Economics 102C:
Advanced Topics in Econometrics
4 - Asymptotics & Large Sample Properties of OLS

Michael Best

Spring 2015

Asymptotics

- So far we have looked at the *finite sample* properties of OLS
- Relied heavily on the assumption that $\varepsilon_i|\mathbf{X} \sim N\left[0, \sigma^2\right]$
- Without this assumption, the t statistic doesn't have a t distribution, and F statistics don't have F distributions.
- Luckily, If we are unwilling to assume normality, we can still use **asymptotic (large sample) properties** of estimators.
- Look at what happens to estimators as sample size $n \to \infty$

Asymptotics

- Most estimators are functions of sample means, so focus on what happens to these means in large samples.
- Turns out things like the t and F statistics will have *approximately* t and F distributions in large samples.
- Most advanced methods justified exclusively on asymptotic grounds. Finite-sample properties are often unknown

Convergence in Probability

- We define asymptotic arguments with respect to the sample size n.
- Let x_n by a sequence random variable (a set of random variables $\{x_1, x_2, \ldots, x_n\}$) indexed by its sample size n

Definition (Convergence in Probability)

The random variable x_n ***converges in probability*** *to a constant c if* $\lim_{n\to\infty} \Pr\left(|x_n - c| > \varepsilon\right) = 0$ *for any $\varepsilon > 0$*

- The probability that x takes values far from c disappears as $n \to \infty$
- If x_n converges in probability to c, we say that plim $x_n = c$

Convergence in Mean Square

- It is often hard to verify convergence in probability, so we usually use a special case.

Definition (Convergence in Mean Square)

*If x_n has mean μ_n and variance σ_n^2 such that with limits c and 0, respectively, then x_n **converges in mean square** to c and plim $x_n = c$*

- Easier to check that the mean and variance have limits
- Note that mean square convergence implies convergence in porbability, but not vice versa

Definition

*An estimator $\hat{\theta}_n$ of a parameter θ is a **consistent** estimator of θ if and only if*

$$\textit{plim}\ \hat{\theta}_n = \theta$$

Consistency: Practice

- Turn to your neighbor: Let's take the example of the sample mean in an i.i.d. random sample: $\bar{x}_n \equiv \frac{1}{n}\sum_{i=1}^{n} x_i$:

1. Find $\mathrm{E}[\bar{x}_n]$ and $\mathrm{Var}[\bar{x}_n]$
2. What happens to $\mathrm{E}[\bar{x}_n]$ and $\mathrm{Var}[\bar{x}_n]$ as $n \to \infty$?
3. Using the definition of Convergence in Mean Square, what is plim $\bar{x}_n$?

Consistency: Practice

- Turn to your neighbor: Let's take the example of the sample mean in an i.i.d. random sample: $\bar{x}_n \equiv \frac{1}{n}\sum_{i=1}^{n} x_i$:

1. Find $\text{E}[\bar{x}_n]$ and $\text{Var}[\bar{x}_n]$

$$\text{E}[\bar{x}_n] = \text{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n}\text{E}[x_i] = \mu$$

$$\text{Var}[\bar{x}_n] = \text{Var}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\text{Var}\left[\sum_{i=1}^{n} x_i\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}[x] = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}$$

2. What happens to $\text{E}[\bar{x}_n]$ and $\text{Var}[\bar{x}_n]$ as $n \to \infty$?

$$\text{E}[\bar{x}_n] \to \mu \quad \text{Var}[\bar{x}_n] \to 0$$

3. Using the definition of Convergence in Mean Square, what is plim $\bar{x}_n$?

$$\bar{x}_n \overset{\text{mean square}}{\to} \mu \Rightarrow \text{plim } \bar{x}_n = \mu$$

Consistency

- In fact the example of the sample mean generalizes:
- For any function $g(x)$, if $\mathsf{E}[g(x)]$ and $\mathsf{Var}[g(x)]$ are finite constants, then

$$\text{plim}\ \frac{1}{n}\sum_{i=1}^{n} g(x_i) = \mathsf{E}[g(x)]$$

The Law of Large Numbers and Slutsky's Theorem

Theorem (*Law of Large Numbers*)

If x_i $i = 1, \ldots, n$ is a random (i.i.d.) sample from a distribution with finite mean E $[x_i] = \mu < \infty$, then

$$\textit{plim}\ \bar{x}_n = \mu$$

Theorem (Slutsky's Theorem)

For a continuous function $g(x_n)$ that is not a function of n,

$$\textit{plim}\ g(x_n) = g(\textit{plim}\ x_n)$$

- Proofs are hard. Not required for 102C
- These results are very powerful and allow us to show that estimators (functions of the data x_n) are consistent

Properties of plims

- probability limits have a number of useful properties: If x_n and y_n are RVs with plim $x_n = c$ and plim $y_n = d$ then

1. plim $(x_n + y_n) = c + d$
2. plim $x_n y_n = cd$
3. plim $x_n / y_n = c/d$ as long as $d \neq 0$

- If $\mathbf{W}_n$ is a matrix whose elements are RVs, and if plim $\mathbf{W}_n = \mathbf{\Omega}$

$$\text{plim } \mathbf{W}_n^{-1} = \mathbf{\Omega}^{-1}$$

- If $\mathbf{X}_n$ and $\mathbf{Y}_n$ are random matrices with plim $\mathbf{X}_n = \mathbf{A}$ and plim $\mathbf{Y}_n = \mathbf{B}$

$$\text{plim } \mathbf{X}_n \mathbf{Y}_n = \mathbf{AB}$$

Convergence in Distribution

- We use the plim to analyze whether estimators are *consistent*
- In order to make inference (for e.g. is a coefficient = 0?) we need to know the *distribution* of the estimator

Definition (Convergence in Distribution)

x_n converges in distribution to a random variable x with CDF $F(x)$ if $\lim_{n\to\infty} |F_n(x_n) - F(x)| = 0$ over the whole support of $F(x)$. We denote this as

$$x_n \xrightarrow{d} x$$

Rules for Convergence in Distribution

- Analogously to the rules for plims, if $x_n \xrightarrow{d} x$ and plim $y_n = c$, then

$$x_n y_n \xrightarrow{d} cx$$

$$x_n + y_n \xrightarrow{d} x + c$$

$$x_n / y_n \xrightarrow{d} x/c \quad \text{if } c \neq 0$$

- If $x_n \xrightarrow{d} x$ and $g\left(x_n\right)$ is a continuous function, then

$$g\left(x_n\right) \xrightarrow{d} g\left(x\right)$$

- If $y_n \xrightarrow{d} y$ and plim $\left(x_n - y_n\right) = 0$, then $x_n \xrightarrow{d} y$
- If $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$ then $\mathbf{c}'\mathbf{x}_n \xrightarrow{d} \mathbf{c}'\mathbf{x}$

Asymptotic Normality and the Central Limit Theorem

- In principle, if $\text{plim}\ \hat{\theta}_n = \theta$, then $\hat{\theta}_n \xrightarrow{d} \theta$ and the limiting distribution of $\hat{\theta}_n$ is a spike.
- Of course, we don't think that in any given sample this is a reasonable thing to assume.
- Instead, to get more reasonable statistical properties of the estimator, we use a **stabilizing** transformation.

Asymptotic Normality and the Central Limit Theorem

Theorem (Univariate Central Limit Theorem)

If $x_1, x_2, \ldots, x_n$ are a random sample from a distribution with mean $\mu < \infty$ and variance $\sigma^2 < \infty$ and $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, then

$$\sqrt{n}\left(\bar{x}_n - \mu\right) \xrightarrow{d} N\left[0, \sigma^2\right]$$

Theorem (Multivariate Central Limit Theorem)

if $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are a random sample from a multivariate distribution with finite mean vector $\boldsymbol{\mu}$ and finite covariance matrix $\mathbf{Q}$, and $\bar{\mathbf{x}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$, then

$$\sqrt{n}\left(\bar{\mathbf{x}}_n - \boldsymbol{\mu}\right) \xrightarrow{d} N\left[\mathbf{0}, \mathbf{Q}\right]$$

Outline

OLS Asymptotics: Consistency

- We need to modify the assumptions we made when studying OLS in finite samples slightly. We assume:
  1. $(\mathbf{x}_i, \varepsilon_i)$ $i = 1, \ldots, n$ is a sequence of *independent* observations
  2. plim $\frac{\mathbf{X}'\mathbf{X}}{n} = \mathbf{Q}$, a non-singular matrix
- Now rewrite the OLS estimate $\hat{\beta}$ as

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

- So that asymptotically:

$$\text{plim } \hat{\beta} = \beta + \mathbf{Q}^{-1} \text{plim } \left(\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{n}\right)$$

OLS Asymptotics: Consistency

- So we need to find plim $\left(\frac{\mathbf{X}'\varepsilon}{n}\right)$. Notice we can write this as

$$\frac{1}{n}\mathbf{X}'\boldsymbol{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\varepsilon_i = \frac{1}{n}\sum_{i=1}^{n}\mathbf{w}_i = \bar{\mathbf{w}}$$

- So that plim $\hat{\beta}$=$\beta + \mathbf{Q}^{-1}$plim $\bar{\mathbf{w}}$
- To find plim $\bar{\mathbf{w}}$ we need to see what happens to E $[\bar{\mathbf{w}}]$ and Var $[\bar{\mathbf{w}}]$ as $n \to \infty$

$$\mathsf{E}[\mathbf{w}_i] = \mathsf{E}_x[\mathsf{E}[\mathbf{w}_i|\mathbf{x}_i]] = \mathsf{E}_x\left[\mathbf{x}_i \underbrace{\mathsf{E}[\varepsilon_i|\mathbf{x}_i]}_{=0 \text{ (exogeneity assn)}}\right] = 0$$

$$\Rightarrow \mathsf{E}[\bar{\mathbf{w}}] = 0 \quad (< \infty)$$

OLS Asymptotics: Consistency

- Turning to $\text{Var}[\bar{\mathbf{w}}]$:

$$\text{Var}[\bar{\mathbf{w}}] = \text{E}[\text{Var}[\bar{\mathbf{w}}|\mathbf{X}]] + \underbrace{\text{Var}[\text{E}[\bar{\mathbf{w}}|\mathbf{X}]]}_{=0\ (\text{E}[\varepsilon_i|\mathbf{X}]=0)}$$

$$\text{Var}[\bar{\mathbf{w}}|\mathbf{X}] = \text{E}\left[\bar{\mathbf{w}}\bar{\mathbf{w}}'|\mathbf{X}\right] = \frac{1}{n}\mathbf{X}'\text{E}\left[\varepsilon\varepsilon'|\mathbf{X}\right]\mathbf{X}\frac{1}{n}$$

$$= \frac{1}{n}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}\frac{1}{n} = \left(\frac{\sigma^2}{n}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)$$

$$\text{Var}[\bar{\mathbf{w}}] = \left(\frac{\sigma^2}{n}\right)\text{E}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)$$

$$\lim_{n\to\infty}\text{Var}[\bar{\mathbf{w}}] = 0 \times \mathbf{Q} = \mathbf{0}$$

- Therefore $\lim_{n\to\infty}\text{E}[\bar{\mathbf{w}}] = 0 < \infty$ and $\lim_{n\to\infty}\text{Var}[\bar{\mathbf{w}}] = 0$ so $\bar{\mathbf{w}}$ converges in mean square to 0 and

$$\text{plim } \bar{\mathbf{w}} = 0$$

- Putting this all together:

$$\text{plim } \hat{\beta} = \beta + \mathbf{Q}^{-1}\cdot\mathbf{0} = \beta$$

OLS Asymptotic Distribution

- We want to relax the normality assumption, but we still want to know the distribution of $\hat{\beta}$ (at least in large enough samples)
- To do this, we will use the central limit theorem, which requires us to assume that the observations are *independent*
- We will focus on

$$\sqrt{n}\left(\hat{\beta}-\beta\right)=\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{1}{\sqrt{n}}\right)\mathbf{X}'\varepsilon$$

- Using the rules of convergence in distribution, if this has a limiting distribution, it's the same as that of

$$\left[\text{plim}\ \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\right]\left(\frac{1}{\sqrt{n}}\right)\mathbf{X}'\varepsilon=\mathbf{Q}^{-1}\left(\frac{1}{\sqrt{n}}\right)\mathbf{X}'\varepsilon$$

OLS Asymptotic Distribution

- Let's try and find the limiting distribution of

$$\left(\frac{1}{\sqrt{n}}\right)\mathbf{X}'\boldsymbol{\varepsilon} = \sqrt{n}\bar{\mathbf{w}}$$

- $\mathsf{E}\left[\mathbf{w}_i\right] = 0$, what about $\mathsf{Var}\left[\mathbf{w}_i\right]$? $\mathbf{w}_i = \mathbf{x}_i\varepsilon_i$ so

$$\mathsf{Var}\left[\mathbf{x}_i\varepsilon_i\right] = \mathsf{E}\left[\mathbf{x}_i\varepsilon_i^2\mathbf{x}_i'\right] = \sigma^2\mathsf{E}\left[\mathbf{x}_i\mathbf{x}_i'\right] = \sigma^2\mathbf{Q}$$

- Applying the central limit theorem:

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

OLS Asymptotic Distribution

- Putting pieces together

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

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

Theorem (Asympototic distribution of $\hat{\beta}$)

If ε_i are i.i.d. with mean 0 and variance σ^2, then

$$\hat{\beta} \overset{a}{\sim} N\left[\beta, \frac{\sigma^2}{n}\mathbf{Q}^{-1}\right]$$

- In practice, of course, we estimate $(1/n)\,\mathbf{Q}^{-1}$ with $(\mathbf{X}'\mathbf{X})^{-1}$ and σ^2 with $\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k)$

OLS Asymptotic Distribution

- So, we have an (asymptotically) normal distribution for $\hat{\beta}$,
  - but it's not because we assumed the ε_i are normally distributed
  - it's coming from the central limit theorem

OLS Asymptotics: Practice

- Turn to your neighbor: let's work out the asymptotics for OLS in the simplest case:

$$y_i = \beta_0 + \beta_1 x_{i1} + u_i$$

1. Write $\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 & \hat{\beta}_1 \end{pmatrix}' = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$ in terms of the raw observations y_i, x_{i1} and their sample averages $\bar{x}_1$
   1.1 What is $(\mathbf{X}'\mathbf{X})$ in this case?
   1.2 Express $(\mathbf{X}'\mathbf{X})^{-1}$ in terms of $\bar{x}_1$, the $x_{i1} - \bar{x}_1$ and $\sum_{i=1}^{n} x_{i1}^2$
   1.3 Express $\mathbf{X}'\mathbf{y}$ in terms of $\sum_{i=1}^{n} x_i y_i$ and $\bar{y}$
   1.4 Express $\hat{\beta}_1$ in terms of $\bar{x}_1$, $\bar{y}$, $\frac{1}{n}\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2$, and $\frac{1}{n}\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)(y_i - \bar{y})$
2. Write plim $\hat{\beta}_1$ in terms of β_1, Cov (x_1, u_i) and Var (x_1) and show it is $= \beta_1$

OLS Asymptotics: Practice

3. We can write

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) = \left(\frac{1}{n}\sum_{i=1}^{n}\left(x_{i1} - \bar{x}_1\right)^2\right)^{-1} \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_{i1} - \bar{x}_1\right)\left(u_i - \bar{u}\right)\right)$$

   Under what condition is it the case that the limiting distribution of $\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right)$ is the same as that of $\sqrt{n}\text{Var}\left(x_1\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_{i1} - \bar{x}_1\right)\left(u_i - \bar{u}\right)\right)$?
4. Use the Law of Iterated Expectations (LIE) and the exogeneity assumption to find $\text{E}\left[\left(x_{i1} - \bar{x}_1\right)\left(u_i - \bar{u}\right)\right]$
5. Use the LIE and the homoskedasticity assumption to find $\text{Var}\left[\left(x_{i1} - \bar{x}_1\right)\left(u_i - \bar{u}\right)\right]$
6. Using the central limit theorem, what is the limiting distribution of $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_{i1} - \bar{x}_1\right)\left(u_i - \bar{u}\right)\right)$?
7. What is the asymptotic distribution of $\hat{\beta}_1$?

OLS Asymptotics: Practice-Solutions

1.1: To find $\mathbf{X}'\mathbf{X}$, first note that

$$\mathbf{X}_{(n\times 2)} = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n1} \end{pmatrix}$$

so

$$\mathbf{X}'\mathbf{X}_{(2\times 2)} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \end{pmatrix} \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n1} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^n 1 & \sum_{i=1}^n x_{i1} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 \end{pmatrix} = n \begin{pmatrix} 1 & \bar{x}_1 \\ \bar{x}_1 & \frac{1}{n}\sum_{i=1}^n x_{i1}^2 \end{pmatrix}$$

OLS Asymptotics: Practice-Solutions

1.2: Recall that by definition

$$\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \frac{1}{|\mathbf{X}'\mathbf{X}|}\begin{pmatrix} \mathbf{X}'\mathbf{X}_{22} & -\mathbf{X}'\mathbf{X}_{12} \\ -\mathbf{X}'\mathbf{X}_{21} & \mathbf{X}'\mathbf{X}_{11} \end{pmatrix}$$

To find $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ start by finding $|\mathbf{X}'\mathbf{X}|$:

$$\left|\mathbf{X}'\mathbf{X}\right| = n^2\left[\frac{1}{n}\sum_{i=1}^{n} x_{i1}^2 - \left(\bar{x}_1\right)^2\right] = n^2 \times \frac{1}{n}\sum_{i=1}^{n}\left(x_{i1} - \bar{x}_1\right)^2$$

which means that we can write $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ as

$$\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \frac{1}{n}\frac{1}{\frac{1}{n}\sum_{i=1}^{n}\left(x_{i1} - \bar{x}_1\right)^2}\begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i1}^2 & -\bar{x}_1 \\ -\bar{x}_1 & 1 \end{pmatrix}$$

OLS Asymptotics: Practice-Solutions

1.3: Fist note that $\mathbf{y} = \begin{pmatrix} y_1 & y_2 & \cdots & y_n \end{pmatrix}'$ so

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1} y_i \end{pmatrix} = n \begin{pmatrix} \bar{y} \\ \frac{1}{n} \sum_{i=1}^n x_{i1} y_i \end{pmatrix}$$

1.4: Putting the preceding pieces together:

$$\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{y} = \left(\frac{1}{n} \sum_{i=1}^n (x_{i1} - \bar{x}_1)^2\right)^{-1} \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1}^2 & -\bar{x}_1 \\ -\bar{x}_1 & 1 \end{pmatrix} \begin{pmatrix} \bar{y} \\ \frac{1}{n} \sum_{i=1}^n x_{i1} y_i \end{pmatrix}$$

$$= \left(\frac{1}{n} \sum_{i=1}^n (x_{i1} - \bar{x}_1)^2\right)^{-1} \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n x_{i1}^2 \bar{y} - \frac{1}{n} \sum_{i=1}^n x_{i1} y_i \bar{x}_1 \\ \frac{1}{n} \sum_{i=1}^n x_{i1} y_i - \bar{x}_n \bar{y} \end{pmatrix}$$

$$= \left(\frac{1}{n} \sum_{i=1}^n (x_{i1} - \bar{x}_1)^2\right)^{-1} \begin{pmatrix} \bar{y} \frac{1}{n} \sum_{i=1}^n (x_{i1} - \bar{x}_1)^2 - \bar{x}_1 \frac{1}{n} \sum_{i=1}^n (x_{i1} - \bar{x}_1)(y_i - \bar{y}) \\ \frac{1}{n} \sum_{i=1}^n (x_{i1} - \bar{x}_1)(y_i - \bar{y}) \end{pmatrix}$$

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \bar{y} - \hat{\beta}_1 \bar{x}_1 \\ \left(\frac{1}{n} \sum_{i=1}^n (x_{i1} - \bar{x}_1)^2\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n (x_{i1} - \bar{x}_1)(y_i - \bar{y})\right) \end{pmatrix}$$

OLS Asymptotics: Practice-Solutions

2: Insert $y_i = \beta_0 + \beta_1 x_{i1} + u_i$ and $\bar{y} = \beta_0 + \beta_1 \bar{x}_1 + \bar{u}$ into the expression for $\hat{\beta}_1$ from above to get

$$\begin{aligned}
\hat{\beta}_1 &= \left(\frac{1}{n}\sum_{i=1}^{n}\left(x_{i1}-\bar{x}_1\right)^2\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_{i1}-\bar{x}_1\right)\left(\beta_1\left(x_{i1}-\bar{x}\right)+u_i-\bar{u}\right)\right)\\
&= \left(\frac{1}{n}\sum_{i=1}^{n}\left(x_{i1}-\bar{x}_1\right)^2\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\beta_1\left(x_{i1}-\bar{x}_1\right)^2+\frac{1}{n}\sum_{i=1}^{n}\left(x_{i1}-\bar{x}_1\right)\left(u_i-\bar{u}\right)\right)\\
&= \beta_1+\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_{i1}-\bar{x}_1\right)^2\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_{i1}-\bar{x}_1\right)\left(u_i-\bar{u}\right)\right)
\end{aligned}$$

Using the law of large numbers,

$$\text{plim}\ \frac{1}{n}\sum_{i=1}^{n}\left(x_{i1}-\bar{x}_1\right)^2 = \mathsf{E}\left[\left(x_{1i}-\mathsf{E}\left[x_1\right]\right)^2\right] = \mathsf{Var}\left[x_1\right]$$

$$\text{plim}\ \frac{1}{n}\sum_{i=1}^{n}\left(x_{i1}-\bar{x}_1\right)\left(u_i-\bar{u}\right) = \mathsf{E}\left[\left(x_{1i}-\mathsf{E}\left[x_1\right]\right)\left(u_i-\mathsf{E}\left[u\right]\right)\right] = \mathsf{Cov}\left(x_1,u\right)$$

and so

$$\begin{aligned}
\text{plim}\ \hat{\beta}_1 &= \beta_1+\left(\text{plim}\ \frac{1}{n}\sum_{i=1}^{n}\left(x_{i1}-\bar{x}_1\right)^2\right)^{-1}\left(\text{plim}\ \frac{1}{n}\sum_{i=1}^{n}\left(x_{i1}-\bar{x}_1\right)\left(u_i-\bar{u}\right)\right)\\
&= \beta_1+\frac{\mathsf{Cov}\left(x_1,u\right)}{\mathsf{Var}\left[x_1\right]}
\end{aligned}$$

OLS Asymptotics: Practice-Solutions

3. First use the rule of convergence in probability that if plim $x_n = c$, then plim $g(x_n) = g(\text{plim } x_n)$.

- From this we see that if plim $\frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2 = \text{Var}(x_1)$, then

$$\text{plim}\left(\frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2\right)^{-1} = [\text{Var}(x_1)]^{-1}$$

- Next, we use rule of convergence in distribution that if plim $x_n = c$ and $y_n \overset{d}{\to} y$, then $x_n y_n \overset{d}{\to} cy$.
- Sinceplim $\left(\frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2\right)^{-1} = [\text{Var}(x_1)]^{-1}$ the limiting distribution of $\sqrt{n}\left(\hat{\beta}_1-\beta_1\right)$ is the same as that of

$$\sqrt{n}\text{Var}(x_1)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(u_i-\bar{u})\right)$$

OLS Asymptotics: Practice-Solutions

4. By the law of iterated expectations

$$\begin{aligned}\mathsf{E}\left[(x_{i1}-\bar{x}_1)(u_i-\bar{u})\right] &= \mathsf{E}_{x_1}\left[\mathsf{E}\left[(x_{i1}-\bar{x}_1)(u_i-\bar{u})\,|x_1\right]\right]\\ &= \mathsf{E}_{x_1}\left[(x_{i1}-\bar{x}_1)\underbrace{\mathsf{E}\left[(u_i-\bar{u})\,|x_1\right]}_{=0\text{ (exogeneity)}}\right]\\ &= \mathsf{E}_{x_1}\left[(x_{i1}-\bar{x}_1)\times 0\right]=0\end{aligned}$$

5. Using the answer to part 4,

$$\begin{aligned}\mathsf{Var}\left[(x_{i1}-\bar{x}_1)(u_i-\bar{u})\right] &= \mathsf{E}\left[\left[(x_{i1}-\bar{x}_1)(u_i-\bar{u})\right]^2\right]-\left[\mathsf{E}\left[(x_{i1}-\bar{x}_1)(u_i-\bar{u})\right]\right]^2\\ &= \mathsf{E}\left[\left[(x_{i1}-\bar{x}_1)(u_i-\bar{u})\right]^2\right]\end{aligned}$$

and using the LIE

$$\begin{aligned}\mathsf{E}\left[\left[(x_{i1}-\bar{x}_1)(u_i-\bar{u})\right]^2\right] &= \mathsf{E}_{x_1}\left[\mathsf{E}\left[\left[(x_{i1}-\bar{x}_1)(u_i-\bar{u})\right]^2|x_1\right]\right]\\ &= \mathsf{E}_{x_1}\left[(x_{i1}-\bar{x}_1)^2\,\mathsf{E}\left[(u_i-\bar{u})^2\,|x_1\right]\right]\\ &= \mathsf{E}_{x_1}\left[(x_{i1}-\bar{x}_1)^2\,\mathsf{Var}\left(u_i|x_1\right)\right]\end{aligned}$$

OLS Asymptotics: Practice-Solutions

by the homoskedasticity assumption $\text{Var}\,(u_i|x_1) = \sigma^2$ so

$$\begin{aligned}\mathsf{E}\left[\left[(x_{i1}-\bar{x}_1)(u_i-\bar{u})\right]^2\right] &= \mathsf{E}_{x_1}\left[(x_{i1}-\bar{x}_1)^2\sigma^2\right] \\ &= \sigma^2\mathsf{E}_{x_1}\left[(x_{i1}-\bar{x}_1)^2\right] \\ &= \sigma^2\text{Var}\,[x_1]\end{aligned}$$

6. Parts 4 and 5 showed that $(x_{i1}-\bar{x}_1)(u_i-\bar{u})$ has a finite mean and variance, so we can apply the central limit theorem to see that

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(u_i-\bar{u})\right) \xrightarrow{d} N\left[0, \sigma^2\text{Var}\,[x_1]\right]$$

7. Using the answers to parts 3–6,
$\sqrt{n}\left(\hat{\beta}_1-\beta_1\right) \xrightarrow{d} N\left[0, \sigma^2\left(\text{Var}\,[x_1]\right)^{-1}\right]$ and so

$$\hat{\beta}_1 \xrightarrow{d} N\left[\beta_1, \frac{\sigma^2}{n}\frac{1}{\text{Var}\,(x_1)}\right]$$

Outline

Asymptotic Inference with OLS: The t statistic

- Above, we showed that

$$\hat{\beta} \xrightarrow{d} N\left[\beta, \frac{\sigma^2}{n}\mathbf{Q}^{-1}\right]$$

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

- Of course, we don't know σ^2, so we need an estimator of it. We will use

$$s^2 = \frac{1}{n-k}\hat{\mathbf{u}}'\hat{\mathbf{u}}$$

where $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\beta} = \mathbf{X}\left(\beta - \hat{\beta}\right) + \mathbf{u}.$

- First rewrite this as

$$s^2 = \frac{n}{n-k}\left(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n}\right)$$

Asymptotic Inference with OLS: The t statistic

- Then break open $\hat{\mathbf{u}}'\hat{\mathbf{u}}$

$$\begin{aligned}\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n} &= \frac{1}{n}\left[\left(\beta-\hat{\beta}\right)'\mathbf{X}'+\mathbf{u}'\right]\left[\mathbf{X}\left(\beta-\hat{\beta}\right)+\mathbf{u}\right] \\ &= \frac{1}{n}\left[\left(\beta-\hat{\beta}\right)'\mathbf{X}'\mathbf{X}\left(\beta-\hat{\beta}\right)+\left(\beta-\hat{\beta}\right)'\mathbf{X}'\mathbf{u}+\mathbf{u}'\mathbf{X}\left(\beta-\hat{\beta}\right)+\mathbf{u}'\mathbf{u}\right]\end{aligned}$$

- Since plim $\left(\beta-\hat{\beta}\right)=0$, the plims of the first 3 terms are 0, and so

$$\text{plim } \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n} = \text{plim } \frac{\mathbf{u}'\mathbf{u}}{n}$$

- $\mathbf{u}'\mathbf{u}/n = \frac{1}{n}\sum_{i=1}^{n} u_i^2$ and so by the law of large numbers,

$$\text{plim } \frac{\mathbf{u}'\mathbf{u}}{n} = \mathsf{E}\left[u_i^2\right] = \mathsf{Var}\left[u_i\right] = \sigma$$

- Combine this with the fact that $n/\left(n-k\right)\to 1$ as $n\to\infty$ and we see that

$$\text{plim } s^2 = \sigma^2$$

Asymptotic Inference with OLS: The t statistic

- Rewrite our familiar t statistic as

$$t_k = \frac{\sqrt{n}\left(\hat{\beta}_k - \beta_k^0\right)}{\sqrt{s^2\left(\mathbf{X}'\mathbf{X}/n\right)_{kk}^{-1}}}$$

- Recall that if we assume $\varepsilon_i \sim N\left(0, \sigma^2\right)$ then $t_k \sim t\left[n-k\right]$
- Using the above results, the denominator has plim $\sqrt{s^2\left(\mathbf{X}'\mathbf{X}/n\right)_{kk}^{-1}} = \sqrt{\sigma^2 \mathbf{Q}_{kk}^{-1}}$
- And under the null hypothesis $\left(\beta_k = \beta_k^0\right)$ the numerator converges in distribution to $N\left[0, \sigma^2 \mathbf{Q}_{kk}^{-1}\right]$.
- Therefore, combining these we see that

$$t_k \xrightarrow{d} N\left[0, 1\right]$$

Asymptotic Inference with OLS: The F statistic

- What about testing multiple (linear) hypotheses? To test the set of J hypotheses $\mathbf{R}\beta - \mathbf{q} = \mathbf{0}$ we study the asymptotic distribution oft the Wald Statistic

$$W = JF = \left(\mathbf{R}\hat{\beta} - \mathbf{q}\right)' \left[\mathbf{R}s^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{R}'\right]^{-1} \left(\mathbf{R}\hat{\beta} - \mathbf{q}\right)$$

- If $\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left[\mathbf{0}, \sigma^2 \mathbf{Q}^{-1}\right]$, and if the null hypothesis $(\mathbf{R}\beta - \mathbf{q} - \mathbf{0})$ is true, then

$$\sqrt{n}\mathbf{R}\left(\hat{\beta} - \beta\right) = \sqrt{n}\left(\mathbf{R}\hat{\beta} - \mathbf{q}\right) \xrightarrow{d} N\left[\mathbf{0}, \mathbf{R}\left(\sigma^2 \mathbf{Q}^{-1}\right)\mathbf{R}'\right]$$

- As shorthand, let $\mathbf{z} = \sqrt{n}\left(\mathbf{R}\hat{\beta} - \mathbf{q}\right)$ and $\mathbf{P} = \mathbf{R}\left(\sigma^2 \mathbf{Q}^{-1}\right)\mathbf{R}'$, so we can restate this as $\mathbf{z} \xrightarrow{d} N\left[\mathbf{0}, \mathbf{P}\right]$

Asymptotic Inference with OLS: The F statistic

- The "inverse square root" of a matrix $\mathbf{P}$ is another matrix $\mathbf{T}$ such that $\mathbf{T}^2 = \mathbf{P}^{-1}$. We call this matrix $\mathbf{P}^{-1/2}$
- Since $\mathbf{z} \overset{d}{\to} N[\mathbf{0}, \mathbf{P}]$,

$$\mathbf{P}^{-1/2}\mathbf{z} \overset{d}{\to} N\left[\mathbf{0}, \mathbf{P}^{-1/2}\mathbf{P}\mathbf{P}^{-1/2}\right] = N[\mathbf{0}, \mathbf{I}]$$

- Recall that if $x_n \overset{d}{\to} x$ then $g(x_n) \overset{d}{\to} g(x)$, and that if $y_i \sim N(0,1)$, then $\sum_{i=1}^{J} y_i^2 \sim \chi^2(J)$. Together, these imply that

$$\left(\mathbf{P}^{-1/2}\mathbf{z}\right)'\left(\mathbf{P}^{-1/2}\mathbf{z}\right) = \mathbf{z}'\mathbf{P}\mathbf{z} \overset{d}{\to} \chi^2(J)$$

Asymptotic Inference with OLS: The F statistic

- Putting all of this together, we have shown that

$$n\left(\mathbf{R}\hat{\beta}-\mathbf{q}\right)'\left[\mathbf{R}\left(\sigma^2\mathbf{Q}^{-1}\right)\mathbf{R}'\right]^{-1}\left(\mathbf{R}\hat{\beta}-\mathbf{q}\right)\xrightarrow{d}\chi^2\left(J\right)$$

- Since plim $s^2\left(\mathbf{X}'\mathbf{X}/n\right)^{-1}=\sigma^2\mathbf{Q}^{-1}$, it is also the case that the Wald Statistic

$$W\xrightarrow{d}\chi^2\left(J\right)$$

Outline

Measurement error

- The most controversial assumption of OLS is *exogeneity*: $\mathsf{E}[\varepsilon_i|\mathbf{X}] = 0$
- Let's think through the consequences of measurement error for the exogeneity assumption and the OLS estimator
- There are 2 basic forms of measurement error:
  1. Measurement error in the outcome variable y_i
  2. Measurement error in one (or more) dependent variable

Measurement error in the dependent variable

- Imagine that the true model is that

$$y_i^* = \mathbf{x}_i'\beta + \varepsilon_i$$

- However, we don't have data on y_i^*, instead we observe $y_i = y_i^* + w_i$ where w_i is uncorrelated with $\mathbf{x}_i$ and ε_i and has $\mathrm{E}[w_i] = 0$. Then we can write this model as

$$\begin{aligned} y_i &= \mathbf{x}_i'\beta + \varepsilon_i + w_i \\ &= \mathbf{x}_i'\beta + v_i \end{aligned}$$

- In this model, if $\mathrm{E}[\varepsilon_i|\mathbf{X}] = 0$ then $\mathrm{E}[v_i|\mathbf{X}] = 0$ also. The exogeneity assumption is still satisfied.

Measurement error in the dependent variable

- With measurement error in the dependent variable, the OLS estimator is still unbiased:

$$\begin{aligned}\text{plim } \hat{\beta} &= \beta + \mathbf{Q}^{-1}\text{plim } \left(\frac{\mathbf{X}'\mathbf{v}}{n}\right) \\ &= \beta + \mathbf{Q}^{-1}\left(\text{plim } \left(\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{n} + \frac{\mathbf{X}'\mathbf{w}}{n}\right)\right) \\ &= \beta\end{aligned}$$

- But, the OLS estimator becomes noisier, it's asymptotic variance is

$$\begin{aligned}\text{Avar}\left[\hat{\beta}\right] &= \frac{\text{Var}\left[v\right]}{n}\mathbf{Q}^{-1} = \frac{\text{Var}\left[\varepsilon + w\right]}{n}\mathbf{Q}^{-1} \\ &= \frac{\sigma_\varepsilon^2 + \sigma_w^2}{n}\mathbf{Q}^{-1}\end{aligned}$$

Measurement error in an independent variable

- If there is measurement error in the explanatory variables we are in much more trouble.
- Start with the case of a single explanatory variable: Suppose the true model is

$$y_i = \beta_0 + \beta_1 x_{1i}^* + \varepsilon_i$$

- But we only have data on $x_{1i} = x_{1i}^* + e_{1i}$ so that we can rewrite the model as

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{1i} + (\varepsilon_i - \beta_1 e_{1i}) \\ &= \beta_0 + \beta_1 x_{1i} + v_i \end{aligned}$$

- And clearly $\mathrm{E}\left[v_i | x_{1i}\right] \neq 0$ since they both contain e_{1i}
- So what happens to the OLS estimator?

Measurement error in an independent variable

- The **classical errors in variables** assumption is that

$$\text{Cov}\left(x_1^*, e_1\right) \qquad \text{E}\left[e_1\right] = 0$$

- We can derive the inconsistency in the OLS estimate $\hat{\beta}_1$:

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}\left(x_1, v\right)}{\text{Var}\left(x_1\right)}$$

- Starting with the denominator:

$$\begin{aligned} \text{Var}\left(x_1\right) &= \text{Var}\left(x_1^* + e_1\right) = \text{Var}\left(x_1^*\right) + \text{Var}\left(e_1\right) + 2\underbrace{\text{Cov}\left(x_1^*, e_1\right)}_{=0} \\ &= \sigma^2_{x_1^*} + \sigma^2_{e_1} \end{aligned}$$

- And the numerator:

$$\text{Cov}\left(x_1, v\right) = \text{Cov}\left(x_1^* + e_1, \varepsilon_i - \beta_1 e_1\right) = \text{Cov}\left(e_1, -\beta_1 e_1\right) = \underbrace{-\beta_1 \sigma^2_{e_1}}_{\neq 0!}$$

Measurement error in an independent variable

- Putting this all together

$$\text{plim } \hat{\beta}_1 = \beta_1 - \frac{\beta_1 \sigma^2_{e_1}}{\sigma^2_{x_1^*} + \sigma^2_{e_1}} = \beta_1 \left[1 - \frac{\sigma^2_{e_1}}{\sigma^2_{x_1^*} + \sigma^2_{e_1}} \right]$$

$$= \beta_1 \frac{\sigma^2_{x_1^*}}{\sigma^2_{x_1^*} + \sigma^2_{e_1}}$$

- Since $0 < \frac{\sigma^2_{x_1^*}}{\sigma^2_{x_1^*} + \sigma^2_{e_1}} < 1$, $\hat{\beta}_1$ is biased towards 0: **attenuation bias**
- The amount of attenuation bias depends on $\sigma^2_{e_1} / \sigma^2_{x_1^*}$:
  - if $\sigma^2_{x_1^*}$ is large relative to $\sigma^2_{e_1}$ then the attenuation bias is small
  - If $\sigma^2_{x_1^*}$ is small relative to $\sigma^2_{e_1}$, then the attenuation bias is large

Measurement error with several independent variables

- Let us generalize the above reasoning:

$$\mathbf{y} = \mathbf{X}^*\beta + \varepsilon \quad \mathbf{X} = \mathbf{X}^* + \mathbf{U}$$

- Assume that $\mathsf{E}\left[\mathbf{X}^{*\prime}\mathbf{U}\right] = \mathbf{0}$ and $\mathsf{E}\left[\mathbf{U}\right] = \mathbf{0}$ (classical measurement error)
- We estimate the model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{V} \quad \mathbf{V} = \varepsilon - \mathbf{U}\beta$$

- The analogs of the above are:

$$\begin{aligned} \text{plim}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right) &= \text{plim}\left(\frac{\left(\mathbf{X}^* + \mathbf{U}\right)'\left(\mathbf{X}^* + \mathbf{U}\right)}{n}\right) \\ &= \mathbf{Q}^* + \mathbf{\Sigma}_{uu} \end{aligned}$$

where

$$\mathbf{Q}^* = \mathsf{E}\left[\mathbf{X}'\mathbf{X}\right] \quad \mathbf{\Sigma}_{uu} = \mathsf{E}\left[\mathbf{U}'\mathbf{U}\right]$$

Measurement error with several independent variables

- And,

$$\text{plim}\left(\frac{\mathbf{X}'\mathbf{V}}{n}\right) = \text{plim}\left(\frac{(\mathbf{X}^* + \mathbf{U})'(\boldsymbol{\varepsilon} - \mathbf{U}\beta)}{n}\right) = -\boldsymbol{\Sigma}_{uu}\beta$$

- As a result,

$$\begin{aligned}
\text{plim}\ \hat{\beta} &= \beta + \left[\text{plim}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)\right]^{-1} \text{plim}\left(\frac{\mathbf{X}'\mathbf{V}}{n}\right) \\
&= \beta - (\mathbf{Q}^* + \boldsymbol{\Sigma}_{uu})^{-1}\boldsymbol{\Sigma}_{uu}\beta \\
&= (\mathbf{Q}^* + \boldsymbol{\Sigma}_{uu})^{-1}\mathbf{Q}^*\beta
\end{aligned}$$

- In general, all bets are off. All the coefficients are biased, and signing them is hard.

Measurement error with several independent variables

- What if only 1 of the variables is measured with error? Is that coefficient attenuated but the rest are fine?
- This is like saying that

$$\mathbf{\Sigma}_{uu} = \begin{bmatrix} \sigma_u^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

- It can be shown that now

$$\text{plim } \hat{\beta}_1 = \frac{\beta_1}{1 + \sigma_u^2 q_{11}^*}$$

where q_{11}^* is the $(1,1)^{th}$ element of $\mathbf{Q}^*$. There is **attenuation bias** in the coefficient on the mismeasured variable
- It can also be shown that for $k \neq 1$

$$\text{plim } \hat{\beta}_k = \beta_k - \beta_1 \left[\frac{\sigma_u^2 q_{k1}^*}{1 + \sigma_u^2 q_{11}^*} \right]$$

where q_{k1}^* is the $(k,1)^{th}$ element of $\mathbf{Q}^*$. The inconsistency here has *unknown sign*, but is not, in general 0.

Measurement Error: Summing up

- In a regression in which only 1 variable is measured with error:
  - the coefficient on that variable is attenuated (biased towards 0)
  - all the other coefficients are biased too, but in unknown directions
- In a regression in which more than 1 variable is measured with error:
  - all the coefficients are biased in unknown directions:
    All bets are off.

Outline

Omitted Variable Bias again

- Let's partition the $\mathbf{X}$ matrix into two parts $\mathbf{X}_1$ and $\mathbf{X}_2$: $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{bmatrix}$
- Revisiting omitted variables: If the real model is $\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon$, but we omit $\mathbf{X}_2$

$$\hat{\beta}_1 = \beta_1 + \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2 + \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\varepsilon$$

- Asymptotically:

$$\begin{aligned}\text{plim } \hat{\beta}_1 &= \beta_1 + \left[\text{plim }\left(\frac{\mathbf{X}_1'\mathbf{X}_1}{n}\right)\right]^{-1}\text{plim }\left(\frac{\mathbf{X}_1'\mathbf{X}_2}{n}\right)\beta_2 + \left[\text{plim }\left(\frac{\mathbf{X}_1'\mathbf{X}_1}{n}\right)\right]^{-1}\text{plim }\left(\frac{\mathbf{X}_1'\varepsilon}{n}\right) \\ &= \beta_1 + \mathbf{Q}^{-1}\mathsf{E}\left[\mathbf{X}_1'\mathbf{X}_2\right]\beta_2\end{aligned}$$

- So even asymptotically, OLS is inconsistent unless
  - $\mathsf{E}\left[\mathbf{X}_1'\mathbf{X}_2\right] = \mathbf{0}$ ($\mathbf{X}_1$ and $\mathbf{X}_2$ uncorrelated)
  - $\beta_2 = \mathbf{0}$ (the true model doesn't include $\mathbf{X}_2$

Omitted Variable Bias: Practice

- Turn to your neighbor: Let's consider the simplest possible case: The true model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \tag{1}$$

but instead we run the OLS regression ignoring x_{2i}:

$$y_i = \beta_0 + \beta_1 x_{1i} + v_i \tag{2}$$

1. Recall from above that in this case we can write

$$\hat{\beta}_1 = \left[\frac{1}{n}\sum_{i=1}^{n}\left(x_{1i} - \bar{x}_1\right)^2\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_{1i} - \bar{x}_1\right)\left(y_i - \bar{y}\right)\right]$$

   Substitute in (1) to express $\hat{\beta}_1$ in terms of β_1, β_2, the x_{1i}, x_{2i} and ε_i and $\bar{x}_1$ and $\bar{x}_2$
2. Apply the law of large numbers to the result to find plim $\hat{\beta}_1$

Omitted Variable Bias: Practice

3. Recall that the plim of the OLS estimator for β_1 in (2) can be written as

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}\,(x_1, v)}{\text{Var}\,(x_1)}$$

   3.1. Write v_i in terms of objects in equation (1)
   3.2. Substitute this into the expression for plim $\hat{\beta}_1$ to find plim $\hat{\beta}_1$ in terms of objects in equation (1)

4. Imagine that y is (log) earnings, $x_1$is years of schooling and x_2 is intelligence.
   4.1. What is the likely sign of Cov (x_1, x_2)?
   4.2. What is the likely sign of β_2?
   4.3. What is the likely sign of the omitted variable bias?

5. In our previous example y is health status, $x_1$is going to hospital and x_2 is y_{0i}, people's latent health status.
   5.1. What is the likely sign of Cov (x_1, x_2)?
   5.2. What is the likely sign of β_2?
   5.3. What is the likely sign of the omitted variable bias?

Omitted Variable Bias: Practice

1. Let's focus first on the numerator:

$$\frac{1}{n}\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)\left(y_i-\bar{y}\right)=\frac{1}{n}\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)\Big[\left(\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\varepsilon_i\right)$$

$$\left(\beta_0+\beta_1\bar{x}_1+\beta_2\bar{x}_2+\bar{\varepsilon}\right)\Big]$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)\left(\beta_1\left(x_{1i}-\bar{x}_1\right)+\beta_2\left(x_{2i}-\bar{x}_2\right)+\left(\varepsilon_i-\bar{\varepsilon}\right)\right.$$

$$=\beta_1\frac{1}{n}\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)^2+\beta_2\frac{1}{n}\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)\left(x_{2i}-\bar{x}_2\right)$$

$$+\frac{1}{n}\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)\left(\varepsilon_i-\bar{\varepsilon}\right)$$

Combining this with the denominator:

$$\hat{\beta}_1=\beta_1+\beta_2\frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)\left(x_{2i}-\bar{x}_2\right)}{\frac{1}{n}\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)^2}+\frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)\left(\varepsilon_i-\bar{\varepsilon}\right)}{\frac{1}{n}\sum_{i=1}^{n}\left(x_{1i}-\bar{x}_1\right)^2}$$

Omitted Variable Bias: Practice

2. Applying the law of large numbers to each part:

$$\text{plim } \frac{1}{n}\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2 = \mathsf{E}\left[(x_{1i}-\mathsf{E}[x_1])^2\right] = \mathsf{Var}[x_1]$$

$$\text{plim } \frac{1}{n}\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2) = \mathsf{E}[(x_{1i}-\mathsf{E}[x_1])(x_{2i}-\mathsf{E}[x_2])] = \mathsf{Cov}(x_1,x_2)$$

$$\text{plim } \frac{1}{n}\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(\varepsilon_i-\bar{\varepsilon}) = \mathsf{E}[(x_{1i}-\mathsf{E}[x_1])(\varepsilon_i-\mathsf{E}[\varepsilon])] = \mathsf{Cov}(x_1,\varepsilon)$$

Combining these,

$$\begin{aligned}\text{plim } \hat{\beta}_1 &= \beta_1 + \beta_2\frac{\mathsf{Cov}(x_1,x_2)}{\mathsf{Var}(x_1)} + \frac{\mathsf{Cov}(x_1,\varepsilon)}{\mathsf{Var}(x_1)} \\ &= \beta_1 + \beta_2\frac{\mathsf{Cov}(x_1,x_2)}{\mathsf{Var}(x_1)}\end{aligned}$$

Omitted Variable Bias: Practice

3. Recall that the plim of the OLS estimator for β_1 in (2) can be written as

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}\,(x_1, v)}{\text{Var}\,(x_1)}$$

   3.1. Write v_i in terms of objects in equation (1)

$$v_i = \varepsilon_i + \beta_2 x_{2i}$$

   3.2. Substitute this into the expression for plim $\hat{\beta}_1$ to find plim $\hat{\beta}_1$ in terms of objects in equation (1)

$$\begin{aligned} \text{plim } \hat{\beta}_1 &= \beta_1 + \frac{\text{Cov}\,(x_1, \varepsilon + \beta_2 x_2)}{\text{Var}\,(x_1)} \\ &= \beta_1 + \frac{\text{Cov}\,(x_1, \varepsilon)}{\text{Var}\,(x_1)} + \frac{\text{Cov}\,(x_1, \beta_2 x_2)}{\text{Var}\,(x_1)} \\ &= \beta_1 + \beta_2 \frac{\text{Cov}\,(x_1, x_2)}{\text{Var}\,(x_1)} \end{aligned}$$

Omitted Variable Bias: Practice

4. Imagine that y is (log) earnings, $x_1$is years of schooling and x_2 is intelligence.
    4.1. What is the likely sign of Cov (x_1, x_2)?: > 0
    4.2. What is the likely sign of β_2? > 0
    4.3. What is the likely sign of the omitted variable bias? > 0
5. In our previous example y is health status, $x_1$is going to hospital and x_2 is y_{0i}, people's latent health status.
    5.1. What is the likely sign of Cov (x_1, x_2)? < 0
    5.2. What is the likely sign of β_2? > 0
    5.3. What is the likely sign of the omitted variable bias? < 0

Outline

Bad Control

- Given the difficulty of ruling out omitted variable bias, you might think *the more controls the better*, right?
- Sadly no, there are 2 kinds of ways this can go wrong: **Bad Controls**

1. Putting variables that are themselves affected by the variable of interest on the right hand side of the regression.
2. Putting bad proxies for unmeasured variables on the right hand side of the regression.

Bad Control 1: Outcome Variables on the RHS

- Imagine the following exercise:
  - We are trying to estimate the effect of going to college on earnings
  - People can work in 2 occupations: Blue Collar and White Collar
  - Clearly occupation is correlated by both college and earnings, so should it be a control?
  - Problem: College affects both occupational choice, and earnings
- Let's use our potential outcomes framework to study this

$$y_i = C_i y_{1i} + (1 - C_i)\, y_{0i}$$
$$W_i = C_i W_{1i} + (1 - C_i)\, W_{0i}$$

where $C_i = 1$ if go to college (0 otherwise) and $W_i = 1$ if white collar (0 if blue collar) and $y_{1i}, y_{0i}, W_{1i}, W_{0i}$ are the potential outcomes.

Bad Control 1: Outcome Variables on the RHS

- Let's assume C_i is randomly assigned, so that it is independent of all the potential outcomes: $\{y_{1i}, y_{0i}, W_{1i}, W_{0i}\} \perp C_i$
- So, we can estimate the effect of college on earnings and occupational choice no problem:

$$\mathsf{E}\left[y_i|C_i = 1\right] - \mathsf{E}\left[y_i|C_i = 0\right] = \mathsf{E}\left[y_{1i} - y_{0i}\right]$$
$$\mathsf{E}\left[W_i|C_i = 1\right] - \mathsf{E}\left[W_i|C_i = 0\right] = \mathsf{E}\left[W_{1i} - W_{0i}\right]$$

- The problem is that the comparison of earnings conditional on W_i is *not* the causal effect of college conditional on occupation because of a selection problem.
- College changes the composition of people in each occupation: College affects white collar earnings, but it also affects *who* becomes a white collar worker

Bad Control 1: Outcome Variables on the RHS

- Imagine regressing y_i on C_i in the subsample of white collar workers:

$$\begin{aligned}&\mathsf{E}\left[y_i|W_i=1,C_i=1\right]-\mathsf{E}\left[y_i|W_i=1,C_i=0\right]\\&\quad=\mathsf{E}\left[y_{1i}|W_i=1,C_i=1\right]-\mathsf{E}\left[y_{0i}|W_i=1,C_i=0\right]\end{aligned}$$

- Since C_i is randomly assigned and independent of the potential outcomes

$$\begin{aligned}&\mathsf{E}\left[y_{1i}|W_i=1,C_i=1\right]-\mathsf{E}\left[y_{0i}|W_i=1,C_i=0\right]\\&\quad=\mathsf{E}\left[y_{1i}|W_{1i}=1\right]-\mathsf{E}\left[y_{0i}|W_{0i}=1\right]\\&\quad=\underbrace{\mathsf{E}\left[y_{1i}-y_{0i}|W_{1i}=1\right]}_{\text{causal effect}}+\underbrace{\mathsf{E}\left[y_{0i}|W_{1i}=1\right]-\mathsf{E}\left[y_{0i}|W_{0i}=1\right]}_{\text{selection bias}}\end{aligned}$$

Bad Control 2: Bad Proxies for Unobserved RHS Variables

- Again, let's take a concrete example: Let's say you want to measure the effect of schooling S_i on earnings y_i:

$$y_i = \alpha + \rho S_i + \gamma a_i + \varepsilon_i$$

- If we don't conrol at all for ability a_i we have omitted variable bias:

$$\hat{\rho} = \rho + \frac{\text{Cov}\left(S, a\right)}{\text{Var}\left(S\right)}\gamma$$

- Imagine we had the scores on an IQ test at age 14, before people make any schooling choices (assume everyone completes 8th grade): a_{ei}
  - then controling for a_{ei} fixes the problem $\text{E}\left[S_i \varepsilon_i\right] = \text{E}\left[a_{ei} \varepsilon_i\right] = 0$

Bad Control 2: Bad Proxies for Unobserved RHS Variables

- These kinds of measures are very hard to come by though.
- Imagine instead that you had test scores on a test employers use to screen applicants a_{li}
  - The problem is that this ability measure is measured after schooling choices have been made
  - If the measure is affected by schooling, then we have a problem:

$$a_{li} = \pi_0 + \pi_1 S_i + \pi_2 a_i$$

- Substituting out a_i we see that

$$y_i = \left(\alpha - \gamma \frac{\pi_0}{\pi_2}\right) + \left(\rho - \gamma \frac{\pi_1}{\pi_2}\right) S_i + \frac{\gamma}{\pi_2} a_{li} + \varepsilon_i$$

Bad Control 2: Bad Proxies for Unobserved RHS Variables

- So what can we do?
- In this example $\gamma > 0$, $\pi_1 > 0$, and $\pi_2 > 0$ so $\rho - \gamma\frac{\pi_1}{\pi_2} < \rho$
- We can regress a_{li} on S_i to get a sense of how large π_1 is likely to be. If π_1 is small, maybe not too much of a problem
- Also, note that
  - regression without ability measure overestimates ρ
  - regression controlling for a_{li} underestimates ρ
  - so we can put bounds on ρ