ECNS 561
OLS Asymptotics

Consistency

- Up to this point, we have only concerned ourselves with the finite sample properties of OLS

- Unbiasedness holds for any N

- $\widehat{\beta}_j$ is **consistent** if the distribution around $\widehat{\beta}_j$ becomes more and more tightly distributed around β_j as N gets large.

- What this means is that, if we collect "enough" data, we can get $\widehat{\beta}_j$ arbitrarily close to β_j.

- Fortunately, our first four Gauss-Markov assumptions, in addition to implying unbiasedness, also imply consistency

- Just to review
  - **MLR.1** The population model is linear in parameters
  - **MLR.2** We have a random sample of n observations, following the population model in assumption MLR.1
  - **MLR.3** No perfect collinearity. In the sample (and therefore in the population), none of the independent variables are constant, and there are no exact linear relationships among the independent variables
  - **MLR.4** Zero conditional mean. The error ε has an expected value of zero given any values of the independent variables

$$E[\varepsilon|x_1, x_2, \ldots, x_k] = 0$$

  If this assumption holds, we have **exogenous explanatory variables.**

Consistency of OLS Theorem

"Under assumptions MLR.1-MLR.4, the OLS estimator $\widehat{\beta}_j$ is consistent for β_j, for all $j = 0, 1, \ldots, k$."
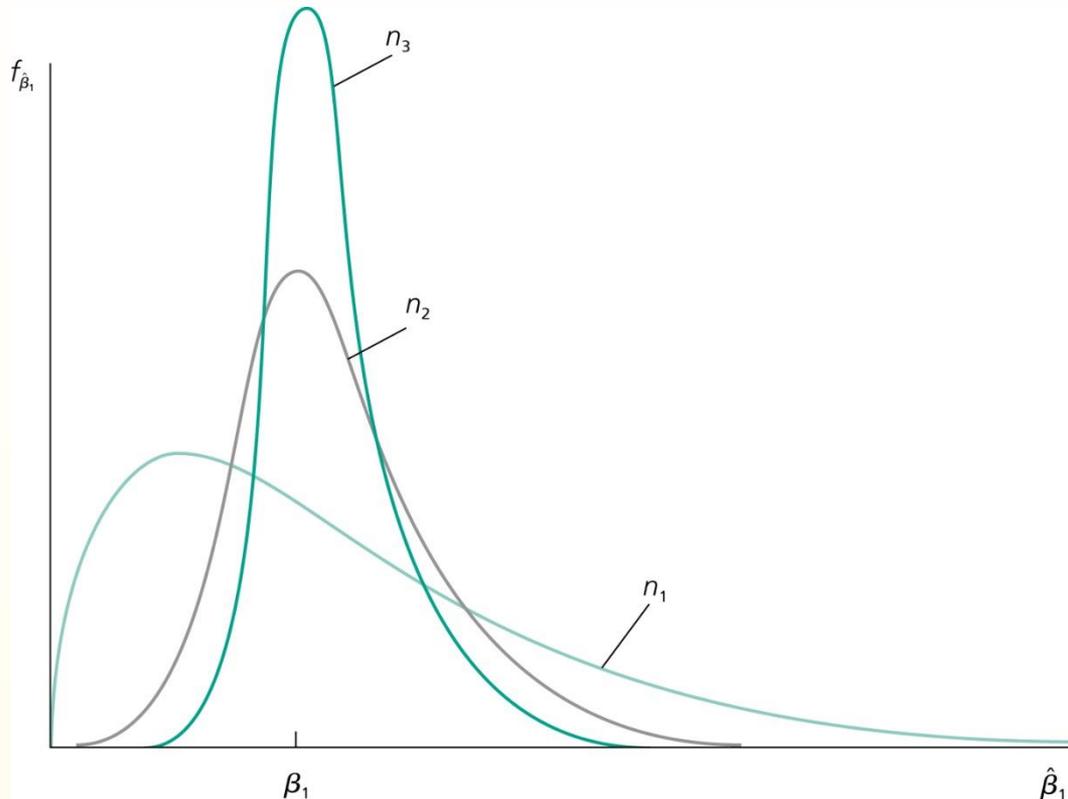
Work through OLS
is consistent proof

Inconsistency

- Correlation between the error and any of the covariates leads to all of the OLS estimators to be inconsistent.

- Moreover, recall that if $E[\varepsilon|x_1, \ldots, x_k] = 0$ fails, then OLS is biased.

- So, we have the following

  "If the error is correlated with any of the independent variables, then OLS is biased and inconsistent."

- In the simple regression case, we can express inconsistency as

  $\text{plim}\widehat{\beta_1} - \beta_1 = \text{Cov}(x_1, \varepsilon)/\text{Var}(x_1)$

- Because $\text{Var}(x_1) > 0$,

  - Inconsistency in $\widehat{\beta_1}$ is positive if x_1 and ε are positively correlated
  - Inconsistency in $\widehat{\beta_1}$ is negative if x_1 and ε are negatively correlated

- If covariance between x_1 and ε is small relative to $\text{Var}(x_1)$, then inconsistency can be negligible.

  - Q. But, why can we not observe how big the covariance between x_1 and ε is?
  - Ans. Because we do not observe ε

- We can derive the asymptotic analog of omitted variable bias. Suppose the true model is as follows

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \upsilon,$$

and follows our first four Gauss-Markov assumptions. So, υ has a zero mean and is uncorrelated with x_1 and x_2. And, $\widehat{\beta_0}$, $\widehat{\beta_1}$, and $\widehat{\beta_2}$ are consistent.

- If we omit x_2 from the regression, then

$$y = \beta_0 + \beta_1 x_1 + \mu, \text{ where } \mu = \beta_2 x_2 + \upsilon.$$

- Let $\widetilde{\beta_1}$ denote the slope estimate from this regression. We can express the plim of this estimator as

$$\text{plim}\,\widetilde{\beta_1} = \beta_1 + \beta_2 \delta_1$$

where $\delta_1 = \text{Cov}(x_1, x_2)/\text{Var}(x_1)$. (recall our "auxiliary" regression)

- A difference between inconsistency and bias is that we express inconsistency in terms of the population variance of x_1 and the population covariance of x_1 and x_2, whereas the bias is based on their **sample** counterparts.
  - Recall that $\text{Bias}(\widetilde{\beta_1}) = E(\widetilde{\beta_1}) - \beta_1 = \beta_2 \widetilde{\delta_1}$

- If the covariance between x_1 and x_2 is small relative to the variance of x_1, the inconsistency can be small

Asymptotic Normality

- We need more than consistency in order to test hypotheses about the parameters of our model

- We also require the sampling distribution of the OLS estimators

- Under our classical linear model assumptions, MLR.1-MLR.6, we know that the sampling distributions are normal.

- Just to review
  - **MLR.5**. Homoskedasticity. The error ε has the same variance given any values of the explanatory variables. That is,
  $$\text{Var}(\varepsilon|x_1, \ldots, x_k) = \sigma^2$$
  - **MLR.6**. The population error ε is independent of the covariates $x_1, x_2, \ldots, x_k$ and is normally distributed with zero mean and variance σ^2: $\varepsilon \sim \text{Normal}(0, \sigma^2)$.

- This was the basis for deriving the t and F distributions we used for hypothesis testing.

- Remember, we showed that the normality of the OLS estimators comes from the assumption that the distribution of the population error is normal

- In practice, if the errors ε_1, ε_2, …, ε_n are random draws from non-normal distributions, then the $\widehat{\beta}_j$ will not be normally distributed.
  - So, the t stats will not have t distributions and the F stats will not have F distributions

- Assumption MLR.6, is equivalent to saying that the distribution of y given x_1, x_2, …, x_k is normal.
  - Because y is observed and ε is not, it is easier to think about whether or not y is normal
  - We can think of many examples where y is not going to be normally distributed (e.g., violent crime)

- In terms of hypothesis testing using t and F stats, are we simply out of luck if y is not normally distributed?

- CLT to the rescue!
  - We can apply the central limit theorem (see our Math Stats notes for a review) to conclude that the OLS estimators are asymptotically normally distributed

Asymptotic Normality of OLS

-see proof in appendix for more details

- From the theorem on asymptotic normality, the standardized estimator has a Normal(0,1) asymptotic distribution

$$\frac{\widehat{\beta_j} - \beta_j}{se(\widehat{\beta_j})} \sim Normal(0, 1)$$

- Does this mean that we should now use the standard normal distribution for inference rather than the t?

- No, it doesn't. In practice, it is ok for us to write

$$\frac{\widehat{\beta_j} - \beta_j}{se(\widehat{\beta_j})} \sim t_{n-k-1} = t_{df}$$

Q.   Why is it ok to do this?

Ans. Because t_{df} approaches the Normal(0,1) as the degrees of freedom get large.

- t-testing is carried out exactly as under the classical linear model assumptions

- If N isn't "big enough" then the t distribution can be a poor approximation to the distribution of t stats when the error is not distributed normally.
  - No real rules of thumb, however, on how large N should be.
  - I think it should be N = 328
    - Just kidding
- The approximation also depends on the number of covariates in the model because this impacts the degrees of freedom…the more covariates, the more observations we need
- Also note that for our theorem of asymptotic normality to hold, we require the assumption of homoskedasticity. If this assumption fails then the t stats are no longer valid…you guys will learn more about this in ECNS 562

- Another conclusion of our theorem on asymptotic normality is that $\widehat{\sigma^2}$ is a consistent estimator of σ^2 (we already showed this is unbiased).

- The estimated variance of $\hat{\beta}_j$ is

$$\widehat{Var(\hat{\beta}_j)} = \frac{\widehat{\sigma^2}}{SST_j(1-R_j^2)}$$

where SST_j is the total sum of squares of the x_j in the sample, and R_j^2 is the R-squared from regressing x_j on all of the other independent variables.

- Let's break down each component (as we did in the case of biasedness)
  - As N gets large, $\widehat{\sigma^2}$ converges in probability to σ^2
  - R_j^2 obviously approaches some number between zero and one
  - The sample variance of x_j is SST_j/n…so SST_j/n approaches $Var(x_j)$ as n gets large. Which implies that SST_j grows at the same rate as the sample size: $SST_j \approx n\sigma_j^2$

- Taken together, we see that $\widehat{Var(\hat{\beta}_j)}$ approaches zero at the rate of $1/n$….hence, we like large samples!

- When the error is not normally distributed, the square root of $\widehat{Var(\hat{\beta}_j)}$ is referred to as the **asymptotic standard error**, and *t* stats are called **asymptotic *t* stats**.

- Asymptotic normality of the OLS estimators also means that the *F* stats have approximate *F* distributions in large sample sizes.

Lagrange Multiplier Stat

- LM stat we discuss below relies on the same Gauss-Markov assumptions that justify the F stat in large samples
  - Do not need normality assumption

- To derive the LM stat, we consider our standard population model

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon$$

- Suppose we would like to test whether, for instance, the last q of these variables all have zero population parameters (so, q exclusion restrictions)

$$H_0: \beta_{k-q+1} = 0, \ldots, \beta_k = 0$$

$$H_1: \text{at least one of the } q \text{ parameters is not equal to zero}$$

- The LM stat requires estimation only of the restricted model (recall the F stat required estimation of both the restricted and unrestricted models)

$$y = \widetilde{\beta_0} + \widetilde{\beta_1} x_1 + \cdots + \widetilde{\beta_{k-q}} x_{k-q} + \tilde{\varepsilon}$$

- If the omitted variables x_{k-q+1} through x_k truly have zero population coefficients, then, at least approximately, $\tilde{\varepsilon}$ should be uncorrelated with each of these variables in the sample.

- This suggests running a regression of these residuals on those independent variables excluded under H_0, which is almost what the LM test does

- However, it turns out, to get a usable test statistic, we must include *all* of the independent variables in the regression

  – We must include all regressors because, in general, the omitted regressors in the restricted model are correlated with the regressors that appear in the restricted model.

- So, we run the auxiliary regression of

$$\tilde{\varepsilon} \text{ on } x_1, x_2, \ldots, x_k$$ (use this to compute test stat, but coefficient estimates in this regression are not of particular interest)

- If our null below is true

$$H_0: \beta_{k-q+1} = 0, \ldots, \beta_k = 0$$

- Then the R^2 from the auxiliary regression should be "close" to zero

- How do we determine if the test stat is "large enough" to reject H_0?

- It turns out that, under H_0, the sample size multiplied by the R^2 from the auxiliary regression is distributed asymptotically as a chi-square r.v. with q degrees of freedom.

- Cookbook procedure
  - 1.) Regress y on the restricted set of independent variables and save the residuals, $\tilde{\varepsilon}$
  - 2.) Regress $\tilde{\varepsilon}$ on all of the independent variables and obtain the R^2
  - 3.) Compute $LM = nR^2$
  - 4.) Compare the LM to the appropriate critical value, c, in a χ^2_q distribution; if LM > c, then reject H_0.
- Unlike the F stat, the degrees of freedom in the unrestricted model plays no role in carrying out the LM test.
  - All that maters is the number of restrictions being tested (q), the size of the auxiliary R^2, and the sample size.

  [show example of computing LM test in STATA]

MONTANA
STATE UNIVERSITY

Mountains & Minds

Asymptotic Efficiency of OLS

- We have already shown that OLS is BLUE

- OLS is also asymptotically efficient among a certain class of estimators

- Consider the simple regression case

$$y = \beta_0 + \beta_1 x + \varepsilon$$

- Let g(x) be any function of x (e.g., $g(x) = x^2$ or $g(x) = 1/(1+|x|)$)

  - Under the zero conditional mean assumption, this implies that ε is uncorrelated not only with x but also with g(x) (see our conditional expectations property we covered in our probability theory review that shows this).

- Let $z_i = g(x_i)$ for all observations i. Then,

$$\widetilde{\beta_1} = \left(\sum_{i=1}^{n}(z_i - \bar{z})y_i\right) / \left(\sum_{i=1}^{n}(z_i - \bar{z})x_i\right)$$

is consistent for β_1.

> -Remember, for an efficiency argument to be made we need to be comparing across estimators that are consistent.

- This can easily be shown by plugging in for y_i

$$\widetilde{\beta_1} = \beta_1 + (\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})\varepsilon_i)/(\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})x_i)$$

- With the LLN, we know that the numerator converges to Cov(z, ε) and the denominator converges to Cov(z, x).

- So, as long as Cov(z, x) ≠ 0, we know that $\widetilde{\beta_1}$ is consistent because Cov(z, ε) = 0 under the zero conditional mean assumption.

- More difficult to show asymptotic normality of $\widetilde{\beta_1}$
  - Can be shown that $\sqrt{n}(\widetilde{\beta_1} - \beta_1)$ is asymptotically normal with mean zero and asymptotic variance $\sigma^2 Var(z)/[Cov(z,x)]^2$
  - And, asymptotic variance of OLS estimator is $\sigma^2/Var(x)$
  - Can be shown that $\sigma^2/Var(x) \leq \sigma^2 Var(z)/[Cov(z,x)]^2$
    - This inequality is obtained via the Cauchy-Schwartz inequality

- Thus, OLS is asymptotically efficient

MONTANA
STATE UNIVERSITY

Mountains & Minds