2HSOE052

Chapter 4_ MULTIPLE REGRESSION ANALYSIS: THE PROBLEM OF ESTIMATION

Population regression function (PRF) : A population regression function is a linear function, which hypothesizes a theoretical relationship between a dependent variable and a set of independent or explanatory variables at a population level.

A. Three Variable Regression Analysis : THE PROBLEM OF ESTIMATION

Let us write three-variable PRF as

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

Homoskedastic (also spelled "homoscedastic") refers to a condition in which the variance of the residual, or error term, in a regression model is constant. That is, the error term does not vary much as the value of the predictor variable changes.

To operate within the framework of the classical linear regression model (CLRM) , we assume the following:

Zero mean value of u_i, or

$$E(u_i \mid X_{2i}, X_{3i}) = 0 \qquad \text{for each } i$$

No serial correlation, or

$$\text{cov}(u_i, u_j) = 0 \qquad i \neq j$$

Homoscedasticity, or

$$\text{var}(u_i) = \sigma^2$$

Zero covariance between u_i and each X variable, or

$$\text{cov}(u_i, X_{2i}) = \text{cov}(u_i, X_{3i}) = 0$$

No specification bias, or

The model is correctly specified

No exact collinearity between the X variables, or

No **exact linear relationship** between X_2 and X_3

we also assume that the multiple regression model is *linear in the parameters*, that the values of the regressors are fixed in repeated sampling, and that there is sufficient variability in the values of the regressors.

Given the assumptions of the classical regression model, it follows that, on taking the conditional expectation of *Y* on both sides of $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$, we obtain

$$E(Y_i \mid X_{2i}, X_{3i}) = \beta_1 + \beta_2 X_{2i} + \beta_{3i} X_{3i}$$

Which gives conditional mean or expected value of *Y* conditional upon the given or fixed values of X_2 and X_3.

$\beta 1$ is the intercept. The regression coefficients β_2 and β_3 are known as partial regression or partial slope coefficients.

For example β_2 measures the *change* in the mean value of *Y* i.e. $E(Y)$, per unit change in X_2, holding the value of X_3 constant.

In econometrics, Ordinary Least Squares (OLS) method is widely used to estimate the parameter of a linear regression model. OLS estimators minimize the sum of the squared errors (a difference between observed values and predicted values). ... The importance of OLS assumptions cannot be overemphasized.

To find the OLS estimators, let us first write the sample regression function (SRF) corresponding to the PRF of as follows:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$$

where $\hat{u}_i$ is the residual term, the sample counterpart (estimate) of the stochastic disturbance term u_i.

the OLS procedure consists in so choosing the values of the unknown parameters that the residual sum of squares (RSS) $\sum \hat{u}_i^2$ is as small as possible. Symbolically,

$$\min \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2$$

For minimum , we have partially differentiate $\sum \hat{u}_i^2$ with respect to the *β* coefficients, and it equal to zero.

Thus for minimum, $\sum \hat{u}_i^2$, we require

$$\frac{\partial \sum \hat{u}_i^2}{\partial \sum \hat{\beta}_1} = 0 \longrightarrow \text{(A)}$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = 0 \longrightarrow \text{(B)}$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_3} = 0 \longrightarrow \text{(C)}$$

Eqn (A) $\Rightarrow$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} = \frac{\partial}{\partial \hat{\beta}_1} \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} \right)^2 = 0$$

$$\Rightarrow 2 \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} \right) = 0$$

$$\Rightarrow \sum Y_i - \sum \hat{\beta}_1 - \hat{\beta}_2 \sum X_{2i} - \hat{\beta}_3 \sum X_{3i} = 0$$

$$\Rightarrow \bar{Y} - \beta_1 - \beta_2 \bar{X}_2 - \beta_3 \bar{X}_3 = 0$$ [Taking Average in both side]

$$\Rightarrow \beta_1 = \bar{Y} - \beta_2 \bar{X}_2 - \beta_3 \bar{X}_3 \longrightarrow (1)$$

Again, $$\frac{\partial \Sigma \hat{u}_i^2}{\partial \beta_2} = \frac{\partial}{\partial \hat{\beta}_2} \Sigma (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2 = 0$$

$$\Rightarrow 2 \Sigma (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \beta_3 X_{3i})(-X_{2i}) = 0$$

$$\Rightarrow \sum Y_i x_{2i} - \hat{\beta}_1 \sum x_{2i} - \hat{\beta}_2 \sum x_{2i}^2 - \hat{\beta}_3 \sum x_{2i} x_{3i} = 0$$

$$\Rightarrow \sum Y_i x_{2i} - (\bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3) \sum x_{2i} - \hat{\beta}_2 \sum x_{2i}^2 - \hat{\beta}_3 \sum x_{2i} x_{3i} = 0$$

[using eqn (1)]

$$\Rightarrow \sum Y_i x_{2i} + \bar{Y} \sum x_{2i} + \hat{\beta}_2 \bar{X} \sum x_{2i} + \hat{\beta}_3 \bar{X}_3 \sum x_{2i} - \hat{\beta}_2 \sum x_{2i}^2 - \hat{\beta}_3 \sum x_{2i} x_{3i} = 0$$

$$\Rightarrow \sum Y_i x_{2i} - \bar{Y} \sum x_{2i} - \hat{\beta}_2 \left(\sum x_{2i}^2 - \bar{X} \sum x_{2i}\right) - \hat{\beta}_3 \left(\sum x_{2i} x_{3i} - \bar{X}_3 \sum x_{2i}\right) = 0$$

$$\Rightarrow \sum (Y_i - \bar{Y})(X_{2i} - \bar{X}_2) - \hat{\beta}_2 \sum (X_{2i} - \bar{X}_2)^2 - \hat{\beta}_3 \sum (X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3) = 0$$

$$\Rightarrow \sum y_i x_{2i} - \hat{\beta}_2 \sum x_{2i}^2 - \hat{\beta}_3 \sum x_{2i} x_{3i} = 0$$

$$\Rightarrow \hat{\beta}_2 \sum x_{2i}^2 + \hat{\beta}_3 \sum x_{2i} x_{3i} = \sum y_i x_{2i} \longrightarrow (E)$$

In similar fashion, from eqn (C), we get following normal equation

$$\hat{\beta}_2 \sum x_{2i} x_{3i} + \hat{\beta}_3 \sum x_{3i}^2 = \sum y_i x_{3i} \longrightarrow (F)$$

Applying Crammer's rule in Eqn (E) & (F), we get

$$\hat{\beta}_2 = \frac{\begin{vmatrix} \sum y_i x_{2i} & \sum x_{2i} x_{3i} \\ \sum y_i x_{3i} & \sum x_{3i}^2 \end{vmatrix}}{\begin{vmatrix} \sum x_{2i}^2 & \sum x_{2i} x_{3i} \\ \sum x_{2i} x_{3i} & \sum x_{3i}^2 \end{vmatrix}}$$

$$= \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

||ly, $$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

A. 1. Expected value of Regressions Coefficients (3-Varaibles)

$$\left.\begin{aligned} E(\hat{\beta}_1) &= \beta_1 \\ E(\hat{\beta}_2) &= \beta_2 \\ E(\hat{\beta}_3) &= \beta_3 \end{aligned}\right\} \text{unbiased estimates.}$$

A.2. Variances and Standard Errors of OLS Estimators (3 Variables)

The relevant formulas are as follows :

$$\operatorname{var}(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}_2^2 \sum x_{3i}^2 + \bar{X}_3^2 \sum x_{2i}^2 - 2\bar{X}_2\bar{X}_3 \sum x_{2i}x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i}x_{3i}\right)^2}\right] \cdot \sigma^2$$

$$\operatorname{se}(\hat{\beta}_1) = +\sqrt{\operatorname{var}(\hat{\beta}_1)}$$

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sum x_{3i}^2}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i}x_{3i}\right)^2}\sigma^2$$

or, equivalently,

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 \left(1 - r_{23}^2\right)}$$

where r_{23} is the sample coefficient of correlation between X_2 and X_3

$$\text{se}(\hat{\beta}_2) = +\sqrt{\text{var}(\hat{\beta}_2)}$$

$$\text{var}(\hat{\beta}_3) = \frac{\sum x_{2i}^2}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i}x_{3i}\right)^2}\sigma^2$$

or, equivalently,

$$\text{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2\left(1 - r_{23}^2\right)}$$

$$\text{se}(\hat{\beta}_3) = +\sqrt{\text{var}(\hat{\beta}_3)}$$

$$\text{cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\sigma^2}{\left(1 - r_{23}^2\right)\sqrt{\sum x_{2i}^2}\sqrt{\sum x_{3i}^2}}$$

In all these formulas σ^2 is the (homoscedastic) variance of the population disturbances u_i. An unbiased estimator of σ^2 is given by

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-3}$$

A.3. Properties of OLS Estimators (3-varaibles)

1. **Given the assumptions of the classical linear regression model, the partial regression coefficients not only are linear and unbiased and efficient (minimum variance). I.e. *they are BLUE :* Put differently, they satisfy the Gauss-Markov theorem.**

2. **The three-variable regression line (surface) passes through the means i.e Y bar, X_2 bar and X_3 bar.**

3. **The mean value of the estimated Y_i (= $\hat{Y}_i$) is equal to the mean value of the actual Y_i**

4. $\sum \hat{u}_i = \bar{\hat{u}} = 0,$

.

5. The residuals $\hat{u}_i$ are uncorrelated with X_{2i} and X_{3i},

6. The residuals $\hat{u}_i$ are uncorrelated with $\hat{Y}_i$; that is, $\sum \hat{u}_i \hat{Y}_i = 0.$

The Gauss-Markov (GM) theorem states that for an additive linear model, and under the ”standard” GM assumptions that the errors are uncorrelated and homoscedastic with expectation value zero, the Ordinary Least Squares (OLS) estimator has the lowest sampling variance within the class of linear unbiased estimators.

In econometrics, Ordinary Least Squares (OLS) method is widely used to estimate the parameter of a linear regression model. OLS estimators minimize the sum of the squared errors (a difference between observed values and predicted values).

A.4. THE MULTIPLE COEFFICIENT OF DETERMINATION R^2

We have S.R.M :

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \longrightarrow (A)$$

$$= \hat{Y}_i + \hat{u}_i \longrightarrow (B)$$

Taking averages along eqn (A), we get

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \beta_3 \bar{X}_3 \longrightarrow (C)$$

Now,

eqn (A) − eqn (C) ⇒

$$(Y_i - \bar{Y}) = \beta_2 (X_{2i} - \bar{X}_2) + \beta_3 (X_{3i} - \bar{X}_3) + \hat{u}_i$$

$$\Rightarrow y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + \hat{u}_i$$

$$\Rightarrow y_i = \hat{y}_i + \hat{u}_i \longrightarrow (D)$$

Squaring & Taking summation of both sides in eqn (II) we get

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2\sum \hat{y}_i \hat{u}_i$$

$$= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 0 \quad \left[\text{since } \hat{y}_i \,\&\, \hat{u}_i \text{ are uncorrelated}\right]$$

$$= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 \longrightarrow (E)$$

Eqn E states that

TSS = ESS + RSS

Now, by definition

$$R^2 = \frac{\text{ESS}}{\text{TSS}}$$

$$= \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2}$$

R^2 can also be computed as follows:

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} = 1 - \frac{(n-3)\hat{\sigma}^2}{(n-1)S_y^2}$$

A. 5. R^2 AND THE ADJUSTED R^2

An important property of R^2 is that it is a nondecreasing function of the number of explanatory variables or regressors present in the model; as the number of regressors increases, R_2 almost invariably increases and never decreases. To understand this, let us consider the he definition of the coefficient of determination:

$$
\begin{aligned}
R^2 &= \frac{\mathrm{ESS}}{\mathrm{TSS}} \\
&= 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}} \\
&= 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2}
\end{aligned}
$$

Now $\sum y_i^2$ is independent of the number of X variables in the model because it is simply $\sum (Y_i - \bar{Y})^2$. The RSS, $\sum \hat{u}_i^2$, however, depends on the number of regressors present in the model. Intuitively, it is clear that as the number of X variables increases, $\sum \hat{u}_i^2$ is likely to decrease (at least it will not increase); hence R^2 as defined in (7.8.1) will increase. In view of this, in comparing

To do away with the limitation We use R^2 , adjusted R^2 **in case of multiple regression model.**

Adjusted R^2 **is given by the following formula.**

$$\bar{R}^2 = 1 - \frac{\sum \hat{u}_i^2/(n-k)}{\sum y_i^2/(n-1)} \qquad \textbf{(7.8.2)}$$

where k = the number of parameters in the model *including the intercept term*. (In the three-variable regression, $k = 3$. Why?) The R^2 thus defined is known as the **adjusted** $\boldsymbol{R^2}$, denoted by $\bar{R}^2$. The term *adjusted* means adjusted for the df associated with the sums of squares entering into (7.8.1): $\sum \hat{u}_i^2$ has $n - k$ df in a model involving k parameters, which include

Which can be written as can also be written as

$$\bar{R}^2 = 1 - \frac{\hat{\sigma}^2}{S_Y^2} \qquad \textbf{(7.8.3)}$$

where $\hat{\sigma}^2$ is the residual variance, an unbiased estimator of true σ^2, and S_Y^2 is the sample variance of Y.

It is easy to see that $\bar{R}^2$ and R^2 are related as follows:

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n - 1}{n - k} \tag{7.8.4}$$

It is immediately apparent from Eq. (7.8.4) that (1) for $k > 1$, $\bar{R}^2 < R^2$ which implies that as the number of X variables increases, the adjusted R^2 increases less than the unadjusted R^2; and (2) $\bar{R}^2$ can be negative, although R^2 is necessarily nonnegative.[10] In case $\bar{R}^2$ turns out to be negative in an application, its value is taken as zero.

Which R^2 should one use in practice? As Theil notes:

> . . . it is good practice to use $\bar{R}^2$ rather than R^2 because R^2 tends to give an overly optimistic picture of the fit of the regression, particularly when the number of explanatory variables is not very small compared with the number of observations.[11]

. B. K- Variable Regression Analysis : THE PROBLEM OF ESTIMATION

THE *k*-VARIABLE LINEAR REGRESSION MODEL

If we generalize the two- and three-variable linear regression models, the k-variable population regression model (PRF) involving the dependent variable Y and $k - 1$ explanatory variables $X_2, X_3, \ldots, X_k$ may be written as

$$\text{PRF: } Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \qquad i = 1, 2, 3, \ldots, n \tag{C.1.1}$$

where β_1 = the intercept, β_2 to β_k = partial slope coefficients, u = stochastic disturbance term, and i = ith observation, n being the size of the population. The PRF (C.1.1) is to be interpreted in the usual manner: It gives the mean or expected value of Y conditional upon the fixed (in repeated sampling) values of $X_2, X_3, \ldots, X_k$, that is, $E(Y \mid X_{2i}, X_{3i}, \ldots, X_{ki})$.

Equation (C.1.1) is a shorthand expression for the following set of n simultaneous equations:

$$
\begin{aligned}
Y_1 &= \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \cdots + \beta_k X_{k1} + u_1 \\
Y_2 &= \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \cdots + \beta_k X_{k2} + u_2 \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
Y_n &= \beta_1 + \beta_2 X_{2n} + \beta_3 X_{3n} + \cdots + \beta_k X_{kn} + u_n
\end{aligned}
\tag{C.1.2}
$$

Let us write the system of equations (C.1.2) in an alternative but more illuminating way as follows[2]:

$$
\begin{array}{ccccccc}
\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} & = & \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix} & \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} & + & \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \\
\mathbf{y} & = & \mathbf{X} & \boldsymbol{\beta} & + & \mathbf{u} \\
n \times 1 & & n \times k & k \times 1 & & n \times 1
\end{array}
\tag{C.1.3}
$$

where $\mathbf{y} = n \times 1$ column vector of observations on the dependent variable Y

$\mathbf{X} = n \times k$ matrix giving n observations on $k - 1$ variables X_2 to X_k, the first column of 1's representing the intercept term (this matrix is also known as the **data matrix**)

$\boldsymbol{\beta} = k \times 1$ column vector of the unknown parameters $\beta_1, \beta_2, \ldots, \beta_k$

$\mathbf{u} = n \times 1$ column vector of n disturbances u_i

Using the rules of matrix multiplication and addition, the reader should verify that systems (C.1.2) and (C.1.3) are equivalent.

System (C.1.3) is known as the *matrix representation of the general (k-variable) linear regression model*. It can be written more compactly as

$$\underset{n \times 1}{\mathbf{y}} = \underset{n \times k}{\mathbf{X}} \underset{k \times 1}{\boldsymbol{\beta}} + \underset{n \times 1}{\mathbf{u}} \tag{C.1.4}$$

Where there is no confusion about the dimensions or orders of the matrix **X** and the vectors **y**, **β**, and **u**, Eq. (C.1.4) may be written simply as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{C.1.5}$$

B.2 ASSUMPTIONS OF THE CLASSICAL LINEAR REGRESSION MODEL IN MATRIX NOTATION

ASSUMPTIONS OF THE CLASSICAL LINEAR REGRESSION MODEL

Scalar notation		Matrix notation
1. $E(u_i) = 0, \quad$ for each i	(3.2.1)	**1.** $E(\mathbf{u}) = \mathbf{0}$ where $\mathbf{u}$ and $\mathbf{0}$ are $n \times 1$ column vectors, $\mathbf{0}$ being a null vector
2. $E(u_i u_j) = 0 \quad i \neq j$ $= \sigma^2 \quad i = j$	(3.2.5) (3.2.2)	**2.** $E(\mathbf{uu}') = \sigma^2 \mathbf{I}$ where $\mathbf{I}$ is an $n \times n$ identity matrix
3. $X_2, X_3, \ldots, X_k$ are nonstochastic or fixed		**3.** The $n \times k$ matrix $\mathbf{X}$ is nonstochastic, that is, it consists of a set of fixed numbers
4. There is no exact linear relationship among the X variables, that is, no multicollinearity	(7.1.7)	**4.** The rank of $\mathbf{X}$ is $p(\mathbf{X}) = k$, where k is the number of columns in $\mathbf{X}$ and k is less than the number of observations, n
5. For hypothesis testing, $u_i \sim N(0, \sigma^2)$	(4.2.4)	**5.** The $\mathbf{u}$ vector has a multivariate normal distribution, i.e., $\mathbf{u} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$

Assumption 1 means that the expected value of the disturbance vector **u**, that is, of each of its elements, is zero. More explicitly, $E(\mathbf{u}) = \mathbf{0}$ means

$$E\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} E(u_1) \\ E(u_2) \\ \vdots \\ E(u_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \textbf{(C.2.1)}$$

Assumption 2. $E(\mathbf{uu}') = \sigma^2\mathbf{I}$
where $\mathbf{I}$ is an $n \times n$ identity matrix

$$E(\mathbf{uu}') = E\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}[u_1 \quad u_2 \quad \cdots \quad u_n]$$

where $\mathbf{u}'$ is the transpose of the column vector $\mathbf{u}$, or a row vector. Performing the multiplication, we obtain

$$E(\mathbf{uu}') = E\begin{bmatrix} u_1^2 & u_1u_2 & \cdots & u_1u_n \\ u_2u_1 & u_2^2 & \cdots & u_2u_n \\ \cdots & \cdots & \cdots & \cdots \\ u_nu_1 & u_nu_2 & \cdots & u_n^2 \end{bmatrix}$$

Applying the expectations operator E to each element of the preceding matrix, we obtain

$$E(\mathbf{uu}') = \begin{bmatrix} E(u_1^2) & E(u_1u_2) & \cdots & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & \cdots & E(u_2u_n) \\ \cdots & \cdots & \cdots & \cdots \\ E(u_nu_1) & E(u_nu_2) & \cdots & E(u_n^2) \end{bmatrix} \qquad \textbf{(C.2.2)}$$

Because of the assumptions of homoscedasticity and no serial correlation, matrix (C.2.2) reduces to

$$
\begin{aligned}
E(\mathbf{u}\mathbf{u}') &= \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \qquad \textbf{(C.2.3)} \\
&= \sigma^2 \mathbf{I}
\end{aligned}
$$

where **I** is an $n \times n$ identity matrix.

Matrix (C.2.2) [and its representation given in (C.2.3)] is called the **variance–covariance matrix** of the disturbances u_i; the elements on the main diagonal of this matrix (running from the upper left corner to the lower right corner) give the variances, and the elements off the main diagonal give the covariances.[4] Note that the variance–covariance matrix is **symmetric:** The elements above and below the main diagonal are reflections of one another.

Assumption 3 states that the $n \times k$ matrix X is nonstochastic; that is, it consists of fixed numbers. The regression analysis is conditional regression analysis, conditional upon the fixed values of the *X* variables.

Assumption 4 states that the X matrix has full column rank equal to *k*, the number of columns in the matrix. This means that the columns of the *X* matrix are linearly independent; that is, there is no exact linear relationship among the *X* variables. In other words there is no multicollinearity. In scalar notation this is equivalent to saying that there exists no set of numbers λ_1, λ_2, . . . , λ_k not all zero such that

$$\lambda_1 X_{1i} + \lambda_2 X_{2i} + \cdots + \lambda_k X_{ki} = 0 \qquad \textbf{(C.2.4)}$$

where $X_{1i} = 1$ for all i (to allow for the column of 1's in the $\mathbf{X}$ matrix). In matrix notation, (C.2.4) can be represented as

$$\boldsymbol{\lambda}'\mathbf{x} = 0 \qquad \textbf{(C.2.5)}$$

where $\boldsymbol{\lambda}'$ is a $1 \times k$ row vector and $\mathbf{x}$ is a $k \times 1$ column vector.

If an exact linear relationship such as (C.2.4) exists, the variables are said to be collinear. If, on the other hand, (C.2.4) holds true only if $\lambda_1 = \lambda_2 = \lambda_3 = \cdots = 0$, then the *X* variables are said to be linearly independent.

B.3. OLS ESTIMATION: K variables

To obtain the OLS estimate of $\boldsymbol{\beta}$, let us first write the k-variable sample regression (SRF):

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + \hat{u}_i \qquad \textbf{(C.3.1)}$$

which can be written more compactly in matrix notation as

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}} \qquad \textbf{(C.3.2)}$$

and in matrix form as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} + \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix} \qquad \textbf{(C.3.3)}$$

$$\underset{n \times 1}{\mathbf{y}} = \underset{n \times k}{\mathbf{X}} \quad \underset{k \times 1}{\hat{\boldsymbol{\beta}}} + \underset{n \times 1}{\hat{\mathbf{u}}}$$

where $\hat{\boldsymbol{\beta}}$ is a k-element column vector of the OLS estimators of the regression coefficients and where $\hat{\mathbf{u}}$ is an $n \times 1$ column vector of n residuals.

As in the two- and three-variable models, in the k-variable case the OLS estimators are obtained by minimizing

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})^2 \qquad \textbf{(C.3.4)}$$

where $\sum \hat{u}_i^2$ is the residual sum of squares (RSS). In matrix notation, this amounts to minimizing $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ since

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = [\hat{u}_1 \quad \hat{u}_2 \quad \cdots \quad \hat{u}_n]\begin{bmatrix}\hat{u}_1\\ \hat{u}_2\\ \vdots\\ \hat{u}_n\end{bmatrix} = \hat{u}_1^2 + \hat{u}_2^2 + \cdots + \hat{u}_n^2 = \sum \hat{u}_i^2 \qquad \textbf{(C.3.5)}$$

Now from (C.3.2) we obtain

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \qquad \textbf{(C.3.6)}$$

Therefore,

$$\begin{aligned}\hat{\mathbf{u}}'\hat{\mathbf{u}} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}\end{aligned} \qquad \textbf{(C.3.7)}$$

where use is made of the properties of the transpose of a matrix, namely, $(\mathbf{X}\hat{\boldsymbol{\beta}})' = \hat{\boldsymbol{\beta}}'\mathbf{X}'$; and since $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$ is a scalar (a real number), it is equal to its transpose $\mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}}$.

Equation (C.3.7) is the matrix representation of (C.3.4). In scalar notation, the method of OLS consists in so estimating $\beta_1, \beta_2, \ldots, \beta_k$ that $\sum \hat{u}_i^2$ is as small as possible. This is done by differentiating (C.3.4) partially with respect to $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ and setting the resulting expressions to zero. This process yields k simultaneous equations in k unknowns, the normal equations of the least-squares theory. As shown in Appendix CA, Section CA.1, these equations are as follows:

$$\begin{aligned}
n\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2i} + \hat{\beta}_3 \sum X_{3i} + \cdots + \hat{\beta}_k \sum X_{ki} &= \sum Y_i \\
\hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i}X_{3i} + \cdots + \hat{\beta}_k \sum X_{2i}X_{ki} &= \sum X_{2i}Y_i \\
\hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{3i}X_{2i} + \hat{\beta}_3 \sum X_{3i}^2 + \cdots + \hat{\beta}_k \sum X_{3i}X_{ki} &= \sum X_{3i}Y_i \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
\hat{\beta}_1 \sum X_{ki} + \hat{\beta}_2 \sum X_{ki}X_{2i} + \hat{\beta}_3 \sum X_{ki}X_{3i} + \cdots + \hat{\beta}_k \sum X_{ki}^2 &= \sum X_{ki}Y_i
\end{aligned}$$

In matrix form, Eq. (C.3.8) can be represented as

$$\underbrace{\begin{bmatrix} n & \sum X_{2i} & \sum X_{3i} & \cdots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} & \cdots & \sum X_{2i}X_{ki} \\ \sum X_{3i} & \sum X_{3i}X_{2i} & \sum X_{3i}^2 & \cdots & \sum X_{3i}X_{ki} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum X_{ki} & \sum X_{ki}X_{2i} & \sum X_{ki}X_{3i} & \cdots & \sum X_{ki}^2 \end{bmatrix}}_{(\mathbf{X}'\mathbf{X})} \underbrace{\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}}_{\hat{\boldsymbol{\beta}}} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{21} & X_{22} & \cdots & X_{2n} \\ X_{31} & X_{32} & \cdots & X_{3n} \\ \cdots & \cdots & \cdots & \cdots \\ X_{k1} & X_{k2} & \cdots & X_{kn} \end{bmatrix}}_{\mathbf{X}'} \underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}}_{\mathbf{y}} \tag{C.3.9}$$

or, more compactly, as

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \tag{C.3.10}$$

Note these features of the $(\mathbf{X}'\mathbf{X})$ matrix: (1) It gives the raw sums of squares and cross products of the X variables, one of which is the intercept term taking the value of 1 for each observation. The elements on the main diagonal give the raw sums of squares, and those off the main diagonal give the raw sums of cross products (by *raw* we mean in original units of measurement). (2) It is symmetrical since the cross product between X_{2i} and X_{3i} is the same as that between X_{3i} and X_{2i}. (3) It is of order $(k \times k)$, that is, k rows and k columns.

In (C.3.10) the known quantities are $(\mathbf{X}'\mathbf{X})$ and $(\mathbf{X}'\mathbf{y})$ (the cross product between the X variables and y) and the unknown is $\hat{\boldsymbol{\beta}}$. Now using matrix algebra, if the inverse of $(\mathbf{X}'\mathbf{X})$ exists, say, $(\mathbf{X}'\mathbf{X})^{-1}$, then premultiplying both sides of (C.3.10) by this inverse, we obtain

$$(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

But since $(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X}) = \mathbf{I}$, an identity matrix of order $k \times k$, we get

$$\mathbf{I}\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

or

$$\underset{k \times 1}{\hat{\boldsymbol{\beta}}} = \underset{k \times k}{(\mathbf{X}'\mathbf{X})^{-1}} \underset{(k \times n)}{\mathbf{X}'} \underset{(n \times 1)}{\mathbf{y}} \tag{C.3.11}$$

Equation (C.3.11) is a fundamental result of the OLS theory in matrix notation. It shows how the $\hat{\boldsymbol{\beta}}$ vector can be estimated from the given data.

B.4 . Variance –Covariance matrix of Coefficients

Variance–Covariance Matrix of $\hat{\boldsymbol\beta}$

Matrix methods enable us to develop formulas not only for the variance of $\hat\beta_i$, any given element of $\hat{\boldsymbol\beta}$, but also for the covariance between any two elements of $\hat{\boldsymbol\beta}$, say, $\hat\beta_i$ and $\hat\beta_j$. We need these variances and covariances for the purpose of statistical inference.

By definition, the variance–covariance matrix of $\hat{\boldsymbol\beta}$ is [cf. (C.2.2)]

$$\text{var-cov}(\hat{\boldsymbol\beta}) = E\{[\hat{\boldsymbol\beta} - E(\hat{\boldsymbol\beta})][\hat{\boldsymbol\beta} - E(\hat{\boldsymbol\beta})]'\}$$

which can be written explicitly as

$$\text{var-cov}(\hat{\boldsymbol\beta}) = \begin{bmatrix} \text{var}(\hat\beta_1) & \text{cov}(\hat\beta_1, \hat\beta_2) & \cdots & \text{cov}(\hat\beta_1, \hat\beta_k) \\ \text{cov}(\hat\beta_2, \hat\beta_1) & \text{var}(\hat\beta_2) & \cdots & \text{cov}(\hat\beta_2, \hat\beta_k) \\ \cdots & \cdots & \cdots & \cdots \\ \text{cov}(\hat\beta_k, \hat\beta_1) & \text{cov}(\hat\beta_k, \hat\beta_2) & \cdots & \text{var}(\hat\beta_k) \end{bmatrix} \tag{C.3.12}$$

It is shown in Appendix CA, Section CA.3, that the preceding variance–covariance matrix can be obtained from the following formula:

$$\text{var-cov}\,(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \tag{C.3.13}$$

where σ^2 is the homoscedastic variance of u_i and $(\mathbf{X}'\mathbf{X})^{-1}$ is the inverse matrix appearing in Eq. (C.3.11), which gives the OLS estimator $\hat{\boldsymbol{\beta}}$.

In the two- and three-variable linear regression models an unbiased estimator of σ^2 was given by $\hat{\sigma}^2 = \sum \hat{u}_i^2/(n-2)$ and $\hat{\sigma}^2 = \sum \hat{u}_i^2/(n-3)$, respectively. In the k-variable case, the corresponding formula is

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\sum \hat{u}_i^2}{n-k}\\ &= \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k}\end{aligned} \tag{C.3.14}$$

where there are now $n-k$ df. (Why?)

Although in principle $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ can be computed from the estimated residuals, in practice it can be obtained directly as follows. Recalling that $\sum \hat{u}_i^2$ $(=\text{RSS}) = \text{TSS} - \text{ESS}$, in the two-variable case we may write

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2 \tag{3.3.6}$$

and in the three-variable case

which can be written explicitly as

$$\text{var-cov}(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} \operatorname{var}(\hat{\beta}_1) & \operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) & \cdots & \operatorname{cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \operatorname{cov}(\hat{\beta}_2, \hat{\beta}_1) & \operatorname{var}(\hat{\beta}_2) & \cdots & \operatorname{cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \cdots & \cdots & \cdots & \cdots \\ \operatorname{cov}(\hat{\beta}_k, \hat{\beta}_1) & \operatorname{cov}(\hat{\beta}_k, \hat{\beta}_2) & \cdots & \operatorname{var}(\hat{\beta}_k) \end{bmatrix} \tag{C.3.12}$$

It is shown in Appendix CA, Section CA.3, that the preceding variance–covariance matrix can be obtained from the following formula:

$$\text{var-cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \tag{C.3.13}$$

where σ^2 is the homoscedastic variance of u_i and $(\mathbf{X}'\mathbf{X})^{-1}$ is the inverse matrix appearing in Eq. (C.3.11), which gives the OLS estimator $\hat{\boldsymbol{\beta}}$.

In the two- and three-variable linear regression models an unbiased estimator of σ^2 was given by $\hat{\sigma}^2 = \sum \hat{u}_i^2/(n-2)$ and $\hat{\sigma}^2 = \sum \hat{u}_i^2/(n-3)$, respectively. In the k-variable case, the corresponding formula is

$$\begin{aligned} \hat{\sigma}^2 &= \frac{\sum \hat{u}_i^2}{n-k} \\ &= \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k} \end{aligned} \tag{C.3.14}$$

where there are now $n - k$ df. (Why?)

Although in principle $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ can be computed from the estimated residuals, in practice it can be obtained directly as follows. Recalling that $\sum \hat{u}_i^2$ $(= \text{RSS}) = \text{TSS} - \text{ESS}$, in the two-variable case we may write

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2 \tag{3.3.6}$$

and in the three-variable case

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i} \tag{7.4.19}$$

By extending this principle, it can be seen that for the k-variable model

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \cdots - \hat{\beta}_k \sum y_i x_{ki} \tag{C.3.15}$$

In matrix notation,

$$\text{TSS:} \quad \sum y_i^2 = \mathbf{y}'\mathbf{y} - n\bar{Y}^2 \tag{C.3.16}$$

$$\text{ESS:} \quad \hat{\beta}_2 \sum y_i x_{2i} + \cdots + \hat{\beta}_k \sum y_i x_{ki} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2 \tag{C.3.17}$$

where the term $n\bar{Y}^2$ is known as the correction for mean.[6] Therefore,

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \tag{C.3.18}$$

Once $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ is obtained, $\hat{\sigma}^2$ can be easily computed from (C.3.14), which, in turn, will enable us to estimate the variance–covariance matrix (C.3.13).

B.3. Properties of OLS ESTIMATION (K variables)

Properties of OLS Vector $\hat{\beta}$

In the two- and three-variable cases we know that the OLS estimators are linear and unbiased, and in the class of all linear unbiased estimators they have minimum variance (the Gauss–Markov property). In short, the OLS estimators are best linear unbiased estimators (BLUE). This property extends to the entire $\hat{\boldsymbol{\beta}}$ vector; that is, $\hat{\boldsymbol{\beta}}$ is linear (each of its elements is a linear function of Y, the dependent variable). $E(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}}$, that is, the expected value of each element of $\hat{\boldsymbol{\beta}}$ is equal to the corresponding element of the true $\boldsymbol{\beta}$, and in the class of all linear unbiased estimators of $\boldsymbol{\beta}$, the OLS estimator $\hat{\boldsymbol{\beta}}$ has minimum variance.

THE COEFFICIENT OF DETERMINATION R^2 IN MATRIX NOTATION

The coefficient of determination R^2 has been defined as

$$R^2 = \frac{\text{ESS}}{\text{TSS}}$$

In the two-variable case,

$$R^2 = \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum y_i^2} \tag{3.5.6}$$

and in the three-variable case

$$R^2 = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2} \tag{7.5.5}$$

Generalizing we obtain for the k-variable case

$$R^2 = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} + \cdots + \hat{\beta}_k \sum y_i x_{ki}}{\sum y_i^2} \tag{C.4.1}$$

By using (C.3.16) and (C.3.17), Eq. (C.4.1) can be written as

$$\boxed{R^2 = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2}{\mathbf{y}'\mathbf{y} - n\bar{Y}^2}} \tag{C.4.2}$$

which gives the matrix representation of R^2.

Source:

D N Gujarati: Basic Econometrics

Entire Note here is based on the above source.

Samir K Mahajan