Multiple Regression Analysis: OLS Asymptotics
Outline

1. Consistency
2. Asymptotic Normality and Large Sample Inference
3. Asymptotic Efficiency of OLS
Finite Sample Properties

The unbiasedness of OLS under the first four Gauss-Markov assumptions is a finite sample property. Why?

- Because it holds for any sample size $n$.

Similarly, the fact that OLS is the best linear unbiased estimator under the full set of Gauss-Markov assumptions is a finite sample property.
Finite Sample Properties

Theorem showed that under the CLM assumptions, the OLS estimators have normal sampling distributions, which led directly to the $t$ and $F$ distributions for $t$ and $F$ statistics.

If the error is not normally distributed, the distribution of a $t$ statistic is not exactly $t$, and an $F$ statistic does not have an exact $F$ distribution for any sample size.
While not all useful estimators are unbiased, virtually all economists agree that **consistency** is a minimal requirement for an estimator.

**What is the meaning of consistency?**

- If an estimator $\hat{\beta}_j$ is consistent, then the distribution of $\hat{\beta}_j$ becomes more and more tightly distributed around $\beta_j$ as the sample size grows.

- As $n$ tends to infinity, the distribution of $\hat{\beta}_j$ collapses to the single point $\beta_j$. 
Consistency

\[ n_1 < n_2 < n_3 \]
Consistency

Theorem: Consistency of OLS

- Under assumptions MLR1 through MLR4, the OLS estimator $\hat{\beta}_j$ is consistent for $\beta_j$, for all $j = 0,1,\ldots,k$.

- Proof (in the case of the simple regression model)?

\[ y_i = \beta_0 + \beta_1 x_{i1} + u_i \]

\[ \hat{\beta}_1 = \frac{\sum (x_{1i} - \bar{x}_1)y_i}{\sum (x_{1i} - \bar{x}_1)^2} \implies \]

\[ \hat{\beta}_1 = \beta_1 + \frac{n^{-1} \sum (x_{1i} - \bar{x}_1)u_i}{n^{-1} \sum (x_{1i} - \bar{x}_1)^2} \]
Consistency

- By LLN, we have:
  
  \[ p\lim \frac{1}{n} \sum (x_{i1} - \bar{x}_1)^2 = E[(x_{i1} - \bar{x}_1)^2] = Var(x_1) \]

  \[ p\lim \frac{1}{n} \sum (x_{i1} - \bar{x}_1)u_i = E[(x_{i1} - \bar{x}_1)u_i] \]
  \[ = \text{cov}(x_1, u) \]

  \[ \Rightarrow p\lim (\hat{\beta}_1) = \beta_1 + \frac{\text{cov}(x_1, u)}{\text{var}(x_1)} \]

  \[ \text{cov}(x_1, u) = 0 \]

  \[ \Rightarrow p\lim (\hat{\beta}_1) = \beta_1 \]
Consistency

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

For consistency, we can have a weaker assumption:

Assumption MLR3\': Zero Mean and Zero Correlation?
\[ E(u) = 0 \text{ and } \text{Cov}(x_j, u) = 0, \text{ for } j = 1, 2, \ldots, k \]

Assumption MLR3 implies MLR3\', but not vice versa.

OLS is biased under assumption MLR3\'!
Deriving the Inconsistency in OLS

Just as failure of $E(u|x_1, x_2, ..., x_k) = 0$ causes bias in the OLS estimators, correlation between $u$ and any of $x_1, x_2, ..., x_k$ generally causes all of the OLS estimators to be inconsistent.

In other words, if the error is correlated with any of the independent variables, the OLS is biased and inconsistent.

This is very unfortunate because it means that any bias persists as the sample size grows.
Deriving the Inconsistency in OLS

Suppose the true model is:
\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + v \]

If we omit \( x_2 \) from the regression and do the simple regression of \( y \) on \( x_1 \), \[ y = \beta_0 + \beta_1 x_1 + u, \]
then \( u = \beta_2 x_2 + v \).

Let \( \tilde{\beta}_1 \) denote the simple regression slope estimator. Then \( \text{plim}(\tilde{\beta}_1) =? \)
\[ \text{plim}(\tilde{\beta}_1) = \beta_1 + \beta_2 \delta \]
where \( \delta = \frac{\text{cov}(x_1, x_2)}{\text{var}(x_1)} \).
Deriving the Inconsistency in OLS

- So, thinking about the direction of the asymptotic bias is just like thinking about the direction of bias for an omitted variable.

- Main difference is that asymptotic bias 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 add data.
Outline

1. Consistency
2. Asymptotic Normality and Large Sample Inference
3. Asymptotic Efficiency of OLS
Large Sample Inference

- Recall that under the CLM assumptions, the sampling distributions are normal, so we could derive $t$ and $F$ distributions for testing.

- This exact normality was due to assuming the population error distribution was normal.

- This assumption of normal errors implied that the distribution of $y$, given the $x$’s, was normal as well.
Large Sample Inference

Easy to come up with examples for which this exact normality assumption will fail.

Any clearly skewed variable, like wages, arrests, savings, etc. can’t be normal, since a normal distribution is symmetric.
Asymptotic Normality

Theorem: Asymptotic Normality

- Under the Gauss-Markov assumptions,

  (i) \( \sqrt{n}(\hat{\beta}_j - \beta_j) \xrightarrow{a} N(0, \frac{\sigma^2}{a_j^2}) \)

\[ a_j^2 = \text{plim}(n^{-1} \sum_{i=1}^{n} \hat{r}_{ij}^2), \] where the \( \hat{r}_{ij} \) are the
residuals from regression \( x_j \) on the other independent
variables.

(ii) \( \hat{\sigma}^2 \) is a consistent estimator of \( \sigma^2 = Var(u) \).
Asymptotic Normality

Theorem: Asymptotic Normality

- Under the Gauss-Markov assumptions,
- (iii) For each \( j \),

\[
\frac{\hat{\beta}_j - \beta_j}{\text{se}(\hat{\beta}_j)} \overset{a}{\sim} \text{Normal}(0,1)
\]

where \( \text{se}(\hat{\beta}_j) \) is the usual OLS standard error.
Asymptotic Normality

Because the $t$ distribution approaches the normal distribution for large degrees of freedom, we can also say that

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

Note that while we no longer need to assume normality with a large sample, we do still need homoskedasticity.
The Lagrange Multiplier Statistic

Once we are using large samples and relying on asymptotic normality for inference, we can use $t$ and $F$ statistics.

The Lagrange multiplier or $LM$ statistic is an alternative for testing multiple exclusion restrictions.

The form of the $LM$ statistic we derive here relies on the Gauss-Markov assumptions, the same assumptions that justify the $F$ statistic in large samples. We do not need to normality assumption.
The Lagrange Multiplier Statistic

Suppose we have a standard model,
\[ y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u \]

Our null hypothesis
\[ H_0: \beta_{k-q+1} = 0, \ldots, \beta_k = 0 \]

The _LM_ statistic requires estimation of the **restricted model** only.
\[ y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \cdots + \tilde{\beta}_{k-q} x_{k-q} + \tilde{u} \]
The Lagrange Multiplier Statistic

- If the omitted variables $x_{k-q+1}$ through $x_k$ truly have zero population coefficients then, at least approximately, $\hat{u}$ 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 $H_0$.
  
  $$\text{regress } \hat{u} \text{ on } x_1, x_2, \ldots, x_k$$

- This is an example of an auxiliary regression.

- $LM = nR_u^2$, where $R_u^2$ is from the regression above.
The Lagrange Multiplier Statistic

- If $H_0$ is true, the R-squared from auxiliary regression should be “close” to zero, subject to sampling error, because $\tilde{u}$ will be approximately uncorrelated with all the independent variables.

- Under the null hypothesis, the $LM$ statistic is distributed asymptotically as a chi-square random variable with $q$ degrees of freedom.

$$LM \sim \chi^2_q$$
Outline

1. Consistency
2. Asymptotic Normality and Large Sample Inference
3. Asymptotic Efficiency of OLS
Asymptotic Efficiency of OLS

- 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.