ECON 351* -- NOTE 11

The Multiple Classical Linear Regression Model (CLRM): Specification and Assumptions

1. Introduction

CLRM stands for the **Classical Linear Regression Model**. The CLRM is also known as the ***standard* linear regression model**.

Three sets of assumptions define the *multiple* CLRM -- essentially the same three sets of assumptions that defined the *simple* CLRM, with one modification to assumption A8.

1. Assumptions respecting the **formulation of the *population regression equation*, or PRE**.

Assumption A1

2. Assumptions respecting the **statistical properties of the *random error term* and the *dependent variable***.

Assumptions A2-A4

- Assumption A2: The Assumption of *Zero Conditional Mean Error*
- Assumption A3: The Assumption of *Constant Error Variances*
- Assumption A4: The Assumption of *Zero Error Covariances*

3. Assumptions respecting the **properties of the *sample data***.

Assumptions A5-A8

- Assumption A5: The Assumption of *Independent Random Sampling*
- Assumption A6: The Assumption of *Sufficient Sample Data (N > K)*
- Assumption A7: The Assumption of *Nonconstant Regressors*
- Assumption A8: The Assumption of *No Perfect Multicollinearity*

2. Formulation of the Population Regression Equation (PRE)

<u>Assumption A1</u>: The population regression equation, or PRE, takes the form

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u = \beta_0 + \sum_{j=1}^{k} \beta_j X_j + u \qquad \textbf{(A1)}$$

or

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} + u_i \qquad \textbf{(A1)}$$

The second form of (A1) writes the PRE for a particular observation i.

As in the simple CLRM, the **PRE (A1) incorporates *three* distinct assumptions**.

A1.1: <u>Assumption of an Additive Random Error Term</u>.

⇒ **The random error term u_i enters the PRE *<u>additively</u>*.**

$$\frac{\partial Y_i}{\partial u_i} = 1 \qquad \text{for all i } (\forall \text{ i}).$$

A1.2: <u>Assumption of Linearity-in-Parameters or Linearity-in-Coefficients</u>.

⇒ **The PRE is *<u>linear</u>* in the population regression coefficients β_j (j = 0, ..., k).**

Let $\underline{x}_i = [1 \;\; X_{1i} \;\; X_{2i} \;\; \cdots \;\; X_{ki}]$ be the (K×1) vector of regressor values for observation i.

$\frac{\partial Y_i}{\partial \beta_j} = f_j(\underline{x}_i)$ where $f_j(\underline{x}_i)$ contains ***no unknown parameters***, j = 0, ..., k.

A1.3: <u>Assumption of Parameter or Coefficient Constancy</u>.

⇒ **The population regression coefficients β_j (j = 0, 1, ..., k) are (unknown) *<u>constants</u>* that do not vary across observations.**

$\beta_{ji} = \beta_j =$ **a *constant* ∀ i** (j = 0, 1, ..., k).

3. Properties of the Random Error Term

<u>Assumption A2</u>: The Assumption of *Zero Conditional Mean Error*

The *conditional mean,* or *conditional expectation,* of the random error terms u_i for any given values X_{ji} of the regressors X_j is equal to zero:

$$E(u \mid X_1, X_2, \ldots, X_k) = E(u \mid \underline{x}) = 0 \qquad \textbf{(A2)}$$

or

$$E(u_i \mid X_{1i}, X_{2i}, \ldots, X_{ki}) = E(u_i \mid \underline{x}_i) = 0 \quad \forall\ i \qquad \textbf{(A2)}$$

where $\underline{x} = [1\ \ X_1\ \ X_2\ \ \cdots\ \ X_k]$ is any (K×1) vector of regressor values, and $\underline{x}_i = [1\ \ X_{1i}\ \ X_{2i}\ \ \cdots\ \ X_{ki}]$ denotes the (K×1) vector of regressor values for a particular observation, namely observation i.

<u>Implications of Assumption A2</u>

- **<u>Implication 1 of A2</u>.** Assumption A2 implies that the ***unconditional* mean** of the population values of **the random error term u equals *zero***:

$$E(u \mid \underline{x}) = 0 \quad \Rightarrow \quad E(u) = 0 \qquad \textbf{(A2-1)}$$

or

$$E(u_i \mid \underline{x}_i) = 0 \quad \Rightarrow \quad E(u_i) = 0 \quad \forall\ i. \qquad \textbf{(A2-1)}$$

This implication follows from the so-called **law of iterated expectations**, which states that $E[E(u \mid \underline{x})] = E(u)$. Since $E(u \mid \underline{x}) = 0$ by A2, it follows that $E(u) = E[E(u \mid \underline{x})] = E[0] = 0$.

The logic of (A2-1) is straightforward: If the conditional mean of u for each and every population value of $\underline{x}$ equals zero, then the mean of these zero conditional means must also be zero.

- **<u>Implication 2 of A2: the Orthogonality Condition</u>.** Assumption A2 also implies that **the population values X_{ji} of the regressor X_j and u_i of the random error term u have *zero covariance* -- i.e., the population values of X_j and u are *uncorrelated***:

$$E(u|\underline{x}) = 0 \;\Rightarrow\; Cov(X_j, u) = E(X_j u) = 0, \; j = 1, 2, \ldots, k \qquad \textbf{(A2-2)}$$

or

$$E(u_i|\underline{x}_i) = 0 \;\Rightarrow\; Cov(X_{ji}, u_i) = E(X_{ji} u_i) = 0 \;\; \forall\, i, \; j = 1, 2, \ldots, k \qquad \textbf{(A2-2)}$$

1. The equality $Cov(X_{ji}, u_i) = E(X_{ji}u_i)$ in (A2-2) follows from the definition of the covariance between X_{ji} and u_i, and from assumption (A2):

$$\begin{aligned}
Cov(X_{ji}, u_i) &\equiv E\{[X_{ji} - E(X_{ji})][u_i - E(u_i|\underline{x}_i)]\} && \text{by definition} \\
&= E\{[X_{ji} - E(X_{ji})]u_i\} && \text{since } E(u_i|\underline{x}_i) = 0 \text{ by A2} \\
&= E[X_{ji}u_i - E(X_{ji})u_i] && \\
&= E(X_{ji}u_i) - E(X_{ji})E(u_i) && \text{since } E(X_{ji}) \text{ is a constant} \\
&= E(X_{ji}u_i) && \text{since } E(u_i) = E(u_i|\underline{x}_i) = 0 \text{ by A2.}
\end{aligned}$$

2. Implication (A2-2) states that the **random error term u** has ***zero covariance with***, or is ***uncorrelated with***, each of the **regressors X_j** (j = 1, …, k) in the population. This assumption means that there exists ***no <u>linear</u> association* between u and any of the k regressors X_j (j = 1, …, k)**.

Note that ***zero covariance* between X_{ji} and u_i** implies ***zero correlation* between X_{ji} and u_i**, since the **simple *correlation coefficient* between X_{ji} and u_i**, denoted as $\rho(X_{ji}, u_i)$, is defined as

$$\rho(X_{ji}, u_i) \equiv \frac{Cov(X_{ji}, u_i)}{\sqrt{Var(X_{ji})\,Var(u_i)}} = \frac{Cov(X_{ji}, u_i)}{sd(X_{ji})\,sd(u_i)}.$$

From this definition of $\rho(X_{ji}, u_i)$, it is obvious that if $Cov(X_{ji}, u_i) = 0$, then $\rho(X_{ji}, u_i) = 0$, i.e.,

$$Cov(X_{ji}, u_i) = 0 \;\Rightarrow\; \rho(X_{ji}, u_i) = 0.$$

- **Implication 3 of A2.** Assumption A2 implies that the ***conditional mean*** **of the population Y_i values** corresponding to given values X_{ji} of the regressors X_j (j = 1, …, k) **equals the *population regression function (PRF)***:

$$E(u \mid \underline{x}) = 0 \;\Rightarrow\; E(Y \mid \underline{x}) = f(\underline{x}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

$$= \beta_0 + \sum_{j=1}^{k} \beta_j X_j \qquad \textbf{(A2-3)}$$

or

$$E(u_i \mid \underline{x}_i) = 0 \;\Rightarrow\; E(Y_i \mid \underline{x}_i) = f(\underline{x}_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$$

$$= \beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} \qquad \forall\, i. \qquad \textbf{(A2-3)}$$

Proof: Take the conditional expectation of the PRE (A1) for some given set of regressor values $\underline{x}_i = \begin{bmatrix} 1 & X_{1i} & X_{2i} & \cdots & X_{ki} \end{bmatrix}$:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} + u_i \qquad \textbf{(A1)}$$

$$\begin{aligned}
E(Y_i \mid \underline{x}_i) &= E(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} \mid \underline{x}_i) + E(u_i \mid \underline{x}_i) \\
&= E(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} \mid \underline{x}_i) \qquad \text{by A2, } E(u_i \mid \underline{x}_i) = 0 \\
&= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} \\
&= \beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} \qquad \text{since } E\left(\beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} \,\middle|\, \underline{x}_i\right) = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ji}.
\end{aligned}$$

- **<u>Meaning of the Zero Conditional Mean Error Assumption A2</u>:**

Each set of regressor values $\underline{x}_i = \begin{bmatrix} 1 & X_{1i} & X_{2i} & \cdots & X_{ki} \end{bmatrix}$ identifies a segment or subset of the relevant population, specifically the segment that has those particular values of the regressors. For each of these population segments or subsets, assumption A2 says that the mean of the random error u is zero.

Assumption A2 rules out both ***linear*** **dependence and** ***nonlinear*** **dependence between each X_j and u;** that is, **it requires that X_j and u be** ***statistically independent*** **for all j = 1, …, k**.

- The ***absence*** **of** ***linear*** **dependence between X_j and u** means that **X_j and u are** ***uncorrelated***, or equivalently that **X_j and u have** ***zero covariance***.
- But linear independence between X_j and u is not sufficient to guarantee the satisfaction of assumption A2. It is possible for X_j and u to be both uncorrelated, or linearly unrelated, and nonlinearly related.
- Assumption A2 therefore also requires that there be **no** ***nonlinear*** **relationship between X_j and u**.

- **<u>Violations of the Zero Conditional Mean Error Assumption A2</u>**

- Remember that the random error term u represents all the unobservable, unmeasured and unknown variables other than the regressors X_j, j = 1, …, k that determine the population values of the dependent variable Y.
- Anything that causes the random error u to be correlated with one or more of the regressors X_j (j = 1, …, k) will violate assumption A2:

$$\text{Cov}(X_j, u) \neq 0 \;\; or \;\; \rho(X_j, u) \neq 0 \quad \Rightarrow \quad E(u \mid \underline{x}) \neq 0 .$$

If X_j and u are correlated, then $E(u \mid \underline{x})$ must depend on X_j and so cannot be zero.

Note that the converse is not true:

$\text{Cov}(X_j, u) = 0 \;\; or \;\; \rho(X_j, u) = 0$ for all j does *not* imply that $E(u \mid \underline{x}) = 0$.

<u>*Reason*</u>: $\text{Cov}(X_j, u)$ measures only ***linear*** **dependence** between u and X_j. But any ***nonlinear*** **dependence** between u and X_j will also cause $E(u \mid \underline{x})$ to depend on X_j, and hence to differ from zero. So $\text{Cov}(X_j, u) = 0$ for all j = 1, …, k is not enough to insure that assumption A2 is satisfied.

- **Common causes of correlation or dependence between the X_j and u** -- i.e., common causes of violations of assumption A2.

  1. **Incorrect specification of the functional form of the relationship between Y and the X_j, j = 1, …, k.**

     Examples: Using Y as the dependent variable when the true model has ln(Y) as the dependent variable. Or using X_j as the independent variable when the true model has $\ln(X_j)$ as the independent variable.

  2. **Omission of relevant variables that are correlated with one or more of the included regressors X_j, j = 1, …, k.**

  3. **Measurement errors in the regressors X_j, j = 1, …, k.**

  4. **Joint determination of one or more X_j and Y.**

<u>Assumption A3</u>: The Assumption of *Constant Error Variances*
The Assumption of *Homoskedastic Errors*
The Assumption of *Homoskedasticity*

The *conditional variances* of the random error terms u_i are identical for all observations -- i.e., for all sets of regressor values $\underline{x} = [1 \;\; X_1 \;\; X_2 \;\; \cdots \;\; X_k]$**) -- and equal the same finite positive constant σ^2 for all i:**

$$\mathrm{Var}(u|\underline{x}) = E(u^2|\underline{x}) = \sigma^2 > 0 \qquad \textbf{(A3)}$$

or

$$\mathrm{Var}(u_i|\underline{x}_i) = E(u_i^2|\underline{x}_i) = \sigma^2 > 0 \quad \forall\, i \qquad \textbf{(A3)}$$

where σ^2 is a *finite positive (unknown) constant* and $\underline{x}_i = [1 \;\; X_{1i} \;\; X_{2i} \;\; \cdots \;\; X_{ki}]$ is the (K×1) vector of regressor values for observation i.

- The **first equality in A3** follows from the definition of the conditional variance of u_i and assumption A2:

$$\mathrm{Var}(u_i|\underline{x}_i) \equiv E\{[u_i - E(u_i|\underline{x}_i)]^2|\,\underline{x}_i\} \qquad \text{by definition}$$
$$= E\{[u_i - 0\,]^2|\,\underline{x}_i\} \qquad \text{because } E(u_i|\underline{x}_i) = 0 \text{ by assumption A2}$$
$$= E(u_i^2|\,\underline{x}_i).$$

- **<u>Implication 1 of A3</u>:** Assumption A3 implies that the ***unconditional* variance of the random error u** is also equal to σ^2:

  $$\text{Var}(u_i) = E\left[(u_i - E(u_i))^2\right] = E\left(u_i^2\right) = \sigma^2 \quad \forall\ i.$$

  where $\text{Var}(u_i) = E\left(u_i^2\right)$ because $E(u_i) = 0$ by A2-1.

  <u>Proof</u>: By assumptions A2 and A3, $E\left(u_i^2 \mid \underline{x}_i\right) = \sigma^2$.

  By ***the law or iterated expectations***, $E\left[E\left(u_i^2 \mid \underline{x}_i\right)\right] = E\left(u_i^2\right)$.
  Thus,

  $$\text{Var}(u_i) = E\left(u_i^2\right) = E\left[E\left(u_i^2 \mid \underline{x}_i\right)\right] = E\left[\sigma^2\right] = \sigma^2 \quad \forall\ i.$$

- **<u>Implication 2 of A3</u>:** Assumption A3 implies that the **conditional variance of the regressand Y_i** corresponding to given set of regressor values $\underline{x}_i = \left[1\ \ X_{1i}\ \ X_{2i}\ \ \cdots\ \ X_{ki}\right]$ ***equals* the conditional error variance σ^2**:

  $$\text{Var}(u \mid \underline{x}) = \sigma^2 > 0 \qquad \Rightarrow \qquad \text{Var}(Y \mid \underline{x}) = \sigma^2 > 0. \qquad \textbf{(A3-2)}$$

  or

  $$\text{Var}(u_i \mid \underline{x}_i) = \sigma^2 > 0 \quad \forall\ i \qquad \Rightarrow \qquad \text{Var}(Y_i \mid \underline{x}_i) = \sigma^2 > 0 \quad \forall\ i. \qquad \textbf{(A3-2)}$$

  <u>Proof</u>: Start with the definition of the conditional variance of Y_i for some given set (vector) of values of the regressors $\underline{x}_i = \left[1\ \ X_{1i}\ \ X_{2i}\ \ \cdots\ \ X_{ki}\right]$.

  $$\text{Var}(Y_i \mid \underline{x}_i) \equiv E\left\{[Y_i - E(Y_i \mid \underline{x}_i)]^2 \mid \underline{x}_i\right\} \quad \text{by definition}$$

  $$= E\left\{\left[Y_i - \beta_0 - \sum_{j=1}^{k}\beta_j X_{ji}\right]^2 \mid \underline{x}_i\right\} \quad \text{since } E(Y_i \mid \underline{x}_i) = \beta_0 + \sum_{j=1}^{k}\beta_j X_{ji} \text{ by A2}$$

  $$= E\left(u_i^2 \mid \underline{x}_i\right) \quad \text{since } u_i = Y_i - \beta_0 - \sum_{j=1}^{k}\beta_j X_{ji} \text{ by A1}$$

  $$= \sigma^2 \quad \text{since } E\left(u_i^2 \mid \underline{x}_i\right) = \sigma^2 \text{ by assumption A3.}$$

- **Meaning of the Homoskedasticity Assumption A3**

- For each set of regressor values, there is a ***conditional*** **distribution of random errors**, and a corresponding ***conditional*** **distribution of population Y values**.

- Assumption A3 says that the ***variance*** **of the random errors for any particular set of regressor values** $\underline{x}_i = \begin{bmatrix} 1 & X_{1i} & X_{2i} & \cdots & X_{ki} \end{bmatrix}$ is the ***same*** as the ***variance*** **of the random errors for any other set of regressor values** $\underline{x}_s = \begin{bmatrix} 1 & X_{1s} & X_{2s} & \cdots & X_{ks} \end{bmatrix}$ (for all $\underline{x}_s \neq \underline{x}_i$).

  In other words, the ***variances*** of the ***conditional*** **random error distributions** corresponding to each set of regressor values in the relevant population **are all *equal* to the *same* finite positive constant** $\boldsymbol{\sigma^2}$.

$$\text{Var}\left(u_i \middle| \underline{x}_i\right) = \text{Var}\left(u_s \middle| \underline{x}_s\right) = \sigma^2 > 0 \quad \text{for all } \underline{x}_s \neq \underline{x}_i .$$

- Implication A3-2 says that the ***variance*** **of the population Y values for** $\underline{x} = \underline{x}_i = \begin{bmatrix} 1 & X_{1i} & X_{2i} & \cdots & X_{ki} \end{bmatrix}$ is the ***same*** as the ***variance*** **of the population Y values for any other set of regressor values** $\underline{x} = \underline{x}_s = \begin{bmatrix} 1 & X_{1s} & X_{2s} & \cdots & X_{ks} \end{bmatrix}$ (for all $\underline{x}_s \neq \underline{x}_i$). The ***conditional distributions*** **of the population Y values** around the PRF have the ***same constant*** **variance** $\boldsymbol{\sigma^2}$ for all sets of regressor values.

$$\text{Var}\left(Y_i \middle| \underline{x}_i\right) = \text{Var}\left(Y_s \middle| \underline{x}_s\right) = \sigma^2 > 0 \quad \text{for all } \underline{x}_s \neq \underline{x}_i .$$

Assumption A4: The Assumption of *Zero Error Covariances*
The Assumption of *Nonautoregressive Errors*
The Assumption of *Nonautocorrelated Errors*

Consider any pair of distinct random error terms u_i and u_s ($i \neq s$) corresponding to two different sets (or vectors) of regressor values $\underline{x}_i \neq \underline{x}_s$. This assumption states that u_i and u_s have *zero covariance*:

$$\text{Cov}\left(u_i, u_s \middle| \underline{x}_i, \underline{x}_s\right) = E\left(u_i u_s \middle| \underline{x}_i, \underline{x}_s\right) = 0 \quad \forall\, i \neq s. \qquad \textbf{(A4)}$$

- The **first equality in (A4)** follows from the definition of the conditional covariance of u_i and u_s and assumption (A2):

$$\text{Cov}\left(u_i, u_s \middle| \underline{x}_i, \underline{x}_s\right) \equiv E\left\{\left[u_i - E(u_i|\underline{x}_i)\right]\left[u_s - E(u_s|\underline{x}_s)\right] \middle| \underline{x}_i, \underline{x}_s\right\} \quad \text{by definition}$$

$$= E\left(u_i u_s \middle| \underline{x}_i, \underline{x}_s\right) \qquad \text{since } E(u_i|\underline{x}_i) = E(u_s|\underline{x}_s) = 0 \text{ by A2.}$$

- The **second equality in (A4)** states the assumption that all pairs of error terms corresponding to different sets of regressor values have zero covariance.

- **Implication of A4:** Assumption A4 implies that the conditional covariance of any two distinct values of the regressand, say Y_i and Y_s where $i \neq s$, is equal to zero:

$$\text{Cov}\left(u_i, u_s \middle| \underline{x}_i, \underline{x}_s\right) = 0 \quad \forall\, i \neq s \quad \Rightarrow \quad \text{Cov}\left(Y_i, Y_s \middle| \underline{x}_i, \underline{x}_s\right) = 0 \quad \forall\, i \neq s.$$

Proof: Show that $\text{Cov}\left(Y_i, Y_s \middle| \underline{x}_i, \underline{x}_s\right) = E\left(u_i u_s \middle| \underline{x}_i, \underline{x}_s\right) = \text{Cov}\left(u_i, u_s \middle| \underline{x}_i, \underline{x}_s\right)$.

(1) Begin with the definition of the conditional covariance for Y_i and Y_s for given $\underline{x}_i$ and $\underline{x}_s$ values where $\underline{x}_i \neq \underline{x}_s$:

$$\begin{aligned}\text{Cov}\left(Y_i, Y_s \middle| \underline{x}_i, \underline{x}_s\right) &\equiv E\left\{\left[Y_i - E(Y_i|\underline{x}_i)\right]\left[Y_s - E(Y_s|\underline{x}_s)\right]\middle|\underline{x}_i, \underline{x}_s\right\} \\ &= E\left(u_i u_s \middle| \underline{x}_i, \underline{x}_s\right)\end{aligned}$$

since

$$Y_i - E\left(Y_i \middle| \underline{x}_i\right) = Y_i - \beta_0 - \sum_{j=1}^{k} \beta_j X_{ji} = u_i \qquad \text{by assumption A1,}$$

$$Y_s - E\left(Y_s \middle| \underline{x}_s\right) = Y_s - \beta_0 - \sum_{j=1}^{k} \beta_j X_{js} = u_s \qquad \text{by assumption A1.}$$

(2) Therefore

$$\text{Cov}\left(Y_i, Y_s \middle| \underline{x}_i, \underline{x}_s\right) = E\left(u_i u_s \middle| \underline{x}_i, \underline{x}_s\right) = 0 \qquad \text{by assumption A4.}$$

- **Meaning of A4:** Assumption A4 means that there is **no systematic *linear association* between u_i and u_s, or between Y_i and Y_s**, where i and s correspond to **different observations** (or different sets of regressor values $\underline{x}_i \neq \underline{x}_s$).

1. Each random error term u_i has ***zero covariance with,*** or ***is uncorrelated with,*** each and every other random error term u_s ($s \neq i$).

2. Equivalently, each regressand value Y_i has ***zero covariance with***, or ***is uncorrelated with***, each and every other regressand value Y_s ($s \neq i$).

- The **assumption of *zero covariance***, or ***zero correlation***, between each pair of distinct observations is ***weaker*** than the **assumption of *independent random sampling A5*** from an underlying population.

- The **assumption of *independent random sampling*** implies that the sample observations are statistically independent. The **assumption of *statistically independent observations* is *sufficient* for the assumption of *zero covariance*** between observations, but is stronger than necessary.

4. Properties of the Sample Data

Assumption A5: Random Sampling *or* Independent Random Sampling

The **sample data** consist of **N *randomly selected observations*** on the regressand Y and the regressors X_j (j = 1, ..., k), the observable variables in the PRE described by A1. These N randomly selected observations can be written as N row vectors:

$$\begin{aligned} \text{Sample data} &\equiv \left[(Y_1, \underline{x}_1), (Y_2, \underline{x}_2), \ldots, (Y_N, \underline{x}_N)\right] \\ &\equiv \left(Y_i, 1, X_{1i}, X_{2i}, \ldots, X_{ki}\right) \qquad i = 1, \ldots, N \\ &\equiv (Y_i, \underline{x}_i) \qquad i = 1, \ldots, N. \end{aligned}$$

- **Implications of the Random Sampling Assumption A5**

  The **assumption of random sampling** implies that **the sample observations are *statistically independent***.

  1. It thus means that the error terms **u_i and u_s are *statistically independent***, and hence **have *zero covariance***, for any two observations i and s.

  $$\text{Random sampling} \Rightarrow \text{Cov}\left(u_i, u_s \mid \underline{x}_i, \underline{x}_s\right) = \text{Cov}\left(u_i, u_s\right) = 0 \qquad \forall \; i \neq s.$$

  2. It also means that the dependent variable values **Y_i and Y_s are *statistically independent***, and hence **have *zero covariance***, for any two observations i and s.

  $$\text{Random sampling} \Rightarrow \text{Cov}\left(Y_i, Y_s \mid \underline{x}_i, \underline{x}_s\right) = \text{Cov}\left(Y_i, Y_s\right) = 0 \qquad \forall \; i \neq s.$$

  The assumption of random sampling is therefore sufficient for assumption A4 of zero covariance between observations, but is stronger than necessary.

- **When is the Random Sampling Assumption A5 Appropriate?**

  The random sampling assumption is often appropriate for ***cross-sectional* regression models**, but is hardly ever appropriate for ***time-series* regression models**.

<u>Assumption A6</u>: The number of sample observations N is greater than the number of unknown parameters K:

number of sample observations > number of unknown parameters

$$\mathbf{N > K.} \qquad \text{(A6)}$$

- **<u>Meaning of A6</u>:** Unless this assumption is satisfied, it is not possible to compute from a given sample of N observations estimates of all the unknown parameters in the model.

<u>Assumption A7</u>: Nonconstant Regressors

The sample values X_{ji} of each regressor X_j (j = 1, …, k) in a given sample (and hence in the population) are not all equal to a constant:

$$X_{ji} \neq c_j \quad \forall \; i = 1, ..., N \quad \text{where the } c_j \text{ are constants } (j = 1, ..., k) \,. \qquad \text{(A7)}$$

- **<u>Technical Form of A7</u>:** Assumption A7 requires that the ***sample*** **variances of all k–1 non-constant regressors X_j** (j = 1, ..., k) must be ***finite positive*** **numbers** for any sample size N; i.e.,

  $$\text{sample variance of } X_{ji} \equiv \text{Var}(X_{ji}) = \frac{\sum_i (X_{ji} - \bar{X}_j)^2}{N-1} = s^2_{X_j} > 0,$$

  where $s^2_{X_j} > 0$ are ***finite positive*** **numbers** for all j = 1, ..., k.

- **<u>Meaning of A7</u>:** Assumption A7 requires that **each nonconstant regressor X_j** (j = 1, …, k) takes **at least *two* different values** in any given sample.

  Unless this assumption is satisfied, it is not possible to compute from the sample data an estimate of the effect on the regressand Y of changes in the value of the regressor X_j. In other words, to calculate the effect of changes in X_j on Y, the sample values X_{ji} of the regressor X_j must vary across observations in any given sample.

<u>Assumption A8</u>: No Perfect Multicollinearity

The sample values of the regressors X_j (j = 1, ..., k) in a multiple regression model do *not* exhibit *perfect or exact multicollinearity*.

This assumption is relevant only in ***multiple* regression models** that contain two or more non-constant regressors.

This assumption is the only new assumption required for the multiple linear regression model.

- **<u>Statement of Assumption A8</u>:** The **absence of *perfect multicollinearity*** means that there exists **no *exact linear* relationship** among the **sample values** of the non-constant regressors X_j (j = 1, ..., k).

  - An exact linear relationship exists among the sample values of the non-constant regressors if the sample values of the regressors X_j (j = 1, ..., k) satisfy a linear relationship of the form

$$\lambda_0 + \lambda_1 X_{1i} + \lambda_2 X_{2i} + \cdots + \lambda_k X_{ki} = 0 \qquad \forall \; i = 1, 2, \ldots, N. \tag{1}$$

    where the λ_j (j = 0, 1, …, k) are **fixed constants**, not all of which equal zero.

  - Assumption A8 – the absence of perfect multicollinearity – means that **there exists no relationship of the form (1)** among the sample values X_{ji} of the regressors X_j (j = 1, ..., k).

- **<u>Meaning of Assumption A8</u>:**

  - Each non-constant regressor X_j (j = 1, ..., k) must exhibit some ***independent* linear variation** in the sample data.

  - Otherwise, it is not possible to estimate the ***separate* linear effect** of each and every non-constant regressor on the regressand Y.

- **Example of Perfect Multicollinearity**

Consider the following multiple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (i = 1, \ldots, N). \qquad (2)$$

Suppose that the sample values of the regressors X_{1i} and X_{2i} satisfy the following linear equality for all sample observations:

$$X_{1i} = 3X_{2i} \qquad \textit{or} \qquad X_{1i} - 3X_{2i} = 0 \;\; \forall \;\; i = 1, \ldots, N. \qquad (3)$$

The exact linear relationship (3) can be written in the general form (1).

1. For the linear regression model given by PRE (2), equation (1) takes the form

$$\lambda_0 + \lambda_1 X_{1i} + \lambda_2 X_{2i} = 0 \qquad \forall \;\; i = 1, 2, \ldots, N.$$

2. Set $\lambda_0 = 0$, $\lambda_1 = 1$, and $\lambda_2 = -3$ in the above equation:

$$X_{1i} - 3X_{2i} = 0 \;\; \forall \;\; i = 1, 2, \ldots, N. \quad \text{(identical to equation (3) above.)}$$

- **Consequences of Perfect Multicollinearity**

1. Substitute for X_{1i} in PRE (2) the equivalent expression $X_{1i} = 3X_{2i}$:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \\ &= \beta_0 + \beta_1 (3X_{2i}) + \beta_2 X_{2i} + u_i \\ &= \beta_0 + 3\beta_1 X_{2i} + \beta_2 X_{2i} + u_i \\ &= \beta_0 + (3\beta_1 + \beta_2) X_{2i} + u_i \\ &= \beta_0 + \alpha_2 X_{2i} + u_i \qquad \text{where } \alpha_2 = 3\beta_1 + \beta_2 \qquad (4a) \end{aligned}$$

- ♦ It is possible to estimate from the sample data the regression coefficients β_0 and α_2.

- ♦ But from the estimate of α_2 it is not possible to compute estimates of the coefficients β_1 and β_2. ***Reason:*** The equation

$$\alpha_2 = 3\beta_1 + \beta_2$$

is *one* equation containing *two* unknowns, namely β_1 and β_2.

Result: It is not possible to compute from the sample data estimates of ***both*** $\boldsymbol{\beta_1}$ ***and*** $\boldsymbol{\beta_2}$, the separate linear effects of X_{1i} and X_{2i} on the regressand Y_i.

2. Alternatively, substitute for X_{2i} in PRE (2) the equivalent expression $X_{2i} = \frac{X_{1i}}{3}$:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \\ &= \beta_0 + \beta_1 X_{1i} + \beta_2 \left(\frac{X_{1i}}{3} \right) + u_i \\ &= \beta_0 + \beta_1 X_{1i} + \frac{\beta_2}{3} X_{1i} + u_i \\ &= \beta_0 + \left(\beta_1 + \frac{\beta_2}{3} \right) X_{1i} + u_i \\ &= \beta_0 + \alpha_1 X_{1i} + u_i \qquad \text{where } \alpha_1 = \beta_1 + \frac{\beta_2}{3}. \qquad (4b) \end{aligned}$$

- It is possible to estimate from the sample data the regression coefficients β_0 and α_1.

- But from the estimate of α_1 it is not possible to compute estimates of the coefficients β_1 and β_2. *__Reason__:* The equation

$$\alpha_1 = \beta_1 + \frac{\beta_2}{3}$$

is *one* equation containing *two* unknowns, namely β_1 and β_2.

__Result__: Again, it is not possible to compute from the sample data estimates of ***both*** $\boldsymbol{\beta_1}$ ***and*** $\boldsymbol{\beta_2}$, the separate linear effects of X_{1i} and X_{2i} on the regressand Y_i.

5. Interpreting Slope Coefficients in Multiple Linear Regression Models

- Consider the multiple linear regression model given by the following **population regression equation (PRE)**:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad (5)$$

X_1, X_2 and X_3 are three distinct independent or explanatory variables that determine the population values of Y.

Because regression equation (5) contains more than one regressor, it is called a ***multiple*** **linear regression model**.

- The **population regression function (PRF)** corresponding to PRE (5) is:

$$E(Y_i \mid \underline{x}_i) = E(Y_i \mid X_{1i}, X_{2i}, X_{3i}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} \qquad (6)$$

where $\underline{x}_i$ is the 1×4 row vector of regressors: $\underline{x}_i = (1\ X_{1i}\ X_{2i}\ X_{3i})$.

❑ Interpreting the Slope Coefficients in Multiple Regression Model (5)

- Each ***slope*** **coefficient β_j** is the ***marginal*** **effect** of the corresponding explanatory variable X_j on the conditional mean of Y. Formally, the ***slope*** **coefficients {β_j : j = 1, 2, 3}** are the ***partial derivatives*** **of the population regression function (PRF) with respect to the explanatory variables {X_j : j = 1, 2, 3}**:

$$\frac{\partial E(Y_i \mid \underline{x}_i)}{\partial X_{ji}} = \frac{\partial E(Y_i \mid X_{1i}, X_{2i}, X_{3i})}{\partial X_{ji}} = \beta_j \qquad j = 1, 2, 3 \qquad (7)$$

For example, for j = 1 in multiple regression model (5):

$$\frac{\partial E(Y_i \mid X_{1i}, X_{2i}, X_{3i})}{\partial X_{1i}} = \frac{\partial(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i})}{\partial X_{1i}} = \beta_1 \qquad (8)$$

- ***Interpretation:*** A ***partial* derivative** isolates the marginal effect on the conditional mean of Y of small variations in one of the explanatory variables, while ***holding constant* the values of the *other* explanatory variables in the PRF**.

Example: In multiple regression model (5)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad (5)$$

with population regression function

$$E(Y_i \mid X_{1i}, X_{2i}, X_{3i}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} \qquad (6)$$

the ***slope* coefficients β_1, β_2 and β_3** are interpreted as follows:

β_1 = the *partial* marginal effect of X_1 on the conditional mean of Y *holding constant* the values of the other regressors X_2 and X_3.

β_2 = the *partial* marginal effect of X_2 on the conditional mean of Y *holding constant* the values of the other regressors X_1 and X_3.

β_3 = the *partial* marginal effect of X_3 on the conditional mean of Y *holding constant* the values of the other regressors X_1 and X_2.

- Including X_2 and X_3 in the regression function allows us to estimate the partial marginal effect of X_1 on $E(Y \mid X_1, X_2, X_3)$ while
  - **holding constant** the values of X_2 and X_3
  - **controlling for** the effects on Y of X_2 and X_3
  - **conditioning** on X_2 and X_3.

❑ Interpreting the Slope Coefficient β_1 in Multiple Regression Model (5)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad (5)$$

$$E(Y_i \mid X_{1i}, X_{2i}, X_{3i}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} \quad (6)$$

- Denote the ***initial* values of the explanatory variables X_1, X_2 and X_3** as X_{10}, X_{20} and X_{30}.

  The ***initial* value of the population regression function for Y** for the initial values of X_1, X_2 and X_3 is:

$$E(Y \mid X_{10}, X_{20}, X_{30}) = \beta_0 + \beta_1 X_{10} + \beta_2 X_{20} + \beta_3 X_{30} \quad (9)$$

- Now change the value of the explanatory variable X_1 by ΔX_1, while holding constant the values of the other two explanatory variables X_2 and X_3 at their initial values X_{20} and X_{30}.

  The ***new* value of X_1** is therefore

$$X_{11} = X_{10} + \Delta X_1$$

  The ***change* in the value of X_1** is thus

$$\Delta X_1 = X_{11} - X_{10}$$

  The ***new* value of the population regression function for Y** at the new value of the explanatory variable X_1 is:

$$\begin{aligned} E(Y \mid X_{11}, X_{20}, X_{30}) &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{20} + \beta_3 X_{30} \\ &= \beta_0 + \beta_1 (X_{10} + \Delta X_1) + \beta_2 X_{20} + \beta_3 X_{30} \\ &= \beta_0 + \beta_1 X_{10} + \beta_1 \Delta X_1 + \beta_2 X_{20} + \beta_3 X_{30} \quad (10) \end{aligned}$$

- The ***change* in the conditional mean value of Y** associated with the change ΔX_1 in the value of X_1 is obtained by subtracting the initial value of the population regression function given by (9) from the new value of the population regression function given by (10):

$$
\begin{aligned}
\Delta E(Y \mid X_1, X_2, X_3) &= E(Y \mid X_{11}, X_{20}, X_{30}) - E(Y \mid X_{10}, X_{20}, X_{30}) \\
&= \beta_0 + \beta_1 X_{10} + \beta_1 \Delta X_1 + \beta_2 X_{20} + \beta_3 X_{30} \\
&\quad - (\beta_0 + \beta_1 X_{10} + \beta_2 X_{20} + \beta_3 X_{30}) \\
&= \beta_0 + \beta_1 X_{10} + \beta_1 \Delta X_1 + \beta_2 X_{20} + \beta_3 X_{30} \\
&\quad - \beta_0 - \beta_1 X_{10} - \beta_2 X_{20} - \beta_3 X_{30} \\
&= \beta_1 \Delta X_1 \qquad (11)
\end{aligned}
$$

- The ***interpretation* of the slope coefficient β_1** is obtained by solving for β_1 in (11):

$$\beta_1 = \left(\frac{\Delta E(Y \mid X_1, X_2, X_3)}{\Delta X_1} \right)_{\Delta X_2 = 0,\ \Delta X_3 = 0} = \frac{\partial E(Y \mid X_1, X_2, X_3)}{\partial X_1}$$

β_1 = the *partial* marginal effect of X_1 on the conditional mean of Y *holding constant* the values of the other regressors X_2 and X_3.

❑ Comparing Slope Coefficients in Simple and Multiple Regression Models

- Compare the ***multiple* linear regression model**

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad (5)$$

with the ***simple* linear regression model**

$$Y_i = \beta_0 + \beta_1 X_{1i} + u_i \qquad (12)$$

- ***Question:*** What is the difference between the slope coefficient β_1 in these two regression models?

- ***Answer:*** Compare the population regression functions for these two models.

For the multiple regression model (5), the population regression function is

$$E(Y \mid X_{1i}, X_{2i}, X_{3i}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} \tag{6}$$

As we have seen, the slope coefficient β_1 in multiple regression model (5) is

$$\beta_1 \text{ in model (5)} = \left(\frac{\Delta E(Y \mid X_1, X_2, X_3)}{\Delta X_1} \right)_{\Delta X_2 = 0,\ \Delta X_3 = 0} = \frac{\partial E(Y \mid X_1, X_2, X_3)}{\partial X_1}$$

For the simple regression model (12), the population regression function is

$$E(Y \mid X_{1i}) = \beta_0 + \beta_1 X_{1i}$$

The slope coefficient β_1 in simple regression model (12) is

$$\beta_1 \text{ in model (12)} = \frac{\Delta E(Y \mid X_1)}{\Delta X_1} = \frac{d\,E(Y \mid X_1)}{d\,X_1}$$

- ***Compare β_1 in model (5) with β_1 in model (12)***

β_1 in multiple regression model (5) controls for – or accounts for – the effects of X_2 and X_3 on the conditional mean value of the dependent variable Y.

β_1 in *multiple* regression model (5) is therefore referred to as the ***adjusted* marginal effect of X_1 on Y**.

β_1 in simple regression model (12) does not control for – or account for – the effects of X_2 and X_3 on the conditional mean value of the dependent variable Y.

β_1 in *simple* regression model (12) is therefore referred to as the ***unadjusted* marginal effect of X_1 on Y**.