ECON 452* -- NOTE 1

Formulation and Specification of the Multiple Linear Regression Model in Vector-Matrix Notation

The population regression equation, or PRE, for the multiple linear regression model can be written in three alternative but equivalent forms:

(1) *scalar* formulation;

(2) *vector* formulation;

(3) *matrix* formulation.

1. Scalar Formulation of the PRE

- **Without observation subscripts**. The scalar formulation of the PRE is written as:

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u \\ &= \beta_0 + \sum_{j=1}^{j=k} \beta_j X_j + u \\ &= \sum_{j=0}^{j=k} \beta_j X_j + u \,, \qquad X_0 = 1 \end{aligned} \tag{1.1}$$

where

$Y \equiv$ the **regressand**, or ***dependent*** **variable**;

$X_j \equiv$ the **j-th regressor**, or ***independent*** **variable**;

$\beta_j \equiv$ the **partial regression coefficient of X_j**;

$u \equiv$ the ***unobservable*** **random error term**.

- **With observation subscripts**. For the i-th observation, the **scalar formulation of the PRE** is written as:

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + u_i \qquad \forall\, i \\
&= \beta_0 + \sum_{j=1}^{j=k} \beta_j X_{ij} + u_i \qquad \forall\, i \\
&= \sum_{j=0}^{j=k} \beta_j X_{ij} + u_i \,, \quad X_{i0} = 1 \;\forall\, i \qquad \forall\, i
\end{aligned}
\tag{1.2}
$$

where

$Y_i \equiv$ the i-th population value of the regressand, or dependent variable;

$X_{ij} \equiv$ the i-th population value of the j-th regressor, or independent variable;

$\beta_j \equiv$ the partial regression coefficient of X_{ij};

$u_i \equiv$ the i-th population value of the *unobservable* random error term.

Note:

(1) **Lower case "k"** denotes the **number of *slope* coefficients** in the PRF.

(2) **Upper case "K"** denotes the ***total* number of regression coefficients** in the PRF.

(3) Therefore: **K = k + 1**.

2. Vector Formulation of the PRE

- Define the following two vectors:

$$x_i^T = \begin{bmatrix} 1 & X_{i1} & X_{i2} & \cdots & X_{ik} \end{bmatrix}$$

= the 1×K or 1×(k+1) row vector of population values of the regressors for observation i, called the **regressor vector for observation i**;

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

= the K×1 or (k+1)×1 column vector of partial regression coefficients β_j, j = 0, 1, ..., k, called the (population) **coefficient vector**.

Note:

(1) **Lower case "k"** denotes the **number of *slope* coefficients** in the PRF.

(2) **Upper case "K"** denotes the ***total* number of regression coefficients** in the PRF.

(3) Therefore: **K = k + 1**.

The vector product $x_i^T\beta$ takes the form

$$x_i^T\beta = \begin{bmatrix} 1 & X_{i1} & X_{i2} & \cdots & X_{ik} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}$$

Note: The vector product $\beta^T x_i = x_i^T\beta = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}$:

$$\beta^T x_i = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \cdots & \beta_k \end{bmatrix} \begin{bmatrix} 1 \\ X_{i1} \\ X_{i2} \\ \vdots \\ X_{ik} \end{bmatrix} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}$$

- **Result:** The **vector form of the PRE for the i-th population observation** is written as:

$$Y_i = x_i^T\beta + u_i \qquad \textit{or} \qquad Y_i = \beta^T x_i + u_i \qquad \forall\ i. \qquad \textbf{(2)}$$

3. Matrix Formulation of the PRE

- The matrix formulation of the PRE is a compact way of writing the population regression equation for a sample of N observations from the relevant population of *all* observations.

- Define the following two vectors:

$$y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_N \end{bmatrix} = \text{the } \mathbf{Nx1 \text{ regressand vector}}$$

= the N×1 column vector of observed sample values of the regressand, or dependent variable, Y_i (i = 1, ..., N);

$$u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_N \end{bmatrix} = \text{the } \mathbf{Nx1 \text{ error vector}}$$

= the N×1 column vector of unobserved random error terms u_i (i = 1, ..., N) corresponding to each of the N sample observations.

- Define the following **NxK regressor matrix X**:

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ x_3^T \\ \vdots \\ x_N^T \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ 1 & X_{31} & X_{32} & \cdots & X_{3k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{N1} & X_{N2} & \cdots & X_{Nk} \end{bmatrix} = \text{ the } \mathbf{N \times K \text{ regressor matrix}}$$

= the N×K matrix of observed sample values of the K = k + 1 regressors, where the first regressor is a constant equal to 1 for all observations ($X_{i0} = 1 \; \forall \; i = 1, ..., N$) and the remaining k = K – 1 regressors $X_{i1}, X_{i2}, ..., X_{ik}$ (i = 1, ..., N) are variables.

The N×K regressor matrix X can be viewed as being assembled in either of two ways: (1) by stacking the N row vectors of K regressor values $\{ x_i^T : i = 1, \ldots, N\}$ for each observation, or (2) by merging the K = k+1 column vectors of N regressor values $\{ x_j : j = 0, \ldots, k\}$ for each regressor (including the intercept constant).

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ x_3^T \\ \vdots \\ x_N^T \end{bmatrix} = \begin{bmatrix} x_0 & x_1 & x_2 & \cdots & x_k \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ 1 & X_{31} & X_{32} & \cdots & X_{3k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{N1} & X_{N2} & \cdots & X_{Nk} \end{bmatrix}$$

(1) **Each row of the regressor matrix X** contains the values of all the regressors X_{i0}, X_{i1}, X_{i2}, ..., X_{ik} for a given sample observation. Since there are N sample observations, X has N rows.

The **i-th row of regressor matrix X** is the 1×K row vector

$$x_i^T = \begin{bmatrix} X_{i0} & X_{i1} & X_{i2} & \cdots & X_{ik} \end{bmatrix} \qquad \text{where} \quad i = 1, \ldots, N.$$

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ x_3^T \\ \vdots \\ x_N^T \end{bmatrix} = \begin{bmatrix} x_0 & x_1 & x_2 & \cdots & x_k \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ 1 & X_{31} & X_{32} & \cdots & X_{3k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{N1} & X_{N2} & \cdots & X_{Nk} \end{bmatrix}$$

(2) **Each column of the regressor matrix X** contains the full set of N sample values for one of the K regressors -- i.e., the j-th column of X contains the sample values X_{ij} (i = 1, ..., N) where j = 0, 1, 2, ..., k. Since there are K regressors (including the intercept constant X_{i0}), X has K = k + 1 columns.

The **j-th column of regressor matrix X** is the N×1 column vector

$$x_j = \begin{bmatrix} X_{1j} \\ X_{2j} \\ X_{3j} \\ \vdots \\ X_{Nj} \end{bmatrix} \qquad \text{where} \quad j = 0, 1, \ldots, k \text{ (and } X_{i0} = 1 \; \forall \text{ i for } j = 0).$$

- **Result: The matrix form of the PRE** for the **full set of N sample observations** is written as:

$$y = X\beta + u\,. \qquad \textbf{(3)}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ 1 & X_{31} & X_{32} & \cdots & X_{3k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{N1} & X_{N2} & \cdots & X_{Nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_N \end{bmatrix}$$

Assumptions of the Classical Linear Regression (CLR) Model

A1 Formulation of the Population Regression Equation (PRE)

The population regression equation, or PRE, takes the form

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + u_i && \forall\, i \\ &= \beta_0 + \sum_{j=1}^{j=k} \beta_j X_{ij} + u_i && \forall\, i \qquad \textbf{(1)} \\ &= \sum_{j=0}^{j=k} \beta_j X_{ij} + u_i \qquad \text{where } X_{i0} = 1 \ \forall\, i && \forall\, i \end{aligned}$$

or

$$Y_i = x_i^T \beta + u_i \qquad \forall\, i \qquad \textbf{(2)}$$

or

$$y = X\beta + u\,. \qquad \textbf{(3)}$$

The population regression equation (PRE) represented by (1), (2), or (3) incorporates ***three*** **distinct assumptions**.

A1.1 Assumption of an Additive Random Error Term

The random error term u_i enters the PRE additively.

Technically, this assumption means that the partial derivative of Y_i with respect to u_i equals 1: i.e.,

$$\frac{\partial Y_i}{\partial u_i} = 1 \qquad \text{for all i } (\forall \text{ i}). \tag{A1-1}$$

A1.2 Assumption of Linearity-in-Parameters or Linearity-in-Coefficients

The PRE is linear in the population regression coefficients β_j (j = 0,1, ..., k).

This assumption means that the partial derivative of Y_i with respect to each of the regression coefficients is a function only of known constants and/or the regressor vector x_i^T; it is not a function of any unknown parameters.

$$\frac{\partial Y_i}{\partial \beta_j} = f_j(x_i^T) \qquad \textit{or} \qquad \frac{\partial E(Y_i|x_i)}{\partial \beta_j} = f_j(x_i^T) \quad j = 0, 1, ..., k \qquad \textbf{(A1-2)}$$

where the functions $f_j(x_i^T)$ **contain no *unknown* parameters**.

A1.3 Assumption of Parameter or Coefficient Constancy

The population regression coefficients β_j (j = 0,1, ..., k) are (unknown) constants.

This assumption means that the unknown **regression coefficients β_j** ***do not vary across observations*** -- i.e., do not vary with the observation subscript "i".

Symbolically, if β_{ij} is the value of the j-th regression coefficient for observation i, then assumption A1.3 states that

$$\beta_{ij} = \beta_j = \text{a constant for all i} \quad (j = 0,1, ..., k). \qquad \textbf{(A1-3)}$$

Alternatively, in vector notation, if β_i is the value of the regression coefficient vector β for observation i, then assumption (A1.3) states that

$$\beta_i = \beta = \text{a vector of constants for all i.} \qquad \textbf{(A1-3)}$$

A2 The Assumption of Zero Conditional Mean Error

The *conditional mean*, or conditional expectation, of the population random errors u_i for any given vector of regressor values x_i^T is equal to *zero*.

- **Scalar formulation of A2**

Without observation subscripts

$$E\left(u \middle| x^T\right) = E\left(u \middle| 1, X_1, X_2, \ldots, X_k\right) = 0 \qquad \text{for all } x^T \qquad \textbf{(A2)}$$

where $x^T = \begin{bmatrix} 1 & X_1 & X_2 & \cdots & X_k \end{bmatrix}$ is any given vector of regressor values.

With observation subscripts

$$E\left(u_i \middle| x_i^T\right) = E\left(u_i \middle| 1, X_{i1}, X_{i2}, \ldots, X_{ik}\right) = 0 \qquad \forall\ i,\ \text{i.e., for all } x_i^T \qquad \textbf{(A2)}$$

where $x_i^T = \begin{bmatrix} 1 & X_{i1} & X_{i2} & \cdots & X_{ik} \end{bmatrix}$ is the (row) vector of regressor values for observation i.

- **Matrix formulation of A2**

$$E(u \mid X) = \begin{bmatrix} E(u_1 \mid X) \\ E(u_2 \mid X) \\ E(u_3 \mid X) \\ \vdots \\ E(u_N \mid X) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \underline{0} = \text{ an N}\times\text{1 vector of zeros.} \qquad \textbf{(A2)}$$

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

$$E(u \mid x^T) = 0 \quad \Rightarrow \quad E(u) = 0 \qquad \textbf{(A2.1)}$$

or

$$E(u_i \mid x_i^T) = 0 \quad \Rightarrow \quad E(u_i) = 0 \quad \forall\, i. \qquad \textbf{(A2.1)}$$

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

  - The logic of (A2.1) is straightforward: If the conditional mean of u for each and every set of regressor values x^T 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_{ij} 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*:**

$$E\left(u \,\middle|\, x^T\right) = 0 \quad \Rightarrow \quad Cov\left(X_j, u\right) = E\left(X_j\, u\right) = 0,\ j = 1, 2, \ldots, k \qquad \textbf{(A2.2)}$$

or

$$E\left(u_i \,\middle|\, x_i^T\right) = 0 \ \Rightarrow \ Cov\left(X_{ij}, u_i\right) = E\left(X_{ij}\, u_i\right) = 0 \ \forall\ i,\ j = 1, 2, \ldots, k \qquad \textbf{(A2.2)}$$

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

$$\begin{aligned}
Cov\left(X_{ij}, u_i\right) &\equiv E\left\{\left[X_{ij} - E(X_{ij})\right]\left[u_i - E(u_i \,|\, x_i^T)\right]\right\} && \text{by definition} \\
&= E\left\{\left[X_{ij} - E(X_{ij})\right] u_i\right\} && \text{since } E(u_i \,|\, x_i^T) = 0 \text{ by A2} \\
&= E\left[X_{ij} u_i - E(X_{ij})\, u_i\right] \\
&= E(X_{ij} u_i) - E(X_{ij}) E(u_i) && \text{since } E(X_{ij}) \text{ is a constant} \\
&= E(X_{ij} u_i) && \text{since } E(u_i) = E(u_i \,|\, x_i^T) = 0 \text{ by A2.}
\end{aligned}$$

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, 2, …, 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: ***Zero covariance*** **between X_{ij} and u_i** implies ***zero correlation*** **between X_{ij} and u_i**, since the **simple *correlation coefficient* between X_{ij} and u_i**, denoted as $\rho(X_{ij}, u_i)$, is defined as

$$\rho(X_{ij}, u_i) \equiv \frac{\text{Cov}(X_{ij}, u_i)}{\sqrt{\text{Var}(X_{ij})\text{Var}(u_i)}} = \frac{\text{Cov}(X_{ij}, u_i)}{\text{sd}(X_{ij})\text{sd}(u_i)}. \qquad \textbf{(4)}$$

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

- **<u>Scalar formulation of A2.2</u>**

For j = 0: $$E\left[\sum_{i=1}^{N} X_{i0}u_i\right] = E\left[\sum_{i=1}^{N} u_i\right] = \sum_{i=1}^{N} E(u_i) = 0 \qquad \text{where } X_{i0} = 1 \ \forall \ i.$$

For j = 1: $$E\left[\sum_{i=1}^{N} X_{i1}u_i\right] = \sum_{i=1}^{N} E(X_{i1}u_i) = 0.$$

For j = 2: $$E\left[\sum_{i=1}^{N} X_{i2}u_i\right] = \sum_{i=1}^{N} E(X_{i2}u_i) = 0.$$

. . .

For j = k: $$E\left[\sum_{i=1}^{N} X_{ik}u_i\right] = \sum_{i=1}^{N} E(X_{ik}u_i) = 0.$$

- **Matrix formulation of A2.2**

- ***Question:*** What vector-matrix product yields a K×1 vector with elements

$$\sum_{i=1}^{N} u_i , \quad \sum_{i=1}^{N} X_{i1} u_i , \quad \sum_{i=1}^{N} X_{i2} u_i , \quad , \ldots , \quad \sum_{i=1}^{N} X_{ik} u_i \, ?$$

- ***Answer:*** The matrix product $X^T u$.

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ 1 & X_{31} & X_{32} & \cdots & X_{3k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{N1} & X_{N2} & \cdots & X_{Nk} \end{bmatrix} \quad \Rightarrow \quad X^T = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_{11} & X_{21} & X_{31} & \cdots & X_{N1} \\ X_{12} & X_{22} & X_{32} & \cdots & X_{N2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{1k} & X_{2k} & X_{3k} & \cdots & X_{Nk} \end{bmatrix}$$

Since X is an N×K matrix, its transpose X^T is a K×N matrix, where K = k+1.

Therefore, the matrix-vector product $X^T u$ is:

$$X^T u = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_{11} & X_{21} & X_{31} & \cdots & X_{N1} \\ X_{12} & X_{22} & X_{32} & \cdots & X_{N2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{1k} & X_{2k} & X_{3k} & \cdots & X_{Nk} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_N \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} u_i \\ \sum_{i=1}^{N} X_{i1} u_i \\ \sum_{i=1}^{N} X_{i2} u_i \\ \vdots \\ \sum_{i=1}^{N} X_{ik} u_i \end{bmatrix}. \quad \textbf{(5)}$$

- **<u>Result</u>: Implication A2.2 of Assumption A2 can be written in vector-matrix form as:**

$E(X^T u) = \underline{0}$, where $\underline{0}$ is a **K×1 *null* vector** (a vector of zeros).

$$E(X^T u) = E\begin{bmatrix} \sum_{i=1}^{N} u_i \\ \sum_{i=1}^{N} X_{i1} u_i \\ \sum_{i=1}^{N} X_{i2} u_i \\ \vdots \\ \sum_{i=1}^{N} X_{ik} u_i \end{bmatrix} = \begin{bmatrix} E\left(\sum_{i=1}^{N} u_i\right) \\ E\left(\sum_{i=1}^{N} X_{i1} u_i\right) \\ E\left(\sum_{i=1}^{N} X_{i2} u_i\right) \\ \vdots \\ E\left(\sum_{i=1}^{N} X_{ik} u_i\right) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} E(u_i) \\ \sum_{i=1}^{N} E(X_{i1} u_i) \\ \sum_{i=1}^{N} E(X_{i2} u_i) \\ \vdots \\ \sum_{i=1}^{N} E(X_{ik} u_i) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \underline{0}. \quad \textbf{(6)}$$

- **Implication 3 of A2.** Assumption A2 implies that the ***conditional mean*** **of the population Y values** corresponding to any given vector of regressor values x_i^T ***equals the corresponding PRF*** $x_i^T \beta$:

$$E\left(u \middle| x^T\right) = 0 \;\Rightarrow\; E\left(Y \middle| x^T\right) = f(x,\beta) = x^T\beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \qquad \textbf{(A2.3)}$$

or

$$E\left(u_i \middle| x_i^T\right) = 0 \;\Rightarrow\; E\left(Y_i \middle| x_i^T\right) = f(x_i,\beta) = x_i^T\beta = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} \qquad \textbf{(A2.3)}$$

The function $f(x_i,\beta) = x_i^T\beta$ is called the ***population regression function (PRF)*** or the **population** ***conditional mean function (CMF)***.

***Proof*:** Take the conditional expectation of the PRE (2) for some given value x_i^T of the regressor vector x^T:

$$Y_i = x_i^T\beta + u_i \qquad \forall\, i \qquad \textbf{(2)}$$

Take the expected value of equation (2) conditional on x_i^T:

$$\begin{aligned} E\left(Y_i \middle| x_i^T\right) &= E\left(x_i^T\beta \middle| x_i^T\right) + E\left(u_i \middle| x_i^T\right) \\ &= E\left(x_i^T\beta \middle| x_i^T\right) && \text{since } E\left(u_i \middle| x_i^T\right) = E(u_i) = 0 \text{ by A2} \\ &= x_i^T\beta && \text{because both } x_i^T \text{ and } \beta \text{ are vectors of constants.} \end{aligned}$$

Meaning of Assumption A2: For each and every vector of regressor values x_i^T, the **conditional** ***mean*** of the **population random errors** is **equal to zero**, and the **conditional** ***mean*** of the **population Y values** is **equal to the corresponding PRF** $x_i^T\beta$.

A3 The Assumption of Homoskedastic and Nonautoregressive Errors -- The Assumption of Spherical Errors

A ***two-part*** **assumption** about the ***second*** **moments** of the random error terms u_i.

A3.1 The Assumption of Homoskedastic Errors or Constant Error Variances

Also called the **Assumption of Constant Conditional Error Variances**
or the **Assumption of Homoskedasticity**

The ***conditional*** **variances of the random error terms u_i are *identical* for all observations** -- i.e., for all values of x_i^T -- and equal the same finite positive constant σ^2 for all i.

- **Scalar formulation of A3.1**

$$\text{Var}\left(u \,|\, x^T\right) = E\left(u^2 \,|\, x^T\right) = E\left(u^2 \,|\, 1, X_1, X_2, \ldots, X_k\right) = \sigma^2 > 0 \qquad \textbf{(A3.1)}$$

or

$$\text{Var}\left(u_i \,|\, x_i^T\right) = E\left(u_i^2 \,|\, x_i^T\right) = E\left(u_i^2 \,|\, 1, X_{i1}, X_{i2}, \ldots, X_{ik}\right) = \sigma^2 > 0 \quad \forall \text{ i.} \qquad \textbf{(A3.1)}$$

where σ^2 is a ***finite positive*** **(unknown)** ***constant***.

- The **first equality** $\mathrm{Var}\left(u_i \middle| x_i^T\right) = \mathrm{E}\left(u_i^2 \middle| x_i^T\right)$ **in (A3.1)** follows from the ***definition* of the conditional error variance** and from ***Assumption A2*:**

$$\mathrm{Var}\left(u_i \middle| x_i^T\right) \equiv \mathrm{E}\left(\left[u_i - \mathrm{E}\left(u_i \middle| x_i^T\right)\right]^2 \middle| x_i^T\right) \qquad \text{by definition.} \qquad \textbf{(7)}$$

- But ***Assumption A2*** states that $\mathrm{E}\left(u_i \middle| x_i\right) = \mathrm{E}\left(u_i\right) = 0 \ \forall$ i, so that:

$$\mathrm{Var}\left(u_i \middle| x_i^T\right) \equiv \mathrm{E}\left(\left[u_i - 0\right]^2 \middle| x_i^T\right) = \mathrm{E}\left(u_i^2 \middle| x_i^T\right).$$

- **Implication of A3.1:** Assumption A3.1 implies that the **conditional variance of the population Y values** corresponding to each and every vector of regressor values x_i^T equals the ***constant* error variance σ^2**:

$$\text{Var}\left(u_i \middle| x_i^T\right) = \sigma^2 \quad \forall \quad x_i^T \qquad \Rightarrow \qquad \text{Var}\left(Y_i \middle| x_i^T\right) = \sigma^2 \quad \forall \quad x_i^T \tag{8}$$

Proof: Start with the definition of the conditional variance of Y_i for some given value x_i^T of the regressor vector x^T:

$$\begin{aligned}
\text{Var}\left(Y_i \middle| x_i^T\right) &\equiv E\left(\left[Y_i - E(Y_i|x_i^T)\right]^2 \middle| x_i^T\right) && \text{by definition} \\
&= E\left(\left[Y_i - x_i^T\beta\right]^2 \middle| x_i^T\right) && \text{since } E(Y_i|x_i^T) = x_i^T\beta \text{ by A2} \\
&= E\left(u_i^2 \middle| x_i^T\right) && \text{since } u_i = Y_i - x_i^T\beta \text{ by A1} \\
&= \sigma^2 && \text{since } E\left(u_i^2 \middle| x_i^T\right) = \sigma^2 \text{ by assumption A3.1.}
\end{aligned}$$

- **Meaning of Assumption A3.1:**

For each and every vector of regressor values x_i^T, the **conditional variance** of the **population random errors** and the conditional variance of the **population Y values** equal the same **constant σ^2**.

The conditional distribution of the **population Y values** around the PRF $E\left(Y_i \middle| x_i^T\right) = x_i^T\beta$ for **any *given* vector of regressor values x_i^T** has the **same *variance*, or same *dispersion***, as the conditional distribution of the **population Y values** for **any *other* vector of regressor values x_s^T**, where $x_s^T \neq x_i^T$.

A3.2 The Assumption of Nonautoregressive Errors or Zero Error Covariances

Also called the **Assumption of Zero Conditional Error Covariances**
or the **Assumption of Nonautocorrelated Errors**
or the **Assumption of Serially Uncorrelated Errors**.

- **Scalar formulation of A3.2**

Consider any pair of distinct random error terms $\mathbf{u_i}$ **and** $\mathbf{u_s}$ ($i \neq s$) corresponding to **two different vectors of regressor values** $\mathbf{x_i^T \neq x_s^T}$. This assumption states that $\mathbf{u_i}$ **and** $\mathbf{u_s}$ **have *zero conditional covariance*** – i.e., that the **population random errors** $\mathbf{u_i}$ corresponding to **any *given* vector of regressor values** $\mathbf{x_i^T}$ have ***zero covariance* with**, or are ***uncorrelated* with**, the **population random errors** $\mathbf{u_s}$ corresponding to **any *other* vector of regressor values** $\mathbf{x_s^T}$ where $x_s^T \neq x_i^T$.

$$\mathrm{Cov}\left(u_i, u_s \middle| x_i^T, x_s^T\right) = \mathrm{E}\left(u_i u_s \middle| x_i^T, x_s^T\right) = 0 \quad \forall\, x_i^T \neq x_s^T \qquad \textbf{(A3.2)}$$

$$\text{Cov}\left(u_i, u_s \middle| x_i^T, x_s^T\right) = E\left(u_i u_s \middle| x_i^T, x_s^T\right) = 0 \quad \forall\, x_i^T \neq x_s^T \qquad \textbf{(A3.2)}$$

- The **first equality in (A3.2)** follows from the definition of the conditional covariance of u_i and u_s and Assumption A2:

$$\text{Cov}\left(u_i, u_s \middle| x_i^T, x_s^T\right) \equiv E\left\{\left[u_i - E(u_i|x_i^T)\right]\left[u_s - E(u_s|x_s^T)\right] \middle| x_i^T, x_s^T\right\}$$

Since $E(u_i \mid x_i^T) = E(u_s \mid x_s^T) = 0$ by A2,

$$\text{Cov}\left(u_i, u_s \middle| x_i^T, x_s^T\right) = E\left(u_i u_s \middle| x_i^T, x_s^T\right).$$

- The **second equality in (A3.2)** states the assumption that the error terms corresponding to different sets of regressor values have zero covariance.

- **Implication of A3.2:** Assumption A3.2 implies that the **population Y values $\mathbf{Y_i}$** corresponding to **any *given* regressor vector $\mathbf{x_i^T}$** have ***zero* covariance with**, or **are *uncorrelated* with**, the **population Y values $\mathbf{Y_s}$** corresponding to ***any other* regressor vector $\mathbf{x_s^T}$**, where $\mathbf{x_s^T \neq x_i^T}$.

$$\mathrm{Cov}\left(u_i, u_s \middle| x_i^T, x_s^T\right) = 0 \quad \forall\, x_i^T \neq x_s^T \qquad \Rightarrow \quad \mathrm{Cov}\left(Y_i, Y_s \middle| x_i^T, x_s^T\right) = 0 \quad \forall\, x_i^T \neq x_s^T . \tag{9}$$

Proof:

Show that $\mathrm{Cov}\left(Y_i, Y_s \middle| x_i^T, x_s^T\right) = \mathrm{E}\left(u_i u_s \middle| x_i^T, x_s^T\right) = \mathrm{Cov}\left(u_i, u_s \middle| x_i^T, x_s^T\right)$.

1. Begin with the definition of the conditional covariance for Y_i and Y_s for given x_i^T and x_s^T values where $x_i^T \neq x_s^T$:

$$\begin{aligned}\mathrm{Cov}\left(Y_i, Y_s \middle| x_i^T, x_s^T\right) &\equiv \mathrm{E}\left\{\left[Y_i - \mathrm{E}(Y_i \,|\, x_i^T)\right]\left[Y_s - \mathrm{E}(Y_s \,|\, x_s^T)\right] \middle| x_i^T, x_s^T\right\} \\ &= \mathrm{E}\left(u_i u_s \middle| x_i^T, x_s^T\right)\end{aligned}$$

since

$$Y_i - \mathrm{E}\left(Y_i \middle| x_i^T\right) = Y_i - x_i^T\beta = u_i \qquad \text{and} \qquad Y_s - \mathrm{E}\left(Y_s \middle| x_s^T\right) = Y_s - x_s^T\beta = u_s \qquad \text{by assumption A1.}$$

2. Therefore

$$\mathrm{Cov}\left(Y_i, Y_s \middle| x_i^T, x_s^T\right) = \mathrm{E}\left(u_i u_s \middle| x_i^T, x_s^T\right) = 0 \qquad \text{by assumption A3.2.}$$

- **Meaning of Assumption A3.2:** Assumption A3.2 means that there is **no *correlation*, or no systematic *linear association*, between $\mathbf{u_i}$ and $\mathbf{u_s}$, or between $\mathbf{Y_i}$ and $\mathbf{Y_s}$,** where i and s correspond to **different observations** (i.e., to **different sets or vectors of regressor values $\mathbf{x_i^T \neq x_s^T}$**).

  1. The conditional distribution of **population random errors $\mathbf{u_i}$** corresponding to **any *given* vector** of regressor values x_i^T has ***zero* covariance with**, or **is *uncorrelated* with**, the conditional distribution of **population random errors $\mathbf{u_s}$** corresponding to **any *other* vector** of regressor values x_s^T, where $x_s^T \neq x_i^T$.

  2. The conditional distribution of the **population Y values $\mathbf{Y_i}$** for **any *given* regressor vector** x_i^T has ***zero* covariance with** the conditional distribution of the **population $\mathbf{Y_s}$ values** for ***any other* regressor vector** x_s^T, where $x_s^T \neq x_i^T$.

- **Relationship Between A3.2 and Assumption of Random Sampling**

  - The **assumption of *zero covariance***, or ***zero correlation***, between each pair of distinct observations is ***weaker*** than the **assumption of *independent random sampling A4*** 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.

A3 Matrix Formulation of Assumptions A3.1 and A3.2

Assumptions A3.1 (homoskedastic errors) and A3.2 (nonautoregressive errors) together specify the form of the **variance-covariance matrix of the error vector u**, which is commonly called the **error covariance matrix**.

- **General Definition of the Variance-Covariance Matrix of Error Vector u**

  The variance-covariance matrix of the N×1 error vector u for any given observation matrix X is defined as:

$$V(u \mid X) \equiv E\left\{[u - E(u \mid X)][u - E(u \mid X)]^T \mid X\right\}. \qquad \textbf{(10)}$$

- **General Form of the Error Covariance Matrix Under Assumption A2**

- Assumption A2 states that $E(u \mid X) = E(u) = \underline{0}$, where $\underline{0}$ is an N×1 vector of zeroes, so that the error covariance matrix simplifies to

$$V(u \mid X) = E\left(uu^T \mid X\right) \qquad \textbf{(11)}$$

  where uu^T is an **N×N (square) symmetric matrix** known as the ***outer product* of the vector u**.

- The form of the N×N outer product matrix uu^T is:

$$uu^T = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_N \end{bmatrix} \begin{bmatrix} u_1 & u_2 & u_3 & \cdots & u_N \end{bmatrix} = \begin{bmatrix} u_1^2 & u_1u_2 & u_1u_3 & \cdots & u_1u_N \\ u_2u_1 & u_2^2 & u_2u_3 & \cdots & u_2u_N \\ u_3u_1 & u_3u_2 & u_3^2 & \cdots & u_3u_N \\ \vdots & \vdots & \vdots & & \vdots \\ u_Nu_1 & u_Nu_2 & u_Nu_3 & \cdots & u_N^2 \end{bmatrix} \quad \textbf{(12)}$$

1. The ***diagonal* elements** of uu^T are the squares of the individual error terms, $\{u_i^2 : i = 1, ..., N\}$.

2. The ***off-diagonal* elements** of uu^T are the products of all pairs of error terms for different observations $\{u_iu_s : i \neq s\}$.

3. The **outer product** uu^T is a ***symmetric* matrix** because $u_iu_s = u_su_i \ \forall\ i \neq s$, where

   u_iu_s is the element in row i and column s of uu^T;
   u_su_i is the element in row s and column i of uu^T.

- To obtain the **conditional error covariance matrix**, take the expectation of the outer product uu^T for any given value of the regressor matrix X:

$$V(u|X) = E(uu^T|X) = \begin{bmatrix} E(u_1^2) & E(u_1u_2) & E(u_1u_3) & \cdots & E(u_1u_N) \\ E(u_2u_1) & E(u_2^2) & E(u_2u_3) & \cdots & E(u_2u_N) \\ E(u_3u_1) & E(u_3u_2) & E(u_3^2) & \cdots & E(u_3u_N) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ E(u_Nu_1) & E(u_Nu_2) & E(u_Nu_3) & \cdots & E(u_N^2) \end{bmatrix}$$

$$= \begin{bmatrix} Var(u_1) & Cov(u_1,u_2) & Cov(u_1,u_3) & \cdots & Cov(u_1,u_N) \\ Cov(u_2,u_1) & Var(u_2) & Cov(u_2,u_3) & \cdots & Cov(u_2,u_N) \\ Cov(u_3,u_1) & Cov(u_3,u_2) & Var(u_3) & \cdots & Cov(u_3,u_N) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Cov(u_N,u_1) & Cov(u_N,u_2) & Cov(u_N,u_3) & \cdots & Var(u_N) \end{bmatrix} \quad \textbf{(13)}$$

$$V(u \mid X) = \begin{bmatrix} \mathrm{Var}(u_1) & \mathrm{Cov}(u_1, u_2) & \mathrm{Cov}(u_1, u_3) & \cdots & \mathrm{Cov}(u_1, u_N) \\ \mathrm{Cov}(u_2, u_1) & \mathrm{Var}(u_2) & \mathrm{Cov}(u_2, u_3) & \cdots & \mathrm{Cov}(u_2, u_N) \\ \mathrm{Cov}(u_3, u_1) & \mathrm{Cov}(u_3, u_2) & \mathrm{Var}(u_3) & \cdots & \mathrm{Cov}(u_3, u_N) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(u_N, u_1) & \mathrm{Cov}(u_N, u_2) & \mathrm{Cov}(u_N, u_3) & \cdots & \mathrm{Var}(u_N) \end{bmatrix} \qquad \textbf{(13)}$$

1. The ***diagonal*** **elements** of error covariance matrix $V(u \mid X)$ are the **conditional error *variances*** of the random error terms u_i:

   $$\mathrm{Var}(u_i) = E(u_i^2) = E(u_i^2 \mid X) \qquad \text{for all i.}$$

2. The ***off-diagonal*** **elements** of error covariance matrix $V(u \mid X)$ are the **conditional error *covariances*** for all distinct pairs of random error terms u_i and u_s for $i \neq s$ and are symmetric in i and s:

   $$\mathrm{Cov}(u_i, u_s) = E(u_i u_s) = E(u_i u_s \mid X) = E(u_s u_i \mid X) = \mathrm{Cov}(u_s, u_i) \quad \text{for } i \neq s.$$

Note: In writing the general error covariance matrix (13) we have suppressed the conditioning operator to keep the notation simpler.

- **Form of the Error Variance-Covariance Matrix Under A3.1 and A3.2**

- ***Assumption A3.1 (homoskedastic errors)*** states that

$$\mathrm{Var}(u_i) = \mathrm{E}(u_i^2) = \sigma^2 \qquad \forall \ \mathrm{i}. \tag{A3.1}$$

Implication: All the diagonal elements of $\mathrm{V}(u \mid X)$ equal the positive constant σ^2.

- ***Assumption A3.2 (nonautoregressive errors)*** states that

$$\mathrm{Cov}(u_i, u_s) = \mathrm{E}(u_i u_s) = \mathrm{E}(u_s u_i) = \mathrm{Cov}(u_s, u_i) = 0 \qquad \forall \ i \neq s. \tag{A3.2}$$

Implication: All the off diagonal elements of $\mathrm{V}(u \mid X)$ equal zero.

- In the general error covariance matrix (13), substitute:

$\mathrm{Var}(u_i) = \sigma^2$ $(i = 1, ..., N)$ for the ***principal diagonal*** elements

and

$\mathrm{Cov}(u_i, u_s) = 0$ $(i \neq s)$ for all ***off-diagonal*** elements.

- **Result:** The **error covariance matrix implied by assumptions A3.1 and A3.2** takes the form

$$V(u) = E\left(uu^T\right) = \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \sigma^2 I_N \qquad \textbf{(14)}$$

where $\mathbf{I_N}$ is an **N×N identity matrix** with 1s along the principal diagonal and 0s in all the off-diagonal cells.

❑ Implication of Assumptions A3.1 and A3.2

Assumptions A3.1 and A3.2 imply that the **conditional variance-covariance matrix of the regressand vector y** is equal to $\sigma^2\mathbf{I}_N$:

$$V(y|X) = E(uu^T|X) = V(u|X) = \sigma^2 I_N. \qquad \textbf{(15)}$$

***Proof*:**

1. By definition, the conditional variance-covariance matrix of y is:

$$V(y|X) \equiv E\left\{[y - E(y|X)][y - E(y|X)]^T|X\right\}.$$

2. But since $y = X\beta + u$ by assumption A1 and $E(y|X) = X\beta$ by assumption A2, it follows that

$$y - E(y|X) = X\beta + u - X\beta = u.$$

3. Therefore, assumption A2 implies that

$$V(y|X) \equiv E\left\{[y - E(y|X)][y - E(y|X)]^T|X\right\} = E(uu^T|X) = V(u|X).$$

4. But assumption A3 states that $V(u|X) = E(uu^T|X) = \sigma^2 I_N$, which implies that

$$V(y|X) = V(u|X) = \sigma^2 I_N.$$

Summary of Assumptions A1 to A3: Specify the assumed statistical properties of the PRE

A1: The population regression equation (PRE) for population member i takes the form

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + u_i = \beta_0 + \sum_{j=1}^{j=k} \beta_j X_{ij} + u_i \quad \forall\, i \qquad \textbf{(1)}$$

or

$$Y_i = x_i^T \beta + u_i \quad \forall\, i \qquad \textbf{(2)}$$

For a sample of N observations $\left(Y_i \;\; x_i^T\right)$ $i = 1, \ldots, N$, the PRE is written in matrix form as

$$y = X\beta + u \qquad \textbf{(3)}$$

A2: Zero conditional error *means*

$$E\left(u_i \middle| x_i^T\right) = E\left(u_i \middle| 1, X_{i1}, X_{i2}, \ldots, X_{ik}\right) = 0 \qquad \forall\, x_i^T \qquad \textbf{(A2)}$$

where $x_i^T = \begin{bmatrix} 1 & X_{i1} & X_{i2} & \cdots & X_{ik} \end{bmatrix}$ is the (row) vector of regressor values for population observation i.

$$\Rightarrow \quad E\left(u_i\right) = 0 \quad \forall\, i \qquad \textbf{(A2.1)}$$

$$\Rightarrow \quad \mathrm{Cov}\left(X_{ij}, u_i\right) = E\left(X_{ij} u_i\right) = 0 \quad j = 1, 2, \ldots, k \qquad \textbf{(A2.2)}$$

$$\Rightarrow \quad E\left(Y_i \middle| x_i^T\right) = f(x_i, \beta) = x_i^T \beta = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} \qquad \textbf{(A2.3)}$$

A3.1: Constant conditional error *variances* (homoskedastic errors)

$$\text{Var}\left(u_i \middle| x_i^T\right) = E\left(u_i^2 \middle| x_i^T\right) = E\left(u_i^2 \middle| 1, X_{i1}, X_{i2}, \ldots, X_{ik}\right) = \sigma^2 > 0 \quad \forall\, x_i^T \qquad \textbf{(A3.1)}$$

where $\boldsymbol{\sigma^2}$ is a ***finite positive*** **(unknown)** ***constant***.

$$\Rightarrow \quad \text{Var}\left(Y_i \middle| x_i^T\right) = E\left(u_i^2 \middle| x_i^T\right) = E\left(u_i^2 \middle| 1, X_{i1}, X_{i2}, \ldots, X_{ik}\right) = \sigma^2 > 0 \quad \forall\, x_i^T$$

A3.2: Zero conditional error *covariances* (nonautoregressive errors)

$$\text{Cov}\left(u_i, u_s \middle| x_i^T, x_s^T\right) = E\left(u_i u_s \middle| x_i^T, x_s^T\right) = 0 \quad \forall\, x_i^T \neq x_s^T \qquad \textbf{(A3.2)}$$

$$\Rightarrow \quad \text{Cov}\left(Y_i, Y_s \middle| x_i^T, x_s^T\right) = E\left(u_i u_s \middle| x_i^T, x_s^T\right) = 0 \quad \forall\, x_i^T \neq x_s^T$$

A3.1 and A3.2 imply that the **error covariance matrix** for a sample of N observations takes the form

$$V\left(u \middle| X\right) = E\left(uu^T \middle| X\right) = \underset{(N\times N)}{\begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix}} = \sigma^2 \underset{(N\times N)}{\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}} = \sigma^2 I_N$$

where $\mathbf{I_N}$ is an **N×N identity matrix** with 1s along the principal diagonal and 0s in all the off-diagonal cells.

Assumptions A4 and A5: Specify the minimal required properties of the sample data

A4 The Assumption of Independent Random Sampling

The **sample data** consist of a ***random*** **sample of N observations** on the regressand Y and the regressors $x^T = (1\ X_1\ X_2\ \cdots\ X_k)$, which together comprise the observable variables in the population regression equation described by A1.

- Each of these N randomly selected observations can be written as:

$$(Y_i\ x_i^T) = (Y_i\ 1\ X_{i1}\ X_{i2}\ \cdots\ X_{ik}) \qquad i = 1, \ldots, N$$

- In matrix form, the full set of N sample observations on all the observable variables can be written as

$(y\ X)$ = the N×(K+1) observation matrix or data matrix, the i-th row of which is $(Y_i\ x_i^T) = (Y_i\ 1\ X_{i1}\ X_{i2}\ \cdots\ X_{ik})$

where

y = the N×1 regressand vector, the i-th row (element) of which is Y_i;

X = the N×K regressor matrix, the i-th row of which is $x_i^T = (1\ X_{i1}\ X_{i2}\ \cdots\ X_{ik})$.

- **Implications of the Random Sampling Assumption A4**

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

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

  $$\text{Random sampling} \Rightarrow \text{Cov}\left(u_i, u_s \,\middle|\, x_i^T, x_s^T\right) = \text{Cov}(u_i, u_s) = 0 \qquad i \neq s. \qquad \textbf{(16)}$$

  2. It also means that the dependent variables $\mathbf{Y_i}$ **and** $\mathbf{Y_s}$ **are *statistically independent***, so that $\mathbf{Y_i}$ **and** $\mathbf{Y_s}$ **have *zero covariance*** for any two distinct sample observations.

  $$\text{Random sampling} \Rightarrow \text{Cov}\left(Y_i, Y_s \,\middle|\, x_i^T, x_s^T\right) = \text{Cov}(Y_i, Y_s) = 0 \qquad i \neq s. \qquad \textbf{(17)}$$

- ***Result:*** The assumption of random sampling is sufficient for assumption A3.2 of zero covariance between observations, but is stronger than necessary.

- **When is the Random Sampling Assumption A4 Appropriate?**

  The random sampling assumption is usually appropriate for ***cross-sectional* regression models**, i.e., for regression models formulated for ***cross section data***.

  The random sampling assumption is hardly ever appropriate for ***time-series* regression models**, i.e., for regression models formulated for ***time series data***.

- **Characteristics of Cross Section and Time Series Data**

  Cross Section Data

  - ***Definition:*** A **cross-sectional data set** consists of a sample of observations on individual economic agents or other units taken at a **single *point* in time** or over a **single *period* of time**.

  - A ***distinguishing characteristic*** of any cross-sectional data set is that the individual observations have **no natural ordering**.

  - A ***common, almost universal characteristic*** of cross-sectional data sets is that they usually are constructed by **random sampling** from underlying populations.

  Time Series Data.

  - ***Definition:*** A **time-series data set** consists of a sample of observations on one or more variables **over several successive periods or intervals of time**.

  - A ***distinguishing characteristic*** of any time-series data set is that the observations have a natural ordering -- specifically a **chronological ordering**.

  - A ***common, almost universal characteristic*** of time-series data sets is that the **sample observations exhibit a high degree of *time dependence***, and therefore cannot be assumed to be generated by random sampling.

A5 The Full Rank Assumption

Also called the **Assumption of No Perfect Multicollinearity.**

- **Formal Statement of A5:** The **regressor matrix X has *full column rank***; i.e.,

$$\textbf{rank(X)} \equiv \rho(\mathbf{X}) = \mathbf{K} = \mathbf{k+1} = \textbf{the number of columns in X}. \qquad \textbf{(A5)}$$

- **Interpretation of A5**

- Assumption A5 is **an assumption about the *population* and *sample data***, as distinct from an assumption about the regression model.
- Assumption A5 means that there is **no exact *linear* relationship** among the population values, and therefore the sample values, of the K = k+1 regressors.
- Assumption A5 also means that **none of the k regressors $\mathbf{X_1}$ $\mathbf{X_2}$ $\cdots$ $\mathbf{X_k}$ is a *constant***, either in the population or in the sample.

- **Why is Assumption A5 Important?**

Answer: Assumption A5 must be satisfied by the sample data if an estimator of the complete regression coefficient vector β is to be computable.

- **Scalar formulation of A5:** There exists no set of K = k+1 constants $\lambda_0, \lambda_1, \lambda_2, ..., \lambda_k$, not all zero, such that

$$\lambda_0 + \lambda_1 X_{i1} + \lambda_2 X_{i2} + \cdots + \lambda_k X_{ik} = 0 \qquad \forall \ i \qquad \textbf{(A5.1)}$$

or

$$\lambda_0 + \sum_{j=1}^{k} \lambda_j X_{ij} = 0 \qquad \text{(where } X_{i0} = 1\text{)} \qquad \forall \ i \qquad \textbf{(A5.1)}$$

- **Vector formulation of A5:** The assumption that X has full column rank equal to K = k + 1 means that the **columns of X are *not* linearly dependent**, which in turn requires that **there exists *no nonzero* vector of constants λ**, where

$$\lambda^T = \begin{bmatrix} \lambda_0 & \lambda_1 & \lambda_2 & \cdots & \lambda_k \end{bmatrix},$$

such that the vector product

$$\lambda^T x_i = 0 \qquad \forall \ i \qquad \textbf{(A5.2)}$$

where x_i is the transpose of the i-th row of the regressor matrix X and is therefore of dimension K×1.

Vector formulation (A5.2) of the full rank assumption A5 is equivalent to the scalar formulation (A5.1):

$$\lambda^T x_i = \begin{bmatrix} \lambda_0 & \lambda_1 & \lambda_2 & \cdots & \lambda_k \end{bmatrix} \begin{bmatrix} 1 \\ X_{i1} \\ X_{i2} \\ \vdots \\ X_{ik} \end{bmatrix} = \lambda_0 + \sum_{j=1}^{k} \lambda_j X_{ij} = 0 \qquad \forall \ i \qquad \textbf{(18)}$$

Note: The row vector of constants λ^T is 1×K and the column vector x_i is K×1, so that the vector product $\lambda^T x_i$ is a 1×1 scalar, namely the linear function on the left-hand side of (A5.1).

Perfect Multicollinearity: Example 1

- **Consider** the simple linear regression model given by population regression equation (PRE) 1:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + u_i \qquad (1)$$

- Suppose the regressors X_{i2} and X_{i3} satisfy the following exact linear relationship:

$$X_{i3} = 2X_{i2} \quad \text{for all } i = 1, \ldots, N \qquad (2)$$

- Re-write this exact linear relationship between X_{i2} and X_{i3} in the general form of any exact linear dependence among the regressors in PRE (1):

$$\lambda_0 + \lambda_1 X_{i1} + \lambda_2 X_{i2} + \lambda_3 X_{i3} = 0 \qquad \forall \ i \qquad (5.1)$$

$X_{i3} = 2X_{i2}$ can be re-written in implicit form as

$$-2X_{i2} + X_{i3} = 0 \quad \text{for all } i = 1, \ldots, N \qquad (2^*)$$

Setting the constants $\lambda_0 = 0$, $\lambda_1 = 0$, $\lambda_2 = -2$, and $\lambda_3 = 1$ in (5.1) yields the exact linear dependence (2*):

$$\lambda_0 + \lambda_1 X_{i1} + \lambda_2 X_{i2} + \lambda_3 X_{i3} = 0$$
$$0 + 0 \cdot X_{i1} - 2 \cdot X_{i2} + 1 \cdot X_{i3} = 0$$
$$-2X_{i2} + X_{i3} = 0 \qquad \text{for all } i = 1, \ldots, N$$

Consequences of Perfect Multicollinearity: Example 1

1. Substitute for X_{i3} in regression equation (1) the expression $X_{i3} = 2X_{i2}$:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + u_i && (1) \\ &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 2X_{i2} + u_i \\ &= \beta_0 + \beta_1 X_{i1} + (\beta_2 + 2\beta_3) X_{i2} + u_i \\ &= \beta_0 + \beta_1 X_{i1} + \alpha_2 X_{i2} + u_i \qquad \text{where by definition } \alpha_2 = \beta_2 + 2\beta_3 && (3) \end{aligned}$$

- Regression equation (3) can be estimated from sample data – i.e., we ***can*** **compute estimates of the regression coefficients β_0, β_1 and α_2**. But we ***cannot*** **compute** from sample data **estimates of the regression coefficients β_2 and β_3**; in other words, the perfect multicollinearity implied by the exact linear relationship $X_{i3} = 2X_{i2}$ for all i means that **the coefficients β_2 and β_3 are *not* estimable or *not* computable** from sample data.

2. Alternatively, substitute for X_{i2} in regression equation (1) the expression $X_{i2} = 0.5X_{i3}$:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + u_i && (1) \\ &= \beta_0 + \beta_1 X_{i1} + \beta_2 0.5 X_{i3} + \beta_3 X_{i3} + u_i \\ &= \beta_0 + \beta_1 X_{i1} + (0.5\beta_2 + \beta_3) X_{i3} + u_i \\ &= \beta_0 + \beta_1 X_{i1} + \alpha_3 X_{i3} + u_i \qquad \text{where by definition } \alpha_3 = 0.5\beta_2 + \beta_3 && (3^*) \end{aligned}$$

- From regression equation (3*), we see again that we ***cannot*** **compute** from sample data **separate estimates of the regression coefficients β_2 and β_3**; **the coefficients β_2 and β_3 are *not* estimable or *not* computable** from sample data.

Perfect Multicollinearity: Example 2

- Consider again the simple linear regression model given by population regression equation (1):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + u_i \qquad (1)$$

- Now suppose the regressors X_{i2} and X_{i3} satisfy the following exact linear relationship:

$$X_{i2} + X_{i3} = 1 \qquad \text{for all } i = 1, \ldots, N \qquad (4)$$

- Re-write this exact linear relationship between X_{i2} and X_{i3} in the general form of any exact linear dependence among the regressors in PRE (1):

$$\lambda_0 + \lambda_1 X_{i1} + \lambda_2 X_{i2} + \lambda_3 X_{i3} = 0 \qquad \forall \ i \qquad (5.1)$$

$X_{i2} + X_{i3} = 1$ can be re-written in implicit form as

$$-1 + X_{i2} + X_{i3} = 0 \text{ for all } i = 1, \ldots, N \qquad (4^*)$$

Setting the constants $\lambda_0 = -1$, $\lambda_1 = 0$, $\lambda_2 = 1$, and $\lambda_3 = 1$ in (5.1) yields the exact linear dependence (4*):

$$\begin{aligned} &\lambda_0 + \lambda_1 X_{i1} + \lambda_2 X_{i2} + \lambda_3 X_{i3} = 0 \\ &-1 + 0 \cdot X_{i1} + 1 \cdot X_{i2} + 1 \cdot X_{i3} = 0 \\ &\qquad -1 + X_{i2} + X_{i3} = 0 \qquad \text{for all } i = 1, \ldots, N \end{aligned}$$

Consequences of Perfect Multicollinearity: Example 2

- Substitute for X_{i3} in regression equation (1) the expression $X_{i3} = 1 - X_{i2}$:

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + u_i && (1)\\
&= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 (1 - X_{i2}) + u_i \\
&= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 - \beta_3 X_{i2} + u_i \\
&= (\beta_0 + \beta_3) + \beta_1 X_{i1} + (\beta_2 - \beta_3) X_{i2} + u_i \\
&= \delta_0 + \beta_1 X_{i1} + \delta_2 X_{i2} + u_i \qquad \text{where by definition } \delta_0 = \beta_0 + \beta_3 \text{ and } \delta_2 = \beta_2 - \beta_3 && (5)
\end{aligned}
$$

Result: Regression equation (5) can be estimated from sample data – i.e., we ***can*** **compute estimates of the regression coefficients δ_0, β_1 and δ_2.** But we ***cannot*** **compute** from sample data separate **estimates of the regression coefficients β_0, β_2 and β_3**; in other words, the perfect multicollinearity implied by the exact linear relationship $X_{i3} = 1 - X_{i2}$ for all i means that **the coefficients β_0, β_2 and β_3 are *not* estimable or *not* computable** from sample data.

Consequences of Perfect Multicollinearity: Example 2 (continued)

- Alternatively, substitute for X_{i2} in regression equation (1) the expression $X_{i2} = 1 - X_{i3}$:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + u_i \qquad (1) \\ &= \beta_0 + \beta_1 X_{i1} + \beta_2 (1 - X_{i3}) + \beta_3 X_{i3} + u_i \\ &= \beta_0 + \beta_1 X_{i1} + \beta_2 - \beta_2 X_{i3} + \beta_3 X_{i3} + u_i \\ &= (\beta_0 + \beta_2) + \beta_1 X_{i1} + (\beta_3 - \beta_2) X_{i3} + u_i \\ &= \gamma_0 + \beta_1 X_{i1} + \gamma_3 X_{i3} + u_i \qquad \text{where by definition } \gamma_0 = \beta_0 + \beta_2 \text{ and } \gamma_3 = \beta_3 - \beta_2 \qquad (5^*) \end{aligned}$$

Result: Regression equation (5*) can be estimated from sample data – i.e., we ***can*** **compute estimates of the regression coefficients γ_0, β_1 and γ_3.** But again we ***cannot*** **compute** from sample data separate **estimates of the regression coefficients β_0, β_2 and β_3**; the perfect multicollinearity implied by the exact linear relationship $X_{i2} = 1 - X_{i3}$ for all i means that **the coefficients β_0, β_2 and β_3 are *not* estimable or *not* computable** from sample data.

Importance and Purposes of Assumptions A1-A5

1. Assumptions A1-A5 specify the set of conditions under which the **OLS (Ordinary Least Squares) estimator** $\hat{\beta}_{OLS}$ of the coefficient vector β and associated statistical inference procedures have several **optimal statistical properties**.

These statistical properties are summarized by the **Gauss-Markov Theorem**, which states that

$\hat{\beta}_{OLS}$ is the **Best Linear Unbiased Estimator (BLUE)** of the coefficient vector β

B = Best = *minimum variance* estimator of β:

$$\text{Var}(\hat{\beta}_j) \leq \text{Var}(\tilde{\beta}_j) \quad j = 0, 1, 2, \ldots, k$$

where

$\hat{\beta}_j$ = the OLS estimator of regression coefficient β_j

$\tilde{\beta}_j$ = any other linear unbiased estimator of β_j

L = Linear = *linear* in the regressand Y_i,
= *linear* in the regressand vector y:

Matrix formula for $\hat{\beta}_{OLS}$ is: $\hat{\beta}_{OLS} = (X^T X)^{-1} X^T y$

U = Unbiased:

In scalar terms: $E(\hat{\beta}_j) = \beta_j$ for all j = 0, 1, 2, …, k

In matrix terms: $E(\hat{\beta}_{OLS}) = \beta$

2. Assumptions A1-A5 highlight the things that can go wrong in applied linear regression analysis.