
- Problem: ability (intelligence) may be correlated with (cause) both wage and education
- Since ability is unobservable, it is contained in the error term
- The education variable is then correlated with the error term (endogenous)
- OLS estimation of the returns to education may be inconsistent

First, let's try OLS.

```
library(AER)
attach(CollegeDistance)
lm(wage ~ urban + gender + ethnicity + unemp + education)
```



Note that the returns to education are not statistically significant.

Now let's try using distance from college (while attending high school) as an instrument for education. For the instrument to be valid, we require that distance and education be correlated:

```
summary(lm(education ~ distance))
```

[1] Card, David. *Using geographic variation in college proximity to estimate the return to schooling*. No. w4483. National Bureau of Economic Research, 1993.

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 13.93861    0.03290 423.683  < 2e-16 ***
distance    -0.07258    0.01127  -6.441  1.3e-10 ***
```

While distance appears to be statistically significant, this isn't quite enough to test for validity (a testing problem we won't address here).

From the 2SLS interpretation, we know that we can get the IV estimator by:

1.) getting the predicted values from a regression of education on distance

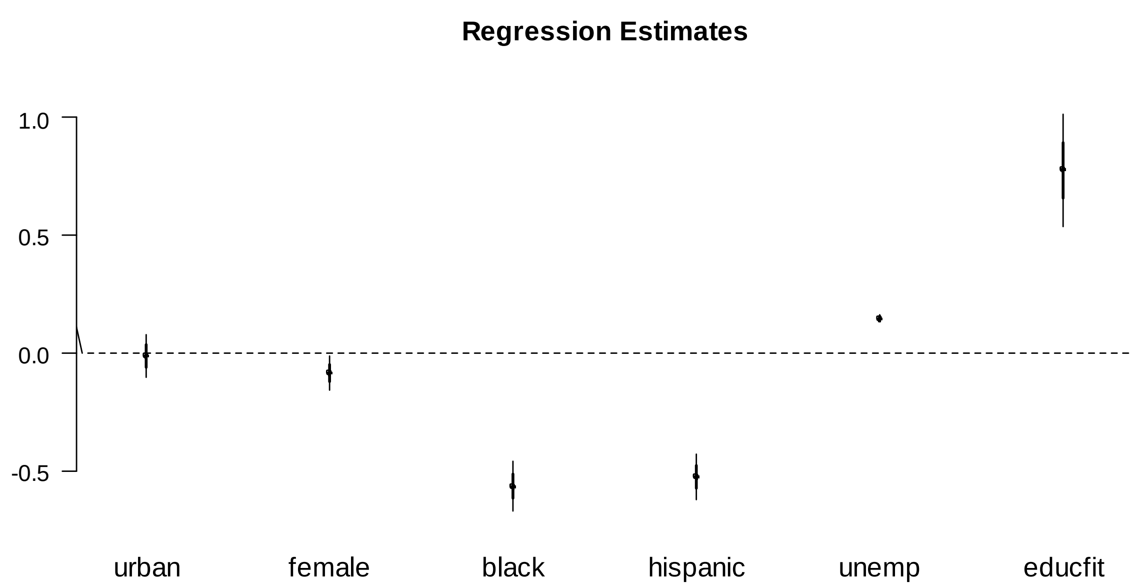
```
educfit = predict(lm(education ~ distance))
```

2.) regressing wage on the same variables, but using educfit instead of education

```
lm(wage ~ urban + gender + ethnicity + unemp + educfit)
```

Note that educfit is the variation in education as it can be explained by distance. These fitted values are uncorrelated with ability, since distance is uncorrelated with ability (by assumption).

Results of IV estimation:



The estimate for the return to education is now positive, and significant.