Multiple Regression Analysis

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u \]

3. Asymptotic Properties
Consistency

- So far, we’ve looked at finite sample properties which hold for any sample size.
- Now we will look at asymptotic properties.
- Under the Gauss-Markov assumptions OLS is BLUE, but in other cases it won’t always be possible to find unbiased estimators.
- In those cases, we may settle for estimators that are biased but consistent; i.e. as $n \to \infty$, the distribution of the estimator collapses to the true parameter value.
Sampling Distributions as $n \rightarrow \infty$

- For each $n$, estimated $\beta_1$ has a probability distribution.
- If the estimated $\beta$ is consistent then the distribution collapses to the true parameter value as $n \rightarrow \infty$.
- Note, this estimator biased in finite samples.
Consistency of OLS

- We can show using assumptions 1-4 that, not only is the OLS estimator unbiased, it is also consistent.
- Consistency can be proved for the simple regression case in a manner similar to the proof of unbiasedness.
- Will need to take probability limit (plim) to establish consistency.

*What is the probability limit?*

- Let \( \hat{A} \) be an estimator of A.
- We want to know how close the estimate is to A as \( n \) gets large.
Probability Limits

- i.e. what is |\hat{A} - A|?
- We can write the probability that this difference is small as \( \Pr(|\hat{A} - A| < \varepsilon) \), where \( \varepsilon \) is some arbitrarily small positive number.
- For consistency: \( \lim_{n \to \infty} \left( \Pr(|\hat{A} - A| < \varepsilon) \right) = 1 \)
or \( \text{plim} \ \hat{A} = A \).
- Properties of plim:
  1) \( \text{plim} \ g(\hat{A}) = g(\text{plim} \ \hat{A}) \)
  2) \( \text{plim} \ (1/\hat{A}) = 1/(\text{plim} \ \hat{A}) \)
  3) \( \text{plim} (\hat{A} \hat{E}) = (\text{plim} \ \hat{A}) \cdot (\text{plim} \ \hat{E}) \)
  4) \( \text{plim} (\hat{A} + \hat{E}) = (\text{plim} \ \hat{A}) + (\text{plim} \ \hat{E}) \)
Law of large numbers

- Another property we will use
- If $b_n$ is iid with mean $\mu$, then
  
  $\text{plim}(\bar{b}_n) = \mu$

- We can get arbitrarily close to estimating the true mean by using a large sample.
- This will also be true for other moments.
- e.g. The probability limit (plim) of the sample variance of $b_n$ will equal the population variance.
Proving Consistency

recall:  \[ \hat{\beta}_1 = \left( \sum (x_{i1} - \bar{x}_1)y_i \right) / \left( \sum (x_{i1} - \bar{x}_1)^2 \right) \]

which we can rewrite as,

\[ = \beta_1 + \left( n^{-1} \sum (x_{i1} - \bar{x}_1)u_i \right) / \left( n^{-1} \sum (x_{i1} - \bar{x}_1)^2 \right) \]

Applying the plim operator to this, we get

\[ \text{plim} \ \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x_1,u)}{\text{Var}(x_1)} = \beta_1 \]

because \( \text{Cov}(x_1,u) = 0. \)
A Weaker Assumption

- For unbiasedness, we assumed a zero conditional mean – $E(u|x_1, x_2, \ldots, x_k) = 0$

- As illustrated above, for consistency, we can have the weaker assumption of zero mean and zero correlation – $E(u)=0$ and $Cov(x_j,u)=0$, for $j=1,2, \ldots,k$.

- Without this assumption, OLS will be biased and inconsistent.

- i.e. If any of the $x$’s are correlated with $u$ all of the estimated parameters are biased and inconsistent.
Deriving the Inconsistency

We derived the omitted variable bias earlier, when \( \text{E}(u|x_1, \ldots, x_k) \) didn’t equal zero

now we want to think about the inconsistency (aka “asymptotic bias”) when \( \text{Cov}(x_j,u) \) doesn’t equal zero

True model: \( y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \nu \)

We estimate: \( y = \beta_0 + \beta_1 x_1 + u \), so that \( u = \beta_2 x_2 + \nu \)

\[
\tilde{\beta}_1 = \beta_1 + \beta_2 \tilde{\delta}_1 \quad \text{where} \quad \tilde{\delta}_1 = \frac{\sum (x_{i1} - \bar{x}_1)x_{i2}}{\sum (x_{i1} - \bar{x}_1)^2}
\]
Deriving the Inconsistency

If we divide the numerator and denominator of $\delta_1$ by $n$:

$$\tilde{\delta}_1 = \frac{\text{sample } \text{Cov}(x_1, x_2)}{\text{sample } \text{Var}(x_1)}$$

so $\text{plim } \tilde{\beta}_1 = \beta_1 + \beta_2 \cdot \delta_1$

where $\delta_1 = \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)}$
Inconsistency (cont)

- So, the direction of the inconsistency is just like the direction of bias for an omitted variable.
- Difference is that inconsistency uses the population variance and covariance, while bias uses the sample counterparts.
- Remember, inconsistency is a large sample problem---it doesn’t go away as we increase the sample size.
- In fact, the more we increase $n$, the closer the estimator will get to $\beta_1 + \beta_2 \delta_1$.
- For the general case of $k$ RHS variables, all of the $x$’s must be uncorrelated with $u$, in order for estimates to be consistent (aka asymptotically unbiased).
Large Sample Inference

- Recall that under the CLM assumptions, the sampling distributions are normal, so we could derive $t$ and $F$ distributions for hypothesis testing.
- This exact normality was due to assuming the population error distribution was normal.
- This assumption implied that the distribution of $y$, given the $x$’s, was normal as well.
- Easy to come up with examples for which this exact normality assumption will fail.
Large Sample Inference (cont)

- Any clearly skewed variable, like wages, arrests, savings, etc. can’t be normal.
- The problem is not that OLS isn’t BLUE in these examples but that we can’t rely on our $t$ and $F$ tests for inference.
- Fortunately, the central limit theorem will allow us to show that OLS estimators are asymptotically normal, even when the error term is not normally distributed.
Asymptotic Normality

- Let \( Z_n \) be a sequence of random variables with sample size \( n \).

- Asymptotic normality implies that \( P(Z<z) \to \Phi(z) \) as \( n \to \infty \), or \( P(Z<z) \approx \Phi(z) \).

- Where \( \Phi \) is the standard normal cumulative distribution function.

- Thus if it is asymptotically normal, probabilities concerning \( Z_n \) can be approximated using the standard normal distribution.
Central Limit Theorem

- The central limit theorem states that the standardized average of any population with mean $\mu$ and variance $\sigma^2$ is asymptotically $\sim N(0,1)$.

- So, suppose that $Y_n$ has mean $\mu_Y$ and variance $\sigma^2$.

- We can write:

$$Z_n = \frac{\bar{Y} - \mu_Y}{\sigma / \sqrt{n}} \sim N(0,1)$$

- Thus, no matter what the population distribution of $Y$ is, $Z_n$ is distributed standard normal as $n$ gets large.
Asymptotic Normality

- So, we can use asymptotic normality in place of our exact normality assumption
- Under the Gauss-Markov assumptions:
  \[ i) \quad \sqrt{n}(\hat{\beta}_j - \beta_j) \xrightarrow{a} N(0, \sigma^2 / a_j^2), \]

  where \( a_j^2 = \text{plim}(\frac{1}{n} \sum \hat{r}_{ij}^2) \)

- The \( r_{ij} \)’s are residuals from regressing \( x_j \) on the other \( x \)’s
- Basically, this is just an application of asymptotic normality to \( (\hat{\beta}_{\text{hat}} - \beta) \) rewriting its variance
Asymptotic Normality (cont)

(ii) $\hat{\sigma}^2$ is a consistent estimator of $\sigma^2$

- This is just an application of large sample properties.
- Tells us we can use sigma-hat to normalize, so that

(iii) $\left(\hat{\beta}_j - \beta_j\right)/se\left(\hat{\beta}_j\right) \sim \text{Normal}(0,1)$

- This follows from the central limit theorem.
- What did we say the distribution of (iii) was in the finite sample case?
  - $t$-distribution with $n-k-1$ degrees of freedom.
Asymptotic Normality (cont)

Because the $t$ distribution approaches the normal distribution for large $df$, we can also say that

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

No difference as $n$ approaches infiniti.

We can use $t$-tests exactly as before if we have a large sample even if $y$ is not normally distributed.

Note that while we no longer need to assume normality with a large sample, we still need to assume homoskedasticity.
Asymptotic Standard Errors

If $u$ is not normally distributed, we will sometimes refer to the standard error as an asymptotic standard error.

We noted earlier that the standard errors tend toward zero as the sample size gets large.

Recall that $se(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{\sqrt{SST_j(1 - R^2)}}$.

In asymptotic analysis this converges to:

$se(\hat{\beta}_j) \approx \frac{c_j}{\sqrt{n}}$

- This shrinks at a rate proportional to root $n$.
- This is why larger sample sizes are better.
Asymptotic Efficiency

- Estimators besides OLS will be consistent
- However, under the Gauss-Markov assumptions, the OLS estimators will have the smallest asymptotic variances
- We say that OLS is asymptotically efficient
- Important to remember our assumptions though, if not homoskedastic, not true
Suppose the Gauss-Markov assumptions hold…

Question: Which of the following statements is FALSE?

1) The OLS estimator is consistent.
2) The OLS estimator is unbiased.
3) The OLS estimator has the lowest variance of consistent estimators.
4) The OLS estimator has the lowest variance of unbiased estimators.
5) None of the above statements is false.
Question: Which of the following is NOT an advantage of working with a large sample size.

1) A large sample size leads to more precise estimates.
2) A large sample size makes bias get very small, for some estimators.
3) When doing OLS estimation, a large sample size eliminates the need for the assumption of normality of the error term (assumption #6).
4) When the OLS estimator is biased, it will be consistent in large samples.