ECON 452* -- NOTE 9

OLS Estimation of the Classical Linear Regression Model: Matrix Notation and Derivations

1. Population Regression Equation (PRE)

The PRE is for a sample of N observations is

$$y = X\beta + u = E(y\,|\,X) + u \tag{1}$$

where

y = the N×1 ***regressand vector***

X = the N×K ***regressor matrix***

β = the K×1 ***regression coefficient vector***

$E(y\,|\,X) = X\beta$ = the N×1 ***population regression function*** (PRF)

$u = y - E(y\,|\,X) = y - X\beta$ = the N×1 ***error vector***

- ***Details***

$$y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_N \end{bmatrix}$$ = the N×1 ***regressand*** **vector**

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

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

= the N×K matrix of observed sample values of the K = k + 1 regressors X_{i0}, X_{i1}, X_{i2}, ..., X_{ik}, where $X_{i0} = 1 \; \forall \; i = 1, ..., N$ and the remaining k = K – 1 regressors are variables.

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$ = the K×1 ***coefficient* vector**

= the K×1 or (k+1)×1 <u>column</u> vector of regression coefficients β_j, j = 0, 1, ..., k.

$$u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_N \end{bmatrix}$$ = the N×1 ***error* vector**

= the N×1 <u>column</u> vector of random errors u_i (i = 1, ..., N) corresponding to each of the N sample observations.

$$E(y\,|\,X) = \begin{bmatrix} E(Y_1\,|\,x_1^T) \\ E(Y_2\,|\,x_2^T) \\ E(Y_3\,|\,x_3^T) \\ \vdots \\ E(Y_N\,|\,x_N^T) \end{bmatrix} = \begin{bmatrix} x_1^T\beta \\ x_2^T\beta \\ x_3^T\beta \\ \vdots \\ x_N^T\beta \end{bmatrix}$$ = the N×1 ***regression function* vector**

= the N×1 <u>column</u> vector of values of the PRF for each of the N sample observations, $E(Y_i\,|\,x_i^T) = x_i^T\beta = \beta_0 + \sum_j \beta_j X_{ij}$, i = 1, ..., N.

The Error Covariance Matrix Under Assumption A3

Assumption A3 is the assumption of ***homoskedastic*** **and** ***nonautoregressive*** **errors**:

- **Assumption A3.1 of Constant Error Variances (Homoskedastic Errors)**

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

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

- **Assumption A3.2 of Zero Error Covariances (Nonautoregressive Errors)**

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

Under Assumption A3 the ***error covariance matrix*** takes the form

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

where

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

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

1. By A3.1, all ***diagonal elements*** of $V(u|X)$ equal the positive constant σ^2, since A3.1 says that $Var(u_i|x_i^T) = E(u_i^2|x_i^T) = \sigma^2 > 0 \quad \forall\ i$.

2. By A3.2, all ***off diagonal elements*** of $V(u|X)$ equal zero, since A3.2 says that $Cov(u_i, u_s|x_i^T, x_s^T) = E(u_i u_s|x_i^T, x_s^T) = 0 \quad \forall\ i \neq s$.

2. OLS Estimator of the Coefficient Vector β

$$\hat{\beta}_{OLS} = \hat{\beta} = \left(X^T X\right)^{-1} X^T y \qquad (2)$$

where

$$\hat{\beta}_{OLS} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \text{ the K×1 OLS coefficient vector estimator}$$

= the K×1 or (k+1)×1 <u>column</u> vector of OLS coefficient estimators $\hat{\beta}_j$ of the unknown partial regression coefficients β_j, j = 0, 1, ..., k.

$X^T X$ = the K×K regressor cross-product matrix, a symmetric matrix with

diagonal elements equal to $\sum_{i=1}^{N} X_{ij}^2$ (j = 0, 1, …, k) and

off-diagonal elements equal to $\sum_{i=1}^{N} X_{ij} X_{ih}$ ($j \neq h$, h = 0, 1, …, k)

$\left(X^T X\right)^{-1}$ = the ***inverse* of the *regressor cross-product matrix*** (a K×K matrix)

$X^T y$ = the K×1 cross-product vector with elements equal to

$\sum_{i=1}^{N} X_{ij} Y_i$ (j = 0, 1, …, k)

The **OLS coefficient estimator $\hat{\beta}_{OLS}$** is derived by ***minimizing* the residual sum of squares function**, denoted as RSS($\hat{\beta}$), for given sample values of the observable variables (Y_i, 1, X_{i1}, X_{i2}, …, X_{ik}) i = 1, …, N. The RSS($\hat{\beta}$) function can be written in both scalar and matrix terms.

❑ Scalar Expression for RSS($\hat{\beta}$)

Since the OLS residuals {$\hat{u}_i$: i = 1, …, N} are

$$\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik} \qquad (i = 1, \ldots, N) \qquad (3)$$

the RSS($\hat{\beta}$) function can be written in scalar terms as

$$\text{RSS}(\hat{\beta}) = \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik} \right)^2 \qquad (4.1)$$

❑ Matrix Expression for RSS($\hat{\beta}$)

- The inner product of the N×1 residual vector $\hat{u}$ with itself is the RSS($\hat{\beta}$):

$$\hat{u}^T \hat{u} = \begin{bmatrix} \hat{u}_1 & \hat{u}_2 & \hat{u}_3 & \cdots & \hat{u}_N \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \vdots \\ \hat{u}_N \end{bmatrix} = \hat{u}_1^2 + \hat{u}_2^2 + \hat{u}_3^2 + \cdots + \hat{u}_N^2 = \sum_{i=1}^{N} \hat{u}_i^2$$

Result: $\text{RSS}(\hat{\beta}) = \hat{u}^T \hat{u}$ (4.2)

- Rewrite $\text{RSS}(\hat{\beta}) = \hat{u}^T \hat{u}$ in terms of the observable variables y and X and the coefficient estimator $\hat{\beta}$.

1. Since the residual vector $\hat{u} = y - X\hat{\beta}$, the transpose of $\hat{u}$ is

$$\hat{u}^T = \left(y - X\hat{\beta} \right)^T = \left(y^T - \hat{\beta}^T X^T \right).$$

2. We can therefore write $\text{RSS}(\hat{\beta})$ in terms of y, X and $\hat{\beta}$ as:

$$\begin{aligned}\text{RSS}(\hat{\beta}) \;=\; \hat{u}^T\hat{u} &= \left(y - X\hat{\beta}\right)^T\left(y - X\hat{\beta}\right) \\ &= \left(y^T - \hat{\beta}^T X^T\right)\left(y - X\hat{\beta}\right) \\ &= y^T y - \hat{\beta}^T X^T y - y^T X\hat{\beta} + \hat{\beta}^T X^T X\hat{\beta}. \qquad (5)\end{aligned}$$

3. Since all four terms on the right-hand side of (5) are scalars (1×1 quantities), the second term $\hat{\beta}^T X^T y$ is equal to its transpose $y^T X\hat{\beta}$, which is the third term. We can therefore sum these two terms as follows:

$$-\hat{\beta}^T X^T y - y^T X\hat{\beta} = -\hat{\beta}^T X^T y - \hat{\beta}^T X^T y = -2\hat{\beta}^T X^T y.$$

- ***Result:*** The residual sum of squares $\text{RSS}(\hat{\beta})$ can therefore be written as:

$$\text{RSS}(\hat{\beta}) \;=\; \hat{u}^T\hat{u} \;=\; y^T y - 2\hat{\beta}^T X^T y + \hat{\beta}^T X^T X\hat{\beta}. \qquad (6)$$

❑ The OLS Estimation Criterion

The OLS coefficient estimator is that expression for $\hat{\beta}$ which ***minimizes*** $\text{RSS}(\hat{\beta})$ for given values of the regressand vector y and the regressor matrix X.

The OLS estimation criterion can therefore be stated as:

$$\begin{aligned}\underset{\{\hat{\beta}\}}{\text{minimize}}\; \text{RSS}(\hat{\beta}) \;&=\; y^T y - 2\hat{\beta}^T X^T y + \hat{\beta}^T X^T X\hat{\beta}. \qquad (7.1) \\ &= \sum_{i=1}^{N}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \cdots - \hat{\beta}_k X_{ik}\right)^2 \qquad (7.2)\end{aligned}$$

3. The OLS Sample Regression Equation (the OLS SRE)

$$y = X\hat{\beta} + \hat{u} = \hat{y} + \hat{u}, \qquad \hat{y} = X\hat{\beta}, \qquad \hat{u} = y - \hat{y} = y - X\hat{\beta} \tag{8}$$

where

$$y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_N \end{bmatrix} = \text{ the N×1 regressand vector}$$

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

$$\hat{y} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \hat{Y}_3 \\ \vdots \\ \hat{Y}_N \end{bmatrix} = \begin{bmatrix} x_1^T\hat{\beta} \\ x_2^T\hat{\beta} \\ x_3^T\hat{\beta} \\ \vdots \\ x_N^T\hat{\beta} \end{bmatrix} = \text{ the N×1 vector of OLS estimated values of } Y_i$$

= the N×1 <u>column</u> vector of OLS estimated (predicted) values $\hat{Y}_i = x_i^T\hat{\beta} = \hat{\beta}_0 + \sum_j \hat{\beta}_j X_{ij}$ (i = 1, ..., N) corresponding to each of the N sample observations.

$$\hat{u} = \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \vdots \\ \hat{u}_N \end{bmatrix} = \text{ the N×1 OLS residual vector}$$

= the N×1 <u>column</u> vector of OLS residuals $\hat{u}_i$ (i = 1, ..., N) corresponding to each of the N sample observations.

Remarks:

1. The OLS sample regression function $\hat{y} = X\hat{\beta}$ is an estimator of the population regression function (PRF) $E(y|X) = X\beta$.

2. The OLS residual vector $\hat{u} = y - \hat{y} = y - X\hat{\beta}$ is, in effect, an estimator of the unobservable random error vector $u = y - E(y|X) = y - X\beta$.

4. Sampling Distribution of the OLS Coefficient Estimator

The **finite sampling distribution** of the OLS coefficient estimator $\hat{\beta}_{OLS}$ has two important sets of moments that are required for performing statistical inference (hypothesis testing and confidence interval estimation):

1. the **mean, or expectation**, of $\hat{\beta}_{OLS}$, denoted as $E(\hat{\beta}_{OLS})$;

2. the **variance-covariance matrix** of $\hat{\beta}_{OLS}$, denoted as $V(\hat{\beta}_{OLS}) = V_{OLS}$.

5. Mean of the OLS Coefficient Estimator

In this section, we derive the mean, or expectation, of $\hat{\beta}_{OLS}$; in the next section we derive the variance-covariance matrix of $\hat{\beta}_{OLS}$ and an unbiased estimator of it.

- The OLS coefficient estimator $\hat{\beta}_{OLS}$ is given by the formula

$$\hat{\beta}_{OLS} = \left(X^T X\right)^{-1} X^T y. \qquad (2)$$

❑ **Proposition:** $E(\hat{\beta}_{OLS}) = \beta$, meaning that $\hat{\beta}_{OLS}$ **is an *unbiased* estimator of β.**

Proof:

1. Substitute for the regressand vector y in the formula for $\hat{\beta}_{OLS}$ the expression for y:

$$\hat{\beta}_{OLS} = \left(X^T X\right)^{-1} X^T y$$
$$\hat{\beta}_{OLS} = \left(X^T X\right)^{-1} X^T (X\beta + u) \qquad \text{since } y = X\beta + u \text{ by A1}$$
$$\hat{\beta}_{OLS} = \left(X^T X\right)^{-1} X^T X\beta + \left(X^T X\right)^{-1} X^T u$$
$$\hat{\beta}_{OLS} = \beta + \left(X^T X\right)^{-1} X^T u \qquad \text{since } \left(X^T X\right)^{-1} X^T X\beta = I_K\beta = \beta$$

Result:

$$\hat{\beta}_{OLS} = \beta + \left(X^T X\right)^{-1} X^T u \qquad (9)$$

Note: The difference between $\hat{\beta}_{OLS}$ and β is called the sampling error of $\hat{\beta}_{OLS}$. The above equation implies that

$$\text{sampling error of } \hat{\beta}_{OLS} = \hat{\beta}_{OLS} - \beta = \left(X^T X\right)^{-1} X^T u. \qquad (10)$$

2. Use equation (9) to take the expectation of $\hat{\beta}_{OLS}$ conditional on given values of the regressors, i.e., conditional on the regressor matrix X.

$$\hat{\beta}_{OLS} = \beta + \left(X^T X\right)^{-1} X^T u \qquad (9)$$

Now take the conditional expectation of $\hat{\beta}_{OLS}$. Since we are conditioning on the regressor matrix X, we can move the expectation operator past all terms in X. We also use assumption A2, the assumption of zero conditional mean errors, which states that $E\left(u \mid X\right) = \underline{0}$.

$$\begin{aligned} E\left(\hat{\beta}_{OLS} \mid X\right) &= E\left(\beta \mid X\right) + E\left[\left(X^T X\right)^{-1} X^T u \mid X\right] \\ &= \beta + \left(X^T X\right)^{-1} X^T E\left(u \mid X\right) && \text{since } \beta = \text{a constant vector} \\ &= \beta + \left(X^T X\right)^{-1} X^T \underline{0} && \text{since } E\left(u \mid X\right) = \underline{0} \text{ by A2} \\ &= \beta \end{aligned}$$

- ***Result:***

$$E\left(\hat{\beta}_{OLS} \mid X\right) = \beta \quad \Rightarrow \quad \hat{\beta}_{OLS} \textbf{ is an } \textbf{\textit{unbiased}} \textbf{ estimator of } \beta. \qquad (11)$$

Note: The unbiasedness property of the OLS coefficient estimator $\hat{\beta}_{OLS}$ depends only on Assumptions A1 and A2 of the classical linear regression model. Assumptions A1 and A2 are the only two assumptions used in the proof of the unbiasedness property.

6. Variance-Covariance Matrix for the OLS Coefficient Estimator

- **Definition:** The variance-covariance matrix of the OLS coefficient estimator $\hat{\beta}_{OLS}$ -- or more briefly the covariance matrix of $\hat{\beta}_{OLS}$ -- conditional on the regressor matrix X is defined as

$$V\left(\hat{\beta}_{OLS}\middle| X\right) = E\left\{\left[\hat{\beta}_{OLS} - E(\hat{\beta}_{OLS}| X)\right]\left[\hat{\beta}_{OLS} - E(\hat{\beta}_{OLS}| X)\right]^T \middle| X\right\} \qquad (12)$$

- **Simplification:** The unbiasedness of $\hat{\beta}_{OLS}$ means that $E\left(\hat{\beta}_{OLS}\middle| X\right) = \beta$. Replacing $E\left(\hat{\beta}_{OLS}\middle| X\right)$ with β in definition (12) permits $V\left(\hat{\beta}_{OLS}\middle| X\right)$ to be written as

$$V\left(\hat{\beta}_{OLS}\middle| X\right) = E\left\{\left[\hat{\beta}_{OLS} - \beta\right]\left[\hat{\beta}_{OLS} - \beta\right]^T \middle| X\right\} \qquad (13)$$

Remarks:

1. Recall that the difference between $\hat{\beta}_{OLS}$ and β is called the sampling error of $\hat{\beta}_{OLS}$:

$$\textbf{sampling error of } \hat{\boldsymbol{\beta}}_{\mathbf{OLS}} = \hat{\beta}_{OLS} - \beta = \left(X^T X\right)^{-1} X^T u. \qquad (10)$$

2. The covariance matrix $V\left(\hat{\beta}_{OLS}\middle| X\right)$ in (13) is thus the expectation, conditional on the regressor matrix X, of the outer product of the K×1 vector of sampling errors $\left(\hat{\beta}_{OLS} - \beta\right)$.

❑ Derivation of the Conditional Covariance Matrix for $\hat{\beta}_{OLS}$, $V(\hat{\beta}_{OLS}|X)$

- From equation (10), the K×1 vector of sampling errors of $\hat{\beta}_{OLS}$ is

$$\hat{\beta}_{OLS} - \beta = (X^TX)^{-1}X^Tu \qquad (10)$$

- The transpose of the K×1 vector $(\hat{\beta}_{OLS} - \beta)$ is therefore the 1×K vector

$$(\hat{\beta}_{OLS} - \beta)^T = \left[(X^TX)^{-1}X^Tu\right]^T = u^TX(X^TX)^{-1} \qquad (14)$$

- The outer product of the K×1 vector $(\hat{\beta}_{OLS} - \beta)$ is therefore the K×K matrix

$$(\hat{\beta}_{OLS} - \beta)(\hat{\beta}_{OLS} - \beta)^T = (X^TX)^{-1}X^Tuu^TX(X^TX)^{-1}. \qquad (15)$$

- Now substitute the right-hand side of equation (15) into equation (13) for $V(\hat{\beta}_{OLS}|X)$:

$$V(\hat{\beta}_{OLS}|X) = E\left\{\left[\hat{\beta}_{OLS} - \beta\right]\left[\hat{\beta}_{OLS} - \beta\right]^T | X\right\} \qquad (13)$$

$$= E\left\{(X^TX)^{-1}X^Tuu^TX(X^TX)^{-1} | X\right\} \qquad \text{from equation (15)}$$

$$= (X^TX)^{-1}X^TE(uu^T | X)X(X^TX)^{-1}$$

$$= (X^TX)^{-1}X^T(\sigma^2 I_N)X(X^TX)^{-1} \qquad \text{since } E(uu^T | X) = \sigma^2 I_N \text{ by A3}$$

$$= \sigma^2(X^TX)^{-1}X^TI_NX(X^TX)^{-1} \qquad \text{since } \sigma^2 \text{ is a scalar constant}$$

$$= \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1} \qquad \text{since } X^TI_NX = X^TX$$

$$= \sigma^2(X^TX)^{-1}I_K \qquad \text{since } X^TX(X^TX)^{-1} = I_K$$

$$= \sigma^2(X^TX)^{-1} \qquad \text{since } (X^TX)^{-1}I_K = (X^TX)^{-1}$$

- ***Result:*** The conditional covariance matrix for the OLS coefficient estimator $\hat{\beta}_{OLS}$ is

$$V\left(\hat{\beta}_{OLS} \middle| X\right) = \sigma^2\left(X^T X\right)^{-1}. \tag{16}$$

Remarks:

1. The constant error variance σ^2 is unknown, meaning we do not know its value.
2. Consequently, it is necessary to obtain an estimator of σ^2 in order to estimate the covariance matrix $V\left(\hat{\beta}_{OLS} \middle| X\right) = \sigma^2\left(X^T X\right)^{-1}$.

❑ An Unbiased Estimator of the Error Variance σ^2

- It is possible to show that an ***unbiased*** **estimator of the constant error variance** $\boldsymbol{\sigma^2}$ is obtained by dividing the OLS residual sum of squares RSS by its degrees of freedom N–K:

$$\hat{\sigma}^2_{OLS} = \frac{RSS(\hat{\beta}_{OLS})}{N-K} = \frac{\hat{u}^T\hat{u}}{N-K} \tag{17}$$

where

$$RSS(\hat{\beta}_{OLS}) = \hat{u}^T\hat{u} = \left(y - X\hat{\beta}_{OLS}\right)^T\left(y - X\hat{\beta}_{OLS}\right)$$
$$= \text{ the residual sum of squares for the coefficient estimator } \hat{\beta}_{OLS}$$

- The unbiasedness of $\hat{\sigma}^2_{OLS}$ follows from the fact that

$$E\left(RSS(\hat{\beta}_{OLS})\right) = E\left(\hat{u}^T\hat{u}\right) = (N-K)\sigma^2$$

which in turn implies that

$$E\left(\hat{\sigma}^2_{OLS}\right) = \frac{E\left(\hat{u}^T\hat{u}\right)}{N-K} = \frac{(N-K)\sigma^2}{N-K} = \sigma^2.$$

- The **square root of $\hat{\sigma}^2_{OLS}$** is the standard deviation of the observed Y_i sample values about the OLS sample regression function $X\hat{\beta}_{OLS}$.

$$\hat{\sigma}_{OLS} = \sqrt{\hat{\sigma}^2_{OLS}} = \sqrt{\frac{RSS(\hat{\beta}_{OLS})}{N-K}} = \sqrt{\frac{\hat{u}^T\hat{u}}{N-K}}. \tag{18}$$

Unfortunately, $\hat{\sigma}_{OLS}$ has a bewildering array of names.

$\hat{\sigma}_{OLS}$ is called (1) the standard error of estimate
(2) the standard error of the regression (SER)
(3) the root mean square error (RMSE)

❑ An Unbiased Estimator of the Covariance Matrix for $\hat{\beta}_{OLS}$

- Start with the formula for the conditional covariance matrix of $\hat{\beta}_{OLS}$ given above in equation (16):

$$V\left(\hat{\beta}_{OLS} \mid X\right) = \sigma^2\left(X^TX\right)^{-1}. \tag{16}$$

- An ***unbiased* estimator of $V\left(\hat{\beta}_{OLS} \mid X\right)$** is obtained by simply substituting for the unknown error variance σ^2 in (16) the unbiased estimator $\hat{\sigma}^2_{OLS}$:

$$\hat{V}_{OLS} = \hat{\sigma}^2_{OLS}\left(X^TX\right)^{-1}. \tag{19}$$

- **Interpreting the Elements of $\hat{\mathbf{V}}_{\mathbf{OLS}}$**

- $\hat{\mathbf{V}}_{\mathbf{OLS}}$ is a **square, symmetric K×K matrix**, the elements of which are the estimated variances and covariances of the OLS coefficient estimates $\hat{\beta}_j$ j = 0, 1, …, k. The symmetry of $\hat{V}_{OLS}$ follows from the fact that

$$\hat{Cov}(\hat{\beta}_f, \hat{\beta}_g) = \hat{Cov}(\hat{\beta}_g, \hat{\beta}_f) \quad \text{for all } f \neq g.$$

- The $\hat{V}_{OLS}$ matrix in general is written as:

$$\hat{V}_{OLS} = \begin{bmatrix} \hat{Var}(\hat{\beta}_0) & \hat{Cov}(\hat{\beta}_0,\hat{\beta}_1) & \hat{Cov}(\hat{\beta}_0,\hat{\beta}_2) & \cdots & \hat{Cov}(\hat{\beta}_0,\hat{\beta}_k) \\ \hat{Cov}(\hat{\beta}_1,\hat{\beta}_0) & \hat{Var}(\hat{\beta}_1) & \hat{Cov}(\hat{\beta}_1,\hat{\beta}_2) & \cdots & \hat{Cov}(\hat{\beta}_1,\hat{\beta}_k) \\ \hat{Cov}(\hat{\beta}_2,\hat{\beta}_0) & \hat{Cov}(\hat{\beta}_2,\hat{\beta}_1) & \hat{Var}(\hat{\beta}_2) & \cdots & \hat{Cov}(\hat{\beta}_2,\hat{\beta}_k) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{Cov}(\hat{\beta}_k,\hat{\beta}_0) & \hat{Cov}(\hat{\beta}_k,\hat{\beta}_1) & \hat{Cov}(\hat{\beta}_k,\hat{\beta}_2) & \cdots & \hat{Var}(\hat{\beta}_k) \end{bmatrix}. \qquad (20)$$

- The software program *Stata* stores and displays the elements of $\hat{V}_{OLS}$ in a slightly different arrangement than that given in (20) above. *Stata* places the estimated variances and covariances involving the intercept coefficient estimate $\hat{\beta}_0$ in the last row and last column of the $\hat{V}_{OLS}$ matrix, rather than in the first row and column as in (20).

$$\textit{Stata}\ \hat{V}_{OLS} = \begin{bmatrix} \hat{Var}(\hat{\beta}_1) & \hat{Cov}(\hat{\beta}_1,\hat{\beta}_2) & \cdots & \hat{Cov}(\hat{\beta}_1,\hat{\beta}_k) & \hat{Cov}(\hat{\beta}_1,\hat{\beta}_0) \\ \hat{Cov}(\hat{\beta}_2,\hat{\beta}_1) & \hat{Var}(\hat{\beta}_2) & \cdots & \hat{Cov}(\hat{\beta}_2,\hat{\beta}_k) & \hat{Cov}(\hat{\beta}_2,\hat{\beta}_0) \\ \hat{Cov}(\hat{\beta}_3,\hat{\beta}_1) & \hat{Cov}(\hat{\beta}_3,\hat{\beta}_2) & \cdots & \hat{Cov}(\hat{\beta}_3,\hat{\beta}_k) & \hat{Cov}(\hat{\beta}_3,\hat{\beta}_0) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat{Cov}(\hat{\beta}_k,\hat{\beta}_1) & \hat{Cov}(\hat{\beta}_k,\hat{\beta}_2) & \cdots & \hat{Var}(\hat{\beta}_k) & \hat{Cov}(\hat{\beta}_k,\hat{\beta}_0) \\ \hat{Cov}(\hat{\beta}_0,\hat{\beta}_1) & \hat{Cov}(\hat{\beta}_0,\hat{\beta}_2) & \cdots & \hat{Cov}(\hat{\beta}_0,\hat{\beta}_k) & \hat{Var}(\hat{\beta}_0) \end{bmatrix}.$$