MLR.1 (linear in parameters): $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$.

MLR.2 (Random sampling)

MLR.3 (No perfect collinearity)

MLR.4 : (Zero conditional mean) $E(u \mid x_1, \ldots, x_k) = 0$

MLR.5 (Homoskedasticity) $Var(u \mid x_1, \ldots, x_k) = \sigma^2$

MLR.6 : $u_i \sim N(0, \sigma^2)$ independently of $x_{i1}, x_{i2}, \ldots x_{ik}$.

$\Rightarrow y \mid x \sim N(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k, \sigma^2)$.

MLR.1 to MLR.6 is called classical linear model assumption. (CLM)

- Properties of OLS that hold for any sample/sample size
  - Expected values/unbiasedness under MLR.1 – MLR.4
  - Variance formulas under MLR.1 – MLR.5
  - Gauss-Markov Theorem under MLR.1 – MLR.5
  - Exact sampling distributions/tests under MLR.1 – MLR.6

- Properties of OLS that hold in large samples
  - Consistency under MLR.1 – MLR.4
  - Asymptotic normality/tests under MLR.1 – MLR.5
    - Note that we drop MLR.6

- **Consistency**

An estimator θ_n is consistent for a population parameter θ if

$$P(|\theta_n - \theta| < \epsilon) \rightarrow 1 \text{ for arbitrary } \epsilon > 0 \text{ and } n \rightarrow \infty.$$

Alternative notation: $plim\ \theta_n = \theta$ ← The estimate converges in probability to the true population value

- Interpretation:
  - Consistency means that the probability that the estimate is arbitrarily close to the true population value can be made arbitrarily high by increasing the sample size
- Consistency is a minimum requirement for sensible estimators

Theorem: OLS estimators are consistent.

$$\operatorname{plim} \hat{\beta}_j = \beta_j \qquad j = 0, 1, 2, \ldots, k$$

Example: $y = \beta_0 + \beta_1 x + u$.

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$= \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2}$$

$$= \frac{\sum (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum (x_i - \bar{x})^2}$$

$$= \frac{\underbrace{\sum (x_i - \bar{x})}_{=0} \beta_0 + \sum (x_i - \bar{x}) \beta_1 x_i + \sum (x_i - \bar{x}) u_i}{\sum (x_i - \bar{x})^2}$$

$$\sum (x_i - \bar{x})^2 = \sum (x_i - \bar{x}) x_i + \underbrace{\sum (x_i - \bar{x})}_{=0} \bar{x}$$

$$= \beta_1 + \frac{\sum (x_i - \bar{x}) u_i}{\sum (x_i - \bar{x})^2}$$

$$\Longrightarrow \operatorname{plim} \hat{\beta}_1 = \operatorname{plim} \beta_1 + \operatorname{plim} \frac{\sum (x_i - \bar{x}) u_i}{\sum (x_i - \bar{x})^2}$$

$$= \beta_1 + \operatorname{plim} \frac{\frac{1}{n} \sum x_i u_i - \bar{x} \cdot \frac{1}{n} \sum u_i}{\frac{1}{n} \sum (x_i - \bar{x})^2}$$

$$= \beta_1 + \frac{\underbrace{\operatorname{Cov}(x, u)}_{=0}}{\operatorname{Var}(x)}$$

$$= \beta_1$$

The correlation between u and any of $x_1, \dots, x_k$ causes biased and inconsistency.

based on sample — is expressed in population variance and covariance.

$$\text{plim}\ \hat{\beta}_1 = \beta_1 + \boxed{\frac{Cov(x,u)}{Var(x)}}$$

the inconsistency is also called asymptotic bias.

Suppose true model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + v.$ $u.$

$$\text{plim}\ \tilde{\beta}_1 = \beta_1 + \frac{Cov(x_1,u)}{Var(x_1)} = \beta_1 + \beta_2 \delta_1$$

$$\delta_1 = \frac{Cov(x_1, x_2)}{Var(x_1)}$$

if $Cov(x_1, x_2) = 0$. $\tilde{\beta}_1$ is consistent estimator of β_1. (may not necessarily unbiased). no matter whether u, x are indep.

inconsistency can't disapper by adding more observations to the sample.

Asymptotic Normality and Large Sample Inference.

MLR.1 → 6. ⟹ sampling distribution is normal.
$y|x$, x , u.

Even though y_i may not from a normal distribution. we can use Central Limit Theorem (CLT) to conclude that the OLS estimators satisfy asymptotic normality.

Theorem: Under the Gauss-Markov Assumption (MLR.1–5).

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

$\frac{\sigma^2}{a_j^2} > 0$. is the asymptotic variance.

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

$\hat{\beta}_j$ is asymptotically normal distributed.

(ii) $\hat{\sigma}^2$ is a consistent estimator of $\sigma^2 = Var(u)$.
$= \frac{SSR}{n-k-1}$

(iii) for each j, $\dfrac{(\hat{\beta}_j - \beta_j)}{sd(\hat{\beta}_j)} \overset{a}{\sim} N(0,1)$

$$\frac{(\hat{\beta}_j - \beta_j)}{se(\hat{\beta}_j)} \overset{a}{\sim} N(0,1)$$

$$t_{\underline{n-k-1} \to +\infty} \overset{a}{\sim} N(0,1).$$

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

SST_j: total sample variation in x_j

R_j^2: R-squared from a regression of x_j on all other independent variables

$\hat{\sigma}^2 \xrightarrow{P} \sigma^2$

R_j^2 approaches a number strictly between 0 and 1.

SST_j approaches $\underline{n\,Var(X_j)}$ as the sample size grows.
$n\sigma_j^2$

$\Longrightarrow \widehat{Var(\hat{\beta}_j)} \approx \frac{1}{n} \cdot \text{constant}$

that is why larger sample size is better.

when u is not normal distributed.

$se(\hat{\beta}_j) = \sqrt{\widehat{Var(\hat{\beta}_j)}}$ is called <u>asymptotic standard error</u>

t statistics are called <u>asymptotic t statistics</u>

$$se(\hat{\beta}_j) \approx \frac{C_j}{\sqrt{n}} \longrightarrow \text{constant.}$$

$$C_j = \frac{\overset{\nearrow\, sd(u)}{\sigma}}{\underset{\downarrow\, sd(u_j)}{\sigma_j}\sqrt{1-\rho_j^2}} \longrightarrow \text{plim } R_j^2$$

(approximation)