CHAPTER 3

THE LINEAR MODEL: METHODS OF ESTIMATION AND SUMS OF SQUARES OF LINEAR FUNCTIONS

3.1 INTRODUCTION

The fitting of models as discussed in this module, consists of the following five steps:

1. *Model specification.* A specification of a statistical model requires an equation that equates the response variable with the explanatory variables, the probability distributions of the response variable and the error term.
2. *Estimation* of the parameters of the model.
3. Determination of the *applicability* of the model.
4. *Inference.* The calculation of confidence intervals or regions and the testing of hypotheses concerning the parameters in the model.
5. *Interpretation* of the results.

3.2 LINEAR STATISTICAL MODELS

A statistical model can, for the goal of this course, be defined as *a mathematical equation containing variates (stochastic variables), ordinary mathematical variables and parameters.* A multivariate linear model can now be described as follows: Given a set D of m-component real vectors so that for every $\underline{x}' = (x_1, x_2, \ldots, x_m) \in D$ a variate $Y_{\underline{x}}$ with p.d.f. $f_{Y_{\underline{x}}}(.)$ exists such that

$$E(Y_{\underline{x}}) \equiv \mu(\underline{x}) \equiv \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m \tag{3.2.1}$$

and $\text{var}(Y_{\underline{x}}) = \sigma^2$ independent from $\underline{x}$

or, also

$$Y_{\underline{x}} = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + E_{\underline{x}} \tag{3.2.2}$$

where $E(E_{\underline{x}}) = 0$ and $\text{var}(E_{\underline{x}}) = \sigma^2$.

$E_{\underline{x}}$ is called the error term in the model.

The goal is now for, say n values $\underline{x}$ in D, to observe $Y_{\underline{x}}$ from which conclusions about the values of $\beta_1, \beta_2, \ldots, \beta_m$, $E(Y_{\underline{x}})$ (for any given $\underline{x} \in D$) and σ^2 are made. Indicate the n $\underline{x}$-values in D, for which $Y_{\underline{x}}$ is observed, by $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ where $\underline{x}_i' = (x_{i1}, x_{i2}, \ldots, x_{im})$ and the variates $Y_{\underline{x}}$ and $E_{\underline{x}}$ associated with $\underline{x} = \underline{x}_i$ by Y_i and E_i respectively, $i = 1, 2, \ldots, n$. The n $\underline{x}$-values are considered to be given. Let $y_1, y_2, \ldots, y_n$ be random observations of $Y_1, Y_2, \ldots, Y_n$ respectively. These observed y-values $y_1, y_2, \ldots, y_n$ will therefore vary from sample to sample while it is assumed that for every sample the same set of $\underline{x}$-values are observed. It is also assumed that not all of the $\underline{x}$-values are equal.

The assumptions of the model can now be summarised as follows:

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_m x_{im} + E_i, \quad i = 1, 2, \ldots, n \tag{3.2.3}$$

where the $\{E_i, \; i = 1, 2, \ldots, n\}$ variates satisfy

$$E(E_i) = 0 \text{ and } \mathrm{var}(E_i) = \sigma^2. \tag{3.2.4}$$

It is also assumed that the variates $E_1, E_2, \ldots, E_n$ are pair-wise uncorrelated, so that

$$E(E_i E_j) = 0 \quad \text{for } i \neq j \text{ and } i, j = 1, 2, \ldots, n. \tag{3.2.5}$$

When the coefficients $\{x_{ij}\}$ are the values of indicator variables (i.e. variables that can only take on the values 0 or 1) which only indicate the absence or presence of the effects $\{\beta_j\}$ in the circumstances under which the observations are made, then we have an *analysis of variance.* If the $\{x_{ij}\}$ are not indicator variables, but take on continuous values like time, temperature, mass, *etc*. then we have *regression analysis.* In the latter case, the $\{x_{ij}\}$ are the independent (explanatory) variables while the $\{Y_i\}$ are the dependent variables. If there are $\{x_{ij}\}$ of both types (*i.e.* indicator variables as well as variables that can take on continuous values), then we have *covariance analysis*.

The unknown effects $\{\beta_j\}$ can be unknown constants which we call *parameters* or unobservable variates which are subject to further assumptions concerning their distribution. This distribution then contains further unknown parameters. Often, one of the $\{\beta_j\}$, say β_k is a constant that has a coefficient of 1 for every observation ($x_{ik} = 1$ for all i). Such a β_k is called an *additive constant.*

A model in which all the $\{\beta_j\}$ are unknown parameters is called a *fixed effects* model. A model in which all the $\{\beta_j\}$, except with the possible exclusion of one additive constant, are variates, is called a *stochastic effects* model. A fixed effects variance analysis model is also called a *Model I variance analysis* while a stochastic variance analysis model is also known as a *variance components model.* Models in which at least one of the $\{\beta_j\}$ is a variate and at least one of the $\{\beta_j\}$ is a constant, but not an additive constant, are called *mixed models*.

A *linear statistical* model is a model which is linear with respect to the effects $\{\beta_j\}$. Examples of linear models are

$$Y_x = \beta_1 + \beta_2 x + E_x \tag{3.2.6}$$

and

$$Y_x = \beta_1 + \beta_2 x + \beta_{22} x^2 + E_x . \tag{3.2.7}$$

The largest power of the independent variable in the case of a linear model is called the *order* of the model. Model (3.2.7) is therefore a second-order linear model.

A *non-linear statistical model* is a model which is not linear in the effects $\{\beta_j\}$. Non-linear statistical models can be grouped into *intrinsic linear models, i.e.* models that can be converted into linear models by using transformations, and *intrinsic non-linear models, i.e.* models that cannot be converted into linear models.

$Y_{\underline{x}} = \beta_0 x_1^{\beta_1} x_2^{\beta_2} x_3^{\beta_3} E_{\underline{x}}$ is an intrinsic linear model, because it can be transformed into the linear model

$$\log(Y_{\underline{x}}) = \log(\beta_0) + \beta_1 \log(x_1) + \beta_2 \log(x_2) + \beta_3 \log(x_3) + \log(E_{\underline{x}}) .$$

The model $Y_x = \beta_0 + \beta_1 x + \beta_2 (\beta_3)^x + E_x$ is intrinsic non-linear.

In this course only linear models will be considered. Further, only *fixed-effects linear models* will be investigated during the first part of the course.

The model (3.2.3) and the assumptions (3.2.4) and (3.2.5) can be written in matrix notation as follows:

$$\underline{Y} = X\underline{\beta} + \underline{E} \tag{3.2.8}$$

where $\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$, $\underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}$, $\underline{E} = \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_n \end{bmatrix}$ and $X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$

with $E(\underline{E}) = \underline{0}$ and $\operatorname{cov}(\underline{E} ; \underline{E}') = E(\underline{E}\,\underline{E}') = \Sigma_{\underline{E}} = I_n\sigma^2$.

In all practical situations which will be considered, the number of observations will be more than the number of parameters which will be estimated, *i.e.* $n > m$.

Because $\underline{\beta}$ and σ^2 are unknown, they must be estimated. The first goal with an analysis with respect to linear statistical models is therefore also the estimation of unknown parameters. The linear model which has just been discussed is called *the general linear model* by Graybill (1976, p. 143-170). Graybill makes the following comparison:

QUANTITATIVE LINEAR MODELS	QUALITATIVE LINEAR MODELS
1. General linear model 2. Linear regression model	3. Design model 4. Variance-components model

Models 2, 3 and 4 can be seen as variations of the general linear model. According to Graybill, the most important characteristics of these 4 models are:

1. *The general linear model* $\underline{Y} = X\underline{\beta} + \underline{E}$.

Here, $\underline{Y}$: $n \times l$ is an observable vector variate, X: $n \times m$ is a matrix of fixed observable values (not variates), $\underline{\beta}$: $m \times 1$ is a vector of unobservable parameters defined in a parameter space Ω_β and $\underline{E}$: $n \times l$ is an unobservable vector variate with

$E(\underline{E}) = \underline{0}$ and $\text{cov}(\underline{E}\,;\underline{E}') = E(\underline{E}\,\underline{E}') = \Sigma_{\underline{E}} = I_n\sigma^2$.

2. *The linear regression model:* Let the k + 1 variates, $x_0, x_1, \ldots, x_k$ have a joint distribution with expected value $\underline{\mu}^*$ and covariance matrix Σ^* such that the conditional p.d.f. of $Y = (X_0 \mid X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k)$ satisfies the following:

(i) $E\{(X_0 \mid X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k)\} = \mu_Y(x_1, \ldots, x_k) = \beta_0 + \sum_{j=1}^{k}\beta_j x_j$

(*i.e.* the expected value of Y is linear in the $\{x_j\}$ and linear in the unknown parameters $\{\beta_j\}$).

(ii) $\text{Var}\{(X_0 \mid X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k)\} = \sigma^2_Y$.

3. *The design model:* The design model satisfies the definition of the general linear model except that rank $(X) = r \le m < n$. Further, the elements of X: $n \times m$ can only be 0 or 1 which form a specific pattern determined by the specific design. As a result, the matrix X is also referred to as the *design matrix.*

4. *Variance-components model:* Let the observable variate $Y_{ij\ldots m}$ be such that $Y_{ij\ldots m} = \mu + A_i + B_{ij} + \ldots + E_{ij\ldots m}$. Here, μ is an unknown constant, A_i a variate with expected value 0 and variance σ^2_A, B_{ij} a variate with expected value 0 and variance σ^2_B, …., $E_{ij\ldots m}$ a variate with expected value 0 and variance σ^2_e. All the variates A_i, B_{ij}, …., $E_{ij\ldots m}$ are uncorrelated.

3.3 THE METHOD OF LEAST SQUARES AND LEAST SQUARES ESTIMATORS

Consider the model (3.2.8) where $\underline{Y}$ is an $(n \times l)$ vector variate, X is an $(n \times m)$ matrix of known (non-stochastic) variables of rank $r \le m < n$, $\underline{\beta}$ is an $(m \times l)$ vector of unknown parameters and $\underline{E}$ is an $(n \times l)$ vector of stochastic error elements.

Let $b_1, b_2, \ldots, b_m$ be estimates of $\beta_1, \beta_2, \ldots, \beta_m$.

For given $\underline{y}$ and $\underline{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}$ form the *sum of squares*

$$\begin{aligned} q &= \sum_{i=1}^{n}(y_i - \sum_{j=1}^{m} x_{ij}\beta_j)^2 \\ &= (\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta}) \qquad (3.3.1) \\ &= \left\| \underline{y} - X\underline{\beta} \right\|^2 . \end{aligned}$$

This sum of squares can also be interpreted as $\sum_{i=1}^{n} e_i^2 = \underline{e}'\underline{e}$.

Definition 3.3.1 (Least squares estimates)

A set of functions of $\underline{y}$*, namely* $\hat{\beta}_1 = \hat{\beta}_1(\underline{y}), \cdots, \hat{\beta}_m = \hat{\beta}_m(\underline{y})$ *such that the values* $b_j = \hat{\beta}_j$ $(j = 1, 2, \ldots, m)$ *minimise* q *defined in (3.3.1) for a given value of* $\underline{y}$ *of the vector variate, is called a set of least squares estimates for the* $\{\beta_j\}$.

A necessary requirement for q to be at a minimum with respect to the variance in $\underline{\beta}$ follows as

$$\frac{\partial q}{\partial \underline{\beta}} = \underline{0}$$

$$\begin{aligned} \textit{i.e.}\ \underline{0} &= \frac{\partial}{\partial \underline{\beta}}\left((\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta})\right) \\ &= \frac{\partial}{\partial \underline{\beta}}\left(\underline{y}'\underline{y} - 2\underline{\beta}'X'\underline{y} + \underline{\beta}'X'X\underline{\beta}\right) \quad \text{(Why?)} \\ &= -2X'\underline{y} + 2X'X\underline{\beta}\,. \qquad (3.3.2) \end{aligned}$$

Now replace $\underline{\beta}$ in (3.3.2) with $\underline{b}$, then the *normal equations* follow as

$$X'X\underline{b} = X'\underline{y}\,. \qquad (3.3.3)$$

It will now be shown that q is indeed minimised if $\underline{b}$ is any solution of the normal equations, *i.e.* that a *sufficient condition* for q to be at a minimum with respect to the variance in $\underline{\beta}$ is that $\underline{b}$ satisfy the normal equations in (3.3.3):

$$
\begin{aligned}
q &= (\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta}) \\
&= (\underline{y} - X\underline{b} + X\underline{b} - X\underline{\beta})'(\underline{y} - X\underline{b} + X\underline{b} - X\underline{\beta}) \\
&= (\underline{y} - X\underline{b})'(\underline{y} - X\underline{b}) + (X\underline{b} - X\underline{\beta})'(X\underline{b} - X\underline{\beta}) + (\underline{y} - X\underline{b})'(X\underline{b} - X\underline{\beta}) + (X\underline{b} - X\underline{\beta})'(\underline{y} - X\underline{b}) \\
&= (\underline{y} - X\underline{b})'(\underline{y} - X\underline{b}) + (\underline{b} - \underline{\beta})'X'X(\underline{b} - \underline{\beta}),
\end{aligned}
$$

because

$$
\begin{aligned}
(\underline{y} - X\underline{b})'(X\underline{b} - X\underline{\beta}) &= (X\underline{b} - X\underline{\beta})'(\underline{y} - X\underline{b}) \\
&= \underline{b}'X'\underline{y} - \underline{b}'X'X\underline{b} - \underline{\beta}'X'\underline{y} + \underline{\beta}'X'X\underline{b} \\
&= \underline{b}'X'\underline{y} - \underline{b}'X'\underline{y} - \underline{\beta}'X'\underline{y} + \underline{\beta}'X'\underline{y} \quad (\underline{b} \text{ satisfies the norm. eqs.}) \\
&= 0.
\end{aligned}
$$

Because $(\underline{b} - \underline{\beta})'X'X(\underline{b} - \underline{\beta}) \geq 0$ it follows that $q \geq (\underline{y} - X\underline{b})'(\underline{y} - X\underline{b})$ and this minimum is attained when $\underline{\beta} = \underline{b}$.

In Theorem 1.11.2 it is also proved that the equations in (3.3.3) are consistent.

The following theorem has therefore been proved:

Theorem 3.3.1

Consider the model (3.2.8). *There always exists a set of least squares estimates for* $\underline{\beta}$ *and any set of least squares estimates satisfy the normal equations. Any solution of the normal equations is a set of least squares estimates.*

Also take note of the following geometric interpretation of the preceding theorem in terms of projections:

The equations in are consistent according to Theorem 1.11.2 and if $\underline{b}$ is any solution, then $X\underline{b}$ is uniquely determined. Further, it follows from Theorem 1.11.3 that $X\underline{b}$ is the projection of $\underline{y}$ onto the vector space generated by the columns of X, namely V(X).

It can now be shown that q is minimised if $\underline{\beta} = \underline{b}$ where $\underline{b}$ is such that $X\underline{b}$ is the projection of $\underline{y}$ onto the vector space generated by the columns of X. Note that $X\underline{\beta} \in$ of the vector space generated by the columns of X for all $\beta_1, \beta_2, \ldots, \beta_m$ and any vector in V(X) can be written in this form for some $\beta_1, \beta_2, \ldots, \beta_m$.

Let $X\underline{b}$ be the projection of $\underline{y}$ on V (X). Then

$$q = (\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta})$$

$$= \{(\underline{y} - X\underline{b}) + (X\underline{b} - X\underline{\beta})\}'\{(\underline{y} - X\underline{b}) + (X\underline{b} - X\underline{\beta})\}$$

$$= (\underline{y} - X\underline{b})'(\underline{y} - X\underline{b}) + (X\underline{b} - X\underline{\beta})'(X\underline{b} - X\underline{\beta})$$

because $(\underline{y} - X\underline{b})$ is orthogonal to V(X) and $(X\underline{b} - X\underline{\beta}) \in V(X)$

i.e. $(\underline{y} - X\underline{b})'(X\underline{b} - X\underline{\beta}) = 0.$

$\therefore\ q = (\underline{y} - X\underline{b})'(\underline{y} - X\underline{b}) + (\underline{b} - \underline{\beta})'X'X(\underline{b} - \underline{\beta}).$

Therefore, $q \geq (\underline{y} - X\underline{b})'(\underline{y} - X\underline{b})$ and the equality is true iff $\underline{\beta} = \underline{b}$. ♦

The minimum value of q is indicated by q_e which is called *the sum of squares for errors*. It follows that q_e *is the square of the length of the projection of* $\underline{y}$ *onto the vector space orthogonal to* V(X) and that q_e is *unique*, *i.e.*

$$\mathrm{Min}(q) \equiv q_e = (\underline{y} - X\underline{b})'(\underline{y} - X\underline{b})$$

$$= \left\| \underline{y} - X\underline{b} \right\|^2 . \qquad (3.3.4)$$

According to Theorem 3.3.1 a solution to the normal equations always exists. This solution is however not necessarily *unique*. Therefore, in the following, methods for obtaining *solutions to the normal equations* are investigated:

(a) The case where X′X is non-singular.

If $r = m$, *i.e.* the matrix X: $n \times m$ is of rank m, it follows that the $(m \times m)$ matrix $X'X$ is of rank m. Therefore, it follows that $X'X$ is non-singular and that $(X'X)^{-1}$ exists. It follows now that for the equation in (3.3.3) that

$$\underline{b} = (X'X)^{-1}X'\underline{y} \qquad (3.3.5)$$

where $\underline{b}$ is the least squares estimate of $\underline{\beta}$. In this case $\underline{b}$ is *unique*. From equation (3.3.5) it follows that $\underline{b}$ is a *linear function* of the observations $y_1, y_2, \ldots, y_n$.

It follows from (3.3.5) that the least squares *point estimators* $B_1, B_2, \ldots, B_m$ of $\beta_1, \beta_2, \ldots, \beta_m$ are given by

$$\underline{B} = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_m \end{bmatrix} = (X'X)^{-1}X'\underline{Y}. \tag{3.3.6}$$

The least squares point estimators $B_1, B_2, \ldots, B_m$ of $\beta_1, \beta_2, \ldots, \beta_m$ can therefore be derived by minimising

$$\begin{aligned} Q &= \sum_{i=1}^{n}(Y_i - \sum_{j=1}^{m} x_{ij}\beta_j)^2 \\ &= (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta}) \\ &= \left\|\underline{Y} - X\underline{\beta}\right\|^2 \end{aligned} \tag{3.3.7}$$

with respect to $\underline{\beta}$. This is done by first finding $\frac{\partial Q}{\partial \underline{\beta}}$, equating it to zero while at the same time replacing $\underline{\beta}$ with $\underline{B}$. In doing so, the normal equations

$$X'X\underline{B} = X'\underline{Y} \tag{3.3.8}$$

are obtained from which $\underline{B}$ can be determined.

Now it follows from (3.3.8) that

$$\begin{aligned} E(\underline{B}) &= (X'X)^{-1}X'E(\underline{Y}) \\ &= (X'X)^{-1}X'X\underline{\beta} \\ &= \underline{\beta} \end{aligned} \tag{3.3.9}$$

so that the least squares estimator $\underline{B}$ is an unbiased estimator of $\underline{\beta}$.

Further, it follows from (3.3.8) that $\text{cov}(\underline{B}; \underline{B}') = (X'X)^{-1}X'\Sigma_{\underline{Y}}\{(X'X)^{-1}X'\}' = (X'X)^{-1}\sigma^2$. (3.3.10)

(b) The case where X′X is singular.

If $X'X$ is singular, then a unique solution for the normal equations in (3.3.3) does not exist. By using a conditional inverse of the matrix $X'X$, the solution can formally be written down as

$$\underline{B} = (X'X)^c X'\underline{Y} \quad (\underline{b} = (X'X)^c X'\underline{y} \quad \text{if estimates are required})$$

by making use of Theorem 1.10.3, but this solution is not unique. In this case unique least squares estimators do not exist for every component of $\underline{\beta}$ and usually unbiased estimators for *linear* *functions of the components of* $\underline{\beta}$ are sought, *i.e.* $k_1\beta_1 + k_2\beta_2 + \cdots + k_m\beta_m$ where the $\{k_i\}$ are constants.

This problem is studied further in the following sections by investigating which linear functions of the parameters are *estimable.*

Remark 3.3.1

In some applications the linear model can take on the following form:

$$\underline{Y} = X\underline{\beta} + \underline{\eta} \tag{3.3.11}$$

where $E(\underline{\eta}) = \underline{0}$, but $\Sigma_{\underline{\eta}} = \sigma^2 T$ with T: $n \times n$ positive definite.

This model can be converted to the form in (3.2.8) by using an appropriate *transformation.*

Because T is positive definite, there exists a non-singular matrix C such that $CC' = T$.

Let $\underline{Z} = C^{-1}\underline{Y}$, then it follows that equation (3.3.11) can be written as

$$\begin{aligned} \underline{Z} &= C^{-1} X\underline{\beta} + C^{-1}\underline{\eta} \qquad (3.3.12) \\ &= C^{-1} X\underline{\beta} + \underline{E} \quad \text{(say)}. \end{aligned}$$

Now, $E(\underline{E}) = C^{-1}E(\underline{\eta}) = \underline{0}$

and
$$\begin{aligned} \Sigma_{\underline{E}} &= C^{-1}\,\Sigma_{\underline{\eta}}\,(C^{-1})' \\ &= C^{-1}T(C^{-1})'\sigma^2 \\ &= C^{-1}(CC')(C^{-1})'\sigma^2 \\ &= I_n\sigma^2 . \end{aligned}$$

The model (3.3.12) is therefore in the form (3.2.8).

3.4 ESTIMABLE LINEAR FUNCTIONS

Assume that the model $\underline{Y} = A\underline{\pi} + \underline{E}$ where $\underline{Y}$ is an $(n \times 1)$ vector variate of independent variables, A: $n \times m$ contains known (non-stochastic) variables, $\underline{\pi}$ is an $(m \times 1)$ column vector of non-stochastic parameters and $\underline{E}$ is an $(n \times 1)$ vector variate of stochastic error elements. It is further assumed that $E(\underline{E}) = \underline{0}$ and $\text{cov}(\underline{E}\,;\underline{E}') = I_n\sigma^2$. The model can also be written as

$$\begin{aligned} E(\underline{Y}) &= A\underline{\pi} \\ &= \underline{a}_{(1)}\pi_1 + \underline{a}_{(2)}\pi_2 + \ \ldots\ + \underline{a}_{(m)}\pi_m , \quad \text{because} \quad A = [\,\underline{a}_{(1)}, \underline{a}_{(2)}, \ \ldots\ , \underline{a}_{(m)}] \end{aligned}$$

or $\quad E(Y_i) = \sum_{j=1}^{m} a_{ij}\pi_j$ for i = 1, 2, . . . , n.

Further, assume that the matrix A is of rank $r \leq m$ where $m < n$. In all practical situations which will be considered, the number of observations will be more than the number of parameters which are to be estimated, *i.e.* $n > m$. If A is an $(n \times m)$ matrix and rank(A) = m, the column vectors of A are linearly independent and the estimation problem does not present any difficulties. However, if rank(A) < m, the column vectors of A are not linearly independent and it is said that the model is *not identifiable.* However, *restrictions* can be introduced that satisfy certain necessary and sufficient conditions for identifiability. As an alternative, it can be determined which linear functions of the parameters are *estimable.*

Now let $\underline{c}: n \times 1$ be any (unknown) n-component column vector of constants. Then the expected value of the linear function $\underline{c}'\underline{Y}$ can be written as

$$E(\underline{c}'\underline{Y}) = \underline{c}'A\underline{\pi}, \quad \text{because } E(\underline{Y}) = A\underline{\pi}.$$

Definition 3.4.1 (Unbiased linear estimator)

The function $\underline{c}'\underline{Y}$ *is an unbiased linear estimator (u.l.e.) of the linear function* $k_1\pi_1 + k_2\pi_2 + \ldots + k_m\pi_m$ *of the parameters* $\pi_1, \pi_2, \ldots, \pi_m$ *if* $E(\underline{c}'\underline{Y}) = \underline{k}'\underline{\pi}$ *is independent of the real values of the parameters* $\pi_1, \pi_2, \ldots, \pi_m$. ♦

A necessary and sufficient condition for $\underline{c}'\underline{Y}$ to be an u.l.e. of $\underline{k}'\underline{\pi}$ is that $\underline{c}'A = \underline{k}'$ (or $A'\underline{c} = \underline{k}$). Therefore, to find an u.l.e. of $\underline{k}'\underline{\pi}$, the linear equations $A'\underline{c} = \underline{k}$ must be solved for $\underline{c}$.

Definition 3.4.2 (Linear estimable function)

The linear function $\underline{k}'\underline{\pi}$ *of the parameters* $\underline{\pi}$ *is linearly estimable if a linear function* $\underline{c}'\underline{Y}$ *of the variates* $Y_1, Y_2, \ldots, Y_n$ *exists such that* $\underline{c}'\underline{Y}$ *is an u.l.e. of* $\underline{k}'\underline{\pi}$. *(Therefore,* $\underline{k}'\underline{\pi}$ *is linearly estimable if a linear function of* $Y_1, Y_2, \ldots, Y_n$ *exists and it is an u.l.e. of* $\underline{k}'\underline{\pi}$.*)*

Theorem 3.4.1

A necessary and sufficient condition for the linear function $\underline{k}'\underline{\pi}$ *of the parameters* $\underline{\pi}$ *to be linearly estimable, is that rank* (A) = *rank* $\begin{bmatrix} A \\ \underline{k}' \end{bmatrix}$ *where* $\begin{bmatrix} A \\ \underline{k}' \end{bmatrix}$ *is the matrix obtained by adjoining the row* *vector* $\underline{k}'$ *to the matrix* A. *This condition can also be written as rank*(A′) = *rank*[A′, $\underline{k}$] *or that* $\underline{k}'$ *is a linear combination of the rows of* A.

Proof

Let $\underline{k}'\underline{\pi}$ be an estimable function of the parameters $\underline{\pi}$.

Therefore, there exists a vector $\underline{c}$: n × 1 such that $E(\underline{c}'\underline{Y}) = \underline{k}'\underline{\pi}$

$$i.e.\quad \underline{c}'E(\underline{Y}) = \underline{k}'\underline{\pi}$$

$$i.e.\quad \underline{c}'A\underline{\pi} = \underline{k}'\underline{\pi}$$

and it is true for all values of $\underline{\pi}$ so that $\underline{c}'A = \underline{k}'$.

Therefore, $\underline{k}' = c_1\,\underline{a}_1' + c_2\,\underline{a}_2' + \ldots + c_n\,\underline{a}_n'$

and $\underline{k}'$ is therefore a linear combination of the rows of A: n × m.

Conversely:

Let rank $\begin{bmatrix} A \\ \underline{k}' \end{bmatrix}$ = rank (A).

Then, $\underline{k}' = \underline{c}'A$

so that $\underline{c}'A'\underline{\pi} = \underline{k}'\underline{\pi}$ for all $\underline{\pi}$: m × 1

i.e. $\underline{c}'E(\underline{Y}) = \underline{k}'\underline{\pi}$

and $E(\underline{c}'\underline{Y}) = \underline{k}'\underline{\pi}$.

Therefore, $\underline{k}'\underline{\pi}$ is estimable. ♦

Note that for any estimable function $\underline{k}'\underline{\pi}$ it is not any specific value of $\underline{c}'$ that is important, but it is the *existence* of $\underline{c}'$ that satisfies equation (3.4.1) below that is important.

$$\underline{k}' = \underline{c}'A = c_1\,\underline{a}_1' + c_2\,\underline{a}_2' + \ldots + c_n\,\underline{a}_n' . \qquad (3.4.1)$$

Certain important consequences of Theorem 3.4.1 are given as:

Remark 3.4.1

1. If rank(A) = m, then every linear function of the parameters are linearly estimable. This follows because it is true in general that rank(A′: m × n) ≤ rank{[A′, $\underline{k}$]: m × (n + 1) }. Therefore, for m < n and rank(A) = rank(A′) = m it follows that rank(A′: m × n) = rank{[A′, $\underline{k}$]: m × (n + 1) }. It is a sufficient condition for any linear function $\underline{k}'\underline{\pi}$ to be linearly estimable.

2. Every estimable linear function of the parameters must be of the form

$$\underline{d}'A\underline{\pi} = \underline{d}'E(\underline{Y}) = \sum_{i=1}^{n} d_i E(Y_i).$$

It follows because if $\underline{k}'\underline{\pi}$ is linearly estimable, there must exist a vector $\underline{d}$: n × 1 such that

$$E(\underline{d}'\underline{Y}) = \underline{k}'\underline{\pi}.$$

Therefore, $\underline{d}'A\underline{\pi} = \underline{k}'\underline{\pi}$

i.e. $\underline{k}'\underline{\pi}$ is of the form $\underline{d}'A\underline{\pi} = \underline{d}'E(\underline{Y}) = \sum_{i=1}^{n} d_i E(Y_i)$.

3. The expected value of any variate Y_i is estimable. It follows from Remark 3.4.1 (2) that by choosing $\underline{d}$ such that only the i^{th} component is equal to one and all the other elements are equal to zero.

Definition 3.4.3 (Error space)

Let V(A) *be as before the vector space generated by the columns of* A: n × m. *The vector space orthogonal to* V(A) *is defined as the error space and it is written as* V(E). (Note that E here is not the error term or the error vector $\underline{E}$ in the model $\underline{Y} = A\underline{\pi} + \underline{E}$, but it is a **symbolic representation** for the orthogonal complement of V(A), namely a symbolic representation of $V^{\perp}(A)$.

Remark 3.4.2

A linear function of the variates $Y_1, Y_2, \ldots, Y_n$, say $\underline{e}'\underline{Y} = e_1Y_1 + e_2Y_2 + \ldots + e_nY_n$, belongs to the error set (shortly: belongs to errors) if the coefficient vector $\underline{e}$ is an element of V(E).

Theorem 3.4.2

A linear function, say $\underline{e}'\underline{Y}$ *of the variates* $Y_1, Y_2, \ldots, Y_n$, *belongs to errors if and only if its expected value is zero, irrespective of the real values of the parameters* $\underline{\pi}$.

Proof

Let the linear function $\underline{e}'\underline{Y}$ belong to errors.

Then $E(\underline{e}'\underline{Y}) = \underline{e}'E(\underline{Y})$

$= \underline{e}'A\underline{\pi}$

$= 0$, because $\underline{e}$ is orthogonal to V(A).

Conversely:

If $E(\underline{e}'\underline{Y}) = \underline{e}'A\underline{\pi} = 0$ irresepective of the values of $\{\pi_1, \pi_2, \ldots, \pi_m\}$,

then $\underline{e}'A = \underline{0}'$

i.e. $\underline{e}$ is orthogonal to V(A) and therefore $\underline{e} \in V(E)$.

Definition 3.4.4 (Linear error set)

The set of all linear function of the variates $Y_1, Y_2, \ldots, Y_n$ *which belong to errors, is called the linear error set,* $L(E'\underline{Y})$, *where* E *is a matrix with (n-component) column vectors which generates the vector space* V(E), *i.e. the columns of the matrix* E *are orthogonal to* V(A).

Remark 3.4.3

Because rank(A) = r, the dimension of V(A) is equal to r and the dimension of V(E) is equal to $n - r = n_e$, say. Therefore, it follows that the number of *degrees of freedom* of $L(E'\underline{Y})$ is equal to $n_e = n - r$.

3.5 THE FUNDAMENTAL THEOREM OF LINEAR ESTIMATION AND THE GAUSS-MARKOV THEOREM

In this section it will be shown that for every estimable function of the parameters a unique unbiased linear estimable function exists that can be determined with the help of the normal equations and the least squares estimates of §3.3.

Theorem 3.5.1 (The fundamental theorem)

If $\underline{k}'\underline{\pi}$ *is any* ***estimable*** *linear function of the parameters* $\pi_1, \pi_2, \ldots, \pi_m$, *then*

(i) *there exists a unique linear function* $\underline{c}'\underline{Y}$ *of the variates* $Y_1, Y_2, \ldots, Y_n$ *such that* $\underline{c}$ *belongs to the column space* V(A) *and* $\underline{c}'\underline{Y}$ *is an u.l.e. of* $\underline{k}'\underline{\pi}$;

(ii) $\text{var}(\underline{c}'\underline{Y}) <$ *variance of any other u.l.e. of* $\underline{k}'\underline{\pi}$.

Proof

(i) Because $\underline{k}'\underline{\pi}$ is linearly estimable, there exists an u.l.e., say $\underline{d}'\underline{Y}$, such that $E(\underline{d}'\underline{Y}) = \underline{k}'\underline{\pi}$. Now, $\underline{d}$: $n \times 1$ can be uniquely written as, say $\underline{d} = \underline{c} + \underline{e}$ where $\underline{e}$ is orthogonal to $\underline{c}$ and $\underline{c}$ is not the zero vector with $\underline{c} \in V(A)$ (compare Theorems 1.7.2 and 1.7.5). Therefore, $\underline{d}'\underline{Y} = \underline{c}'\underline{Y} + \underline{e}'\underline{Y}$ where $\underline{c} \in V(A)$ and $\underline{e}$ belongs to the error vector space V(E) which is orthogonal to V(A). Therefore, it is true that

$E(\underline{d}'\underline{Y}) = E(\underline{c}'\underline{Y}) + 0$

or $E(\underline{c}'\underline{Y}) = \underline{k}'\underline{\pi}$ so that $\underline{c}'\underline{Y}$ is also an u.l.e. of $\underline{k}'\underline{\pi}$.

The *uniqueness* of $\underline{c}'\underline{Y}$ can be demonstrated as follows:

Assume that there exists another vector $\underline{c}_0$ which belongs to V(A) such that

$E(\underline{c}_0'\underline{Y}) = \underline{k}'\underline{\pi}$.

Let $\underline{c}_1 = \underline{c} - \underline{c}_0$, so that $\underline{c}_1$ also belongs to V(A). Then,

$$\begin{aligned} E(\underline{c}_1'\underline{Y}) &= E(\underline{c}'\underline{Y}) - E(\underline{c}_0'\underline{Y}) \\ &= \underline{k}'\underline{\pi} - \underline{k}'\underline{\pi} \\ &= 0 \end{aligned}$$

and $\underline{c}_1$ therefore belongs to V(E). This is not possible unless $\underline{c}_1 = \underline{0}$ and then it follows that $\underline{c} = \underline{c}_0$. Therefore it follows that $\underline{c}$ that belongs to V(A) such that $E(\underline{c}'\underline{Y}) = \underline{k}'\underline{\pi}$, is unique.

(ii) Assume, as in (i), that $\underline{d}'\underline{Y}$ is an arbitrary u.l.e. of $\underline{k}'\underline{\pi}$ and that

$$\begin{aligned} \text{var}(\underline{d}'\underline{Y}) = \underline{d}'\underline{d}\sigma^2 &= (\underline{c} + \underline{e})'(\underline{c} + \underline{e})\sigma^2 \\ &= \underline{c}'\underline{c}\sigma^2 + \underline{e}'\underline{e}\sigma^2 \quad \text{because } \underline{e}'\underline{c} = \underline{c}'\underline{e} = 0 \text{ (Why?)} \\ &= \text{var}(\underline{c}'\underline{Y}) + \text{var}(\underline{e}'\underline{Y}) \\ &\geq \text{var}(\underline{c}'\underline{Y}). \end{aligned}$$

The equality is only true if $\text{var}(\underline{e}'\underline{Y}) = 0$ for all $\underline{Y}$, i.e. if $\underline{e}'\underline{e} = \sum_{i=1}^{n} e_i^2 = 0$, i.e. if $\underline{e}$ is the zero vector. In the latter case, $\underline{d}'\underline{Y} = \underline{c}'\underline{Y}$.

Definition 3.5.1 (Best estimator)

The estimator $\underline{c}'\underline{Y}$ *which satisfies all the conditions of Theorem 3.5.1 is called a best estimator. This estimator is denoted by b.l.u.e.*

Because Theorem 3.5.1 alleges that for every *estimable* function $\underline{k}'\underline{\pi}$ a best estimator $\underline{c}'\underline{Y}$ exists with $\underline{c} \in$ V(A) it follows:

Definition 3.5.2 (Estimation space and estimation set)

The vector space V(A) *is called the (linear) estimation space and the linear set* L(A′$\underline{Y}$) *the (linear) estimation set where* L(A′$\underline{Y}$) *is defined as the linear set generated by the set of linear functions* $\{\underline{a}_{(1)}'\underline{Y}, \underline{a}_{(2)}'\underline{Y}, \ldots, \underline{a}_{(m)}'\underline{Y}\}$ where $[\underline{a}_{(1)}, \underline{a}_{(2)}, \ldots, \underline{a}_{(m)}] = A$.

Remark 3.5.1

It follows from Theorem 3.5.1 as well as Definitions 3.5.1 and 3.5.2 that the best estimator for any linear estimable function of the parameters belongs to the estimation set $L(A'\underline{Y})$.

The matrix $A : n \times m = [\underline{a}_{(1)}, \underline{a}_{(2)}, \cdots, \underline{a}_{(m)}]$

implies that $A'\underline{Y} = \begin{bmatrix} \underline{a}'_{(1)}\underline{Y} \\ \underline{a}'_{(2)}\underline{Y} \\ \cdots \\ \underline{a}'_{(m)}\underline{Y} \end{bmatrix}$.

Further, because any best estimator is an element of $L(A'\underline{Y})$, it follows that any best estimator is of the form $p_1\underline{a}'_{(1)}\underline{Y} + p_2\underline{a}'_{(2)}\underline{Y} + \cdots + p_m\underline{a}'_{(m)}\underline{Y} = \underline{p}'A'\underline{Y}$.

Further, the uniqueness of the best estimator $\underline{c}'\underline{Y}$ with $\underline{c} \in V(A)$ ensures that

$$\text{Best estimator} = \underline{c}'\underline{Y} = \underline{p}'A'\underline{Y}. \qquad (3.5.1)$$

The following theorem can now be easily proved:

Theorem 3.5.2

If $\underline{k}'\underline{\pi}$ is linearly estimable, then its best estimator is $\underline{p}'A'\underline{Y}$ where the m-component vector $\underline{p}$ satisfies the equations $\underline{p}'A'A = \underline{k}'$.

Proof

It follows from (3.5.1) that the best estimator of estimable $\underline{k}'\underline{\pi}$ must be of the form $\underline{p}'A'\underline{Y}$ and because it is also an unbiased estimator, $E(\underline{p}'A'\underline{Y}) = \underline{k}'\underline{\pi}$.

Therefore, $\underline{p}'A'E(\underline{Y}) = \underline{p}'A'A\underline{\pi} = \underline{k}'\underline{\pi}$, but because this result is true for all $\underline{\pi}$: $m \times 1$, $\underline{p}'A'A = \underline{k}'$.

Theorem 3.5.3

If $\underline{k}'\underline{\pi}$ is linearly estimable, its best estimator is $\underline{k}'\hat{\underline{\Pi}}$, where $\hat{\underline{\Pi}}$ is any solution of the equations $A'A\underline{\pi} = A'\underline{Y}$ is.

Proof

It follows from Theorem 3.5.2 that the best estimator of $\underline{k}'\underline{\pi}$ is given by $\underline{p}'A'\underline{Y}$ where the m-component vector $\underline{p}$ satisfies the equations $\underline{p}'A'A = \underline{k}'$. If $\underline{\hat{\Pi}}$ is any solution of the equations $A'A\underline{\pi} = A'\underline{Y}$, then $A'A\underline{\hat{\Pi}} = A'\underline{Y}$ so that $\underline{p}'A'\underline{Y} = \underline{p}'A'A\underline{\hat{\Pi}}$. Because $\underline{p}'A'A = \underline{k}'$, it follows that the best estimator of estimable $\underline{k}'\underline{\pi}$ is $\underline{k}'\underline{\hat{\Pi}}$.

Remark 3.5.2

1. A solution $\underline{\hat{\Pi}}$ which satisfies $A'A\underline{\hat{\Pi}} = A'\underline{Y}$, can be found as follows: If $(A'A)$ is non-singular, the solution of the normal equations is $\underline{\hat{\Pi}} = (A'A)^{-1}A'\underline{Y}$.

   If $(A'A)$ is singular, a solution is given by $\underline{\hat{\Pi}} = (A'A)^c A'\underline{Y}$ where $(A'A)^c$ is any conditional inverse of $(A'A)$. As has already been proved, such a solution always exists.

2. The best estimator $\underline{k}'\underline{\hat{\Pi}}$ of $\underline{k}'\underline{\pi}$ is unique. According to Theorem 3.5.1 there exists a unique best estimator of $\underline{k}'\underline{\pi}$, and in agreement with Theorem 3.5.2 this estimator is of the form $\underline{p}'A'\underline{Y}$ where $\underline{p}'A'A = \underline{k}'$. It follows that $\underline{k}'\underline{\hat{\Pi}} = \underline{p}'A'\underline{Y}$ is a unique estimator.

Theorem 3.5.4 (The Gauss-Markov Theorem)

The best estimator of estimable $\underline{k}'\underline{\pi}$ *is* $\underline{k}'\underline{\hat{\Pi}}$ *where* $\hat{\Pi}_1, \hat{\Pi}_2, \cdots, \hat{\Pi}_m$ *are the least squares estimators of* $\pi_1, \pi_2, \ldots, \pi_m$. *(I.e. the least squares estimator of* $\underline{\pi}$ *leads to the best estimator of estimable* $\underline{k}'\underline{\pi}$ *.)*

Proof

The least squares estimators are obtained by making

$$Q = \{\underline{Y} - E(\underline{Y})\}'\{\underline{Y} - E(\underline{Y})\}$$
$$= (\underline{Y}' - \underline{\pi}'A')(\underline{Y} - A\underline{\pi}) \text{ a minimum with respect to } \underline{\pi}.$$

Now it follows from $\dfrac{\partial Q}{\partial \underline{\pi}} = \underline{0}$ and replacing $\underline{\pi}$ with $\underline{\hat{\Pi}}$ that $A'A\underline{\hat{\Pi}} = A'\underline{Y}$, the so-called normal equations. Just as in Theorem 3.3.1 it can now be shown that the solution $\underline{\hat{\Pi}}$ of $A'A\underline{\hat{\Pi}} = A'\underline{Y}$ will give a minimum value for Q. But, $A'A\underline{\hat{\Pi}} = A'\underline{Y}$, the normal equations, is the system of equations of Theorem 3.5.3 and the solution thereof gives the best estimator $\underline{k}'\underline{\hat{\Pi}}$ of estimable $\underline{k}'\underline{\pi}$. ♦

The following theorem considers the case of a *set of estimable functions* and alleges *inter alia* that any linear combination of estimable functions is also estimable.

Theorem 3.5.5

Let $K: m \times t = [\underline{k}_{(1)}, \underline{k}_{(2)}, \ . \ . \ . \ , \underline{k}_{(t)}]$ *and* $C: n \times t = [\underline{c}_{(1)}, \underline{c}_{(2)}, \ . \ . \ . \ , \underline{c}_{(t)}]$.

If the elements (rows) of

$$K'\underline{\pi} = \begin{bmatrix} \underline{k}'_{(1)} \\ \underline{k}'_{(2)} \\ \vdots \\ \underline{k}'_{(t)} \end{bmatrix} \underline{\pi} = \begin{bmatrix} \underline{k}'_{(1)}\underline{\pi} \\ \underline{k}'_{(2)}\underline{\pi} \\ \vdots \\ \underline{k}'_{(t)}\underline{\pi} \end{bmatrix} \tag{3.5.2}$$

are estimable functions, then any linear combination of the rows of $K'\underline{\pi}$ *, say* $\underline{h}'K'\underline{\pi} = \sum_{i=1}^{t} h_i \underline{k}'_{(i)}\underline{\pi}$ *is also an estimable function.*

If the best estimators of the rows of $K'\underline{\pi}$ *are given by the rows of*

$$C'\underline{Y} = \begin{bmatrix} \underline{c}'_{(1)}\underline{Y} \\ \underline{c}'_{(2)}\underline{Y} \\ \vdots \\ \underline{c}'_{(t)}\underline{Y} \end{bmatrix}, \tag{3.5.3}$$

then $\underline{h}'C'\underline{Y} = \sum_{i=1}^{t} h_i \underline{c}'_{(i)}\underline{Y}$ *is the best estimator of* $\underline{h}'K'\underline{\pi} = \sum_{i=1}^{t} h_i \underline{k}'_{(i)}\underline{\pi}$.

Further, it follows that rank(C) = t *if and only if rank*(K) = t, *i.e. the best estimators of the estimable functions* $\underline{k}'_{(1)}\underline{\pi}, \underline{k}'_{(2)}\underline{\pi}, \cdots, \underline{k}'_{(t)}\underline{\pi}$ *are linearly independent if and only if the estimable functions are linearly independent.*

Proof

Because $E(\underline{h}'C'\underline{Y}) = \underline{h}'K'\underline{\pi}$, $\underline{h}'K'\underline{\pi}$ is estimable because $\underline{h}'C'\underline{Y}$ is an unbiased estimator. But, $(\underline{h}'C')' = \sum_{i=1}^{t} h_i \underline{c}_{(i)}$ is an element of V(A) because $\underline{c}_{(i)} \in V(A)$, $i = 1, 2, \cdots, t$ since $\underline{c}'_{(i)}\underline{Y}$ is the best estimator of $\underline{k}'_{(i)}\underline{\pi}$. Therefore, $\underline{h}'C'\underline{Y}$ is the best estimator of $\underline{h}'K'\underline{\pi}$.

Further, it follows that if the rows of $K'\underline{\pi}$ are linearly dependent, then, for example, $\underline{k}'_{(t)}\underline{\pi}$ can be expressed as

$$\underline{k}'_{(t)}\underline{\pi} = \sum_{i=1}^{t-1} b_i \underline{k}'_{(i)}\underline{\pi} \quad \text{for one or another } \{b_1, b_2, \ . \ . \ . \ , b_{t-1}\}.$$

Therefore, the best estimator of $\underline{k}'_{(t)}\underline{\pi}$ is according to the first part of the proof

$$\underline{c}'_{(t)}\underline{Y} = \sum_{i=1}^{t-1} b_i \underline{c}'_{(i)}\underline{Y}$$

from which it follows that the rows of $C'\underline{Y}$ are also linearly dependent.

Conversely: If the rows of $C'\underline{Y}$ are linearly dependent, then, for example,

$$\underline{c}'_{(t)}\underline{Y} = \sum_{i=1}^{t-1} b_i \underline{c}'_{(i)}\underline{Y} \quad \text{for one or another } \{b_1, b_2, \ \ldots\ , b_{t-1}\}.$$

By taking expected values it follows that

$$E(\underline{c}'_{(t)}\underline{Y}) = \sum_{i=1}^{t-1} b_i E(\underline{c}'_{(i)}\underline{Y}).$$

Therefore, $\underline{k}'_{(t)}\underline{\pi} = \sum_{i=1}^{t-1} b_i \underline{k}'_{(i)}\underline{\pi}$ from which it follows that the rows of $K'\underline{\pi}$ are also linearly dependent.

But linear functions are linearly dependent if and only if their coefficient vectors are linearly dependent.

Therefore: $\{\underline{k}_{(1)}, \underline{k}_{(2)}, \ \ldots\ , \underline{k}_{(t)}\}$ is linearly dependent if and only if $\{\underline{c}_{(1)}, \underline{c}_{(2)}, \ \ldots\ , \underline{c}_{(t)}\}$ is linearly dependent, so that $\{\underline{k}_{(1)}, \underline{k}_{(2)}, \ \ldots\ , \underline{k}_{(t)}\}$ is linearly independent if and only if $\{\underline{c}_{(1)}, \underline{c}_{(2)}, \ \ldots\ , \underline{c}_{(t)}\}$ is linearly independent, i.e. rank(K) = t iff rank(C) = t.

3.6 VARIANCES AND COVARIANCES OF BEST ESTIMATORS

If $\underline{k}'\underline{\pi}$ is linearly estimable, its best estimator is given by $\underline{k}'\hat{\underline{\Pi}} = \underline{p}'A'\underline{Y}$ where $\underline{p}'A'A = \underline{k}'$. Now it follows that

$$\begin{aligned}
\text{var(best estimator)} &= \text{var}(\underline{k}'\hat{\underline{\Pi}}) \\
&= \underline{p}'A' \,\text{cov}(\underline{Y};\underline{Y}')\, A\,\underline{p} \\
&= \underline{p}'A'A\,\underline{p}\ \sigma^2 \\
&= \underline{k}'\underline{p}\ \sigma^2. \qquad (3.6.1)
\end{aligned}$$

Further, the best estimator is also given by $\underline{k}'\hat{\underline{\Pi}}$, where $A'A\hat{\underline{\Pi}} = A'\underline{Y}$ and $\hat{\underline{\Pi}} = (A'A)^c A'\underline{Y}$ is a solution of the normal equations. Therefore, an alternative form of the variance of a best estimator is:

$$\begin{aligned}
\text{var(best estimator)} &= \text{var}(\underline{k}'\hat{\underline{\Pi}}) \\
&= \underline{k}' \text{cov}\{(A'A)^c A'\underline{Y} \,;\, ((A'A)^c A'\underline{Y})'\}\underline{k} \\
&= \underline{k}'(A'A)^c A'A\{(A'A)^c\}'\underline{k}\sigma^2 \\
&= \underline{d}'A(A'A)^c A'A\{(A'A)^c\}'A'\underline{d}\sigma^2 \quad \text{(see Theorem 3.4.1)} \\
&= \underline{d}'A(A'A)^c A'A\,(A'A)^c A'\underline{d}\sigma^2 \quad (\textit{Why?}) \\
&= \underline{d}'A(A'A)^c A'\underline{d}\sigma^2 \quad (\textit{Why?}) \\
&= \underline{k}'(A'A)^c \underline{k}\sigma^2 . \qquad (3.6.2)
\end{aligned}$$

Because $\underline{k}'(A'A)^c \underline{k} = \underline{d}'A(A'A)^c A'\underline{d}$, the variance is unique, *i.e.* independent of the choice of the conditional inverse of $A'A$.

The best estimator of linearly estimable $\underline{k}'\underline{\pi}$ is also given by $\underline{c}'\underline{Y}$ with $\underline{c} \in V(A)$ so that a third expression for var(best estimator of $\underline{k}'\underline{\pi}$) can be written as $\text{var}(\underline{c}'\underline{Y}) = \underline{c}'\underline{c}\sigma^2$.

Now consider the t estimable functions which are the elements of the vector

$$K'\underline{\pi} = \begin{bmatrix} \underline{k}'_{(1)} \\ \underline{k}'_{(2)} \\ \vdots \\ \underline{k}'_{(t)} \end{bmatrix} \underline{\pi} = \begin{bmatrix} \underline{k}'_{(1)}\underline{\pi} \\ \underline{k}'_{(2)}\underline{\pi} \\ \vdots \\ \underline{k}'_{(t)}\underline{\pi} \end{bmatrix}$$

where $K: m \times t = [\underline{k}_{(1)}, \underline{k}_{(2)}, \ . \ . \ . \ , \underline{k}_{(t)}]$:

According to Theorem 3.4.1, $\underline{k}'_{(j)} = \underline{d}'_{(j)} A$ for one or another $\underline{d}'_{(j)}$, $j = 1, 2, \cdots, t$.

Therefore, $K' = \begin{bmatrix} \underline{d}'_{(1)} \\ \underline{d}'_{(2)} \\ \vdots \\ \underline{d}'_{(t)} \end{bmatrix} A = D'A$ where $D : n \times t = \lfloor \underline{d}_{(1)}, \underline{d}_{(2)}, \cdots, \underline{d}_{(t)} \rfloor$

The best estimators of the elements of $K'\underline{\pi}$ are the elements of $K'\hat{\underline{\Pi}}$ and now it is true that

$$\begin{aligned}
\text{Cov}[(K'\hat{\underline{\Pi}}) \,;\, (K'\hat{\underline{\Pi}})'] &= \text{cov}\{K'(A'A)^c A'\underline{Y} \,;\, (K'(A'A)^c A'\underline{Y})'\} \\
&= D'A(A'A)^c A'A\{(A'A)^c\}'A'D\sigma^2 \quad (\textit{Motivate in detail}) \\
&= D'A(A'A)^c A'A\,(A'A)^c A'D\sigma^2 \quad (\textit{Why?}) \\
&= D'A(A'A)^c A'D\sigma^2 \quad (\textit{Why?}) \\
&= K'(A'A)^c K\sigma^2 \qquad (3.6.3)
\end{aligned}$$

$$= \begin{bmatrix} \underline{k}'_{(1)} \\ \underline{k}'_{(2)} \\ \vdots \\ \underline{k}'_{(t)} \end{bmatrix} (A'A)^c \left[\underline{k}_{(1)}, \underline{k}_{(2)}, \cdots, \underline{k}_{(t)} \right] \sigma^2$$

$$= \begin{bmatrix} \underline{k}'_{(1)} (A'A)^c \\ \underline{k}'_{(2)} (A'A)^c \\ \vdots \\ \underline{k}'_{(t)} (A'A)^c \end{bmatrix} \left[\underline{k}_{(1)}, \underline{k}_{(2)}, \cdots, \underline{k}_{(t)} \right] \sigma^2$$

$$= \begin{bmatrix} \underline{k}'_{(1)} (A'A)^c \underline{k}_{(1)} & \underline{k}'_{(1)} (A'A)^c \underline{k}_{(2)} & \cdots & \underline{k}'_{(1)} (A'A)^c \underline{k}_{(t)} \\ \underline{k}'_{(2)} (A'A)^c \underline{k}_{(1)} & \underline{k}'_{(2)} (A'A)^c \underline{k}_{(2)} & \cdots & \underline{k}'_{(2)} (A'A)^c \underline{k}_{(t)} \\ \cdots & \cdots & \cdots & \cdots \\ \underline{k}'_{(t)} (A'A)^c \underline{k}_{(1)} & \underline{k}'_{(t)} (A'A)^c \underline{k}_{(2)} & \cdots & \underline{k}'_{(t)} (A'A)^c \underline{k}_{(t)} \end{bmatrix} \sigma^2.$$

It follows from the preceding symmetric matrix that $\mathrm{cov}\left(\underline{k}'_{(i)} \underline{\hat{\Pi}}; \underline{k}'_{(j)} \underline{\hat{\Pi}}\right) = \underline{k}'_{(i)} (A'A)^c \underline{k}_{(j)} \sigma^2$.

Example 3.6.1

Consider the model $\underline{Y} = X\underline{\beta} + \underline{E}$ with $E(\underline{E}) = \underline{0}$, $E(\underline{E}\,\underline{E}') = I_n\sigma^2$ and where X: $n \times m$ has a rank of m with $m < n$. All linear functions of $\underline{\beta}$ are therefore linearly estimable. The normal equations are $X'X\underline{b} = X'\underline{y}$ and $\underline{b} = (X'X)^{-1} X'\underline{y}$ where $\underline{b}$ is the (unique) least squares estimate as well as the best estimate of $\underline{\beta}$. Then, $\mathrm{var}(\underline{k}'\underline{B}) = \underline{k}'(X'X)^{-1}\underline{k}\,\sigma^2$ where $\underline{B}$ is the least squares estimator as well as the best estimator of $\underline{\beta}$.

In the case of one independent variable the regression equation $\underline{Y} = X\underline{\beta} + \underline{E}$ is often written in the following form:

$$Y_i = \beta_0 + \beta_1 (x_i - \bar{x}) + E_i, \text{ for } i = 1, 2, \cdots, n.$$

Then, $X = \begin{bmatrix} 1 & x_1 - \bar{x} \\ 1 & x_2 - \bar{x} \\ \vdots & \vdots \\ 1 & x_n - \bar{x} \end{bmatrix}$, $\underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$, $X'X = \begin{bmatrix} n & 0 \\ 0 & \sum_i (x_i - \bar{x})^2 \end{bmatrix}$,

$$(X'X)^{-1} = \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{\sum_i (x_i - \bar{x})^2} \end{bmatrix} \text{ and } X'\underline{y} = \begin{bmatrix} \sum_i y_i \\ \sum_i y_i (x_i - \bar{x}) \end{bmatrix}$$

so that $\underline{b} = (X'X)^{-1} X'\underline{y} = \begin{bmatrix} \frac{1}{n}\sum_i y_i \\ \frac{\sum_i y_i (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} \end{bmatrix}$.

Therefore, $b_0 = \bar{y}$ and $b_1 = \dfrac{\sum_i y_i (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}$ = regression coefficient of y on x. Here, b_0 and b_1 are the best estimates (b.l.u.e.s) of β_0 and β_1 while $B_0 = \bar{Y}$ and $B_1 = \dfrac{\sum_i Y_i (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}$ are the best estimators of β_0 and β_1.

$$\text{Var}(B_0) = \text{var}\{(1, 0)(B_0, B_1)'\} = (1, 0)\begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{\sum_i (x_i - \bar{x})^2} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix}\sigma^2 = \sigma^2/n.$$

Similarly, it can be derived that $\text{var}(B_1) = \dfrac{\sigma^2}{\sum_i (x_i - \bar{x})^2}$ and

$$\text{var}\{B_0 + B_1(x - \bar{x})\} = \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} \right)\sigma^2.$$

3.7 <u>SUMS OF SQUARES</u>

Consider again the linear model as given in equation (3.2.8). For every linear function $\underline{c}'\underline{Y}$ there exists a quantity called the sum of squares.

Definition 3.7.1 (Sum of squares)

If $\underline{c}'\underline{Y} = c_1 Y_1 + c_2 Y_2 + \ldots + c_n Y_n$ *is a linear function of the variates* $Y_1, Y_2, \ldots, Y_n$, *then the sum of squares of* $\underline{c}'\underline{Y}$ *is defined as*

$$Q = \frac{(\underline{c}'\underline{Y})^2}{\underline{c}'\underline{c}} = \frac{(c_1 Y_1 + \cdots + c_n Y_n)^2}{c_1^2 + \cdots + c_n^2} \ .$$

The quantity Q is the sum of squares with one *degree of freedom* associated with $\underline{c}'\underline{Y}$.

Example 3.7.1

The sum of squares of $\overline{Y} = \frac{1}{n}\sum_i Y_i$ is $n(\overline{Y})^2$. By letting $c_i = 1/n$ for $i = 1, 2, \cdots, n$ it follows that

$$Q_{\overline{Y}} = \frac{(\frac{1}{n})^2 (Y_1 + \cdots + Y_n)^2}{(\frac{1}{n})^2 + \cdots + (\frac{1}{n})^2} = (\tfrac{1}{n})^2 (n\overline{Y})^2 / \tfrac{1}{n} = n\overline{Y}^2 \,.$$

Corollary 3.7.1

1. *The sum of squares of* $\underline{c}'\underline{Y}$ *is the same as the sum of squares of* $k\underline{c}'\underline{Y}$ *where* k *is a non-zero scalar.*

2. $E(Q) = \sigma^2 + Q_m$ *where* Q_m *is obtained from* Q *by replacing the variates with their expected values. This can be shown as follows:*

   Because $\mathrm{var}(\underline{c}'\underline{Y}) = E\{(\underline{c}'\underline{Y})^2\} - \{E(\underline{c}'\underline{Y})\}^2$

   $$\textit{it follows that}\ \ E\left(\frac{(\underline{c}'\underline{Y})^2}{\underline{c}'\underline{c}}\right) = \frac{\mathrm{var}(\underline{c}'\underline{Y})}{\underline{c}'\underline{c}} + \frac{\{E(\underline{c}'\underline{Y})\}^2}{\underline{c}'\underline{c}}$$

   $$\textit{i.e.}\ \ E(Q) = \frac{\underline{c}'\underline{c}\sigma^2}{\underline{c}'\underline{c}} + \frac{\{\underline{c}'E(\underline{Y})\}^2}{\underline{c}'\underline{c}}$$

   $$= \sigma^2 + Q_m \,.$$

3. *If* $\underline{c}'\underline{Y}$ *belongs to the error set, i.e.* $E(\underline{c}'\underline{Y}) = 0$, *it follows that* $E(Q) = \sigma^2$. *Therefore, the expected value of the sum of squares of a linear function which belongs to the error set and which has a degree of freedom of one is equal to* σ^2 .

4. *In Theorem 1.7.1 it is proved that a vector* $\underline{Y}$ *can be written in the form:* $\underline{Y} = \left(\frac{\underline{c}'\underline{Y}}{\underline{c}'\underline{c}}\right)\underline{c} + \{\underline{Y} - \left(\frac{\underline{c}'\underline{Y}}{\underline{c}'\underline{c}}\right)\underline{c}\}$ *where* $\left(\frac{\underline{c}'\underline{Y}}{\underline{c}'\underline{c}}\right)\underline{c}$ *is the projection of* $\underline{Y}$ *on* $\underline{c}$. *Now write* $\underline{Z}$ *for the projection of the vector variate* $\underline{Y}$ *onto the coefficient vector* $\underline{c}$ *of the linear function* $\underline{c}'\underline{Y}$. *Now it follows that:*

   $$\underline{Z}' = \left(\frac{\underline{c}'\underline{Y}}{\underline{c}'\underline{c}}\right)\underline{c}' \ \textit{and}\ \underline{Z}'\underline{Z} = \frac{(\underline{c}'\underline{Y})^2}{\underline{c}'\underline{c}} = Q \,.$$

Further, $\dfrac{\underline{c}'\underline{Y}}{\sqrt{\underline{c}'\underline{c}}}$ *is the length of the projection* $\underline{Z}$ *so that* $Q = \dfrac{(\underline{c}'\underline{Y})^2}{\underline{c}'\underline{c}} = \underline{Z}'\,\underline{Z}$ = *Square of the length of the projection of* $\underline{Y}$ *onto the coefficient vector* $\underline{c}$.

Therefore, the sum of squares of the linear function $\underline{c}'\underline{Y}$ *is the* ***square of the length of the projection of*** $\underline{\mathbf{Y}}$ ***onto*** $\underline{\mathbf{c}}$.

Definition 3.7.2

Let $C: n \times t = [\underline{c}_{(1)}, \underline{c}_{(2)}, \ldots, \underline{c}_{(t)}]$ *and consider the linear set generated by the rows of*

$$C'\underline{Y} = \begin{bmatrix} \underline{c}'_{(1)}\underline{Y} \\ \underline{c}'_{(2)}\underline{Y} \\ \vdots \\ \underline{c}'_{(t)}\underline{Y} \end{bmatrix}.$$

Now, let this linear set be represented by $L(C'\underline{Y})$. *The sum of squares of the linear set* $L(C'\underline{Y})$ *is defined as the square of the length of the projection of* $\underline{Y}$ *onto the vector space generated by the columns of* C . ♦

The sum of squares of the linear set $L(C'\underline{Y})$ can be simplified as follows:

$$\begin{aligned} Q &= \{C(C'C)^c C'\underline{Y}\}'\{C(C'C)^c C'\underline{Y}\} \text{ (compare with Theorem 1.11.4)} \\ &= \underline{Y}'C\{(C'C)^c\}'C'C(C'C)^c C'\underline{Y} \\ &= \underline{Y}'C(C'C)^c C'\underline{Y} \qquad (3.7.1) \end{aligned}$$

because $C(C'C)^c C'$ is unique, symmetric and idempotent (Theorem 1.11.5).

This sum of squares is a quadratic function of the variates $Y_1, Y_2, \ldots, Y_n$. Because the matrix $C(C'C)^c C$ is independent of the specific column vectors $\underline{c}_{(1)}, \underline{c}_{(2)}, \ldots, \underline{c}_{(t)}$ of C which generate V(C) (Theorem 1.11.6), it follows that this sum of squares is independent of the specific linear functions $\underline{c}'_{(1)}\underline{Y}, \underline{c}'_{(2)}\underline{Y}, \ldots, \underline{c}'_{(t)}\underline{Y}$ which generate $L(C'\underline{Y})$. The number of degrees of freedom associated with Q, is the number of degrees of freedom of the linear set $L(C'\underline{Y})$.

Theorem 3.7.1

Let $\underline{d}'_{(1)}\underline{Y}, \underline{d}'_{(2)}\underline{Y}, \ldots, \underline{d}'_{(k)}\underline{Y}$ *be* k *mutually orthogonal linear functions which are elements of the linear set* $L(C'\underline{Y})$ *with* k *degrees of freedom, then the sum of squares of* $L(C'\underline{Y})$ *is the aggregate of the individual sums of squares of the linear functions* $\underline{d}'_{(1)}\underline{Y}, \underline{d}'_{(2)}\underline{Y}, \ldots, \underline{d}'_{(k)}\underline{Y}$.

Proof

According to Theorems 1.11.6 and 1.7.4, k mutually orthogonal column vectors $\underline{d}_{(1)}, \underline{d}_{(2)}, \ldots, \underline{d}_{(k)}$ can be chosen which generate V(C). Therefore, there exists k mutually orthogonal linear functions $\underline{d}'_{(1)}\underline{Y}, \underline{d}'_{(2)}\underline{Y}, \ldots, \underline{d}'_{(k)}\underline{Y}$ which generate $L(C'\underline{Y})$.

Let $D: n \times k = [\underline{d}_{(1)}, \underline{d}_{(2)}, \ldots, \underline{d}_{(k)}]$, then $L(D'\underline{Y})$ is the same as $L(C'\underline{Y})$ and therefore it follows that :

$$\begin{aligned}
Q &= \underline{Y}'C(C'C)^c C'\underline{Y} \\
&= \underline{Y}'D(D'D)^c D'\underline{Y} \\
&= \underline{Y}'D(D'D)^{-1} D'\underline{Y}, \text{ because } \operatorname{rank}(D'D) = \operatorname{rank}(D) = k \text{ so that } D'D \text{ is non-singular} \\
&= \underline{Y}'D\begin{bmatrix} \underline{d}'_{(1)}\underline{d}_{(1)} & 0 & 0 & \cdots & 0 \\ 0 & \underline{d}'_{(2)}\underline{d}_{(2)} & 0 & \cdots & 0 \\ 0 & 0 & \underline{d}'_{(3)}\underline{d}_{(3)} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \underline{d}'_{(k)}\underline{d}_{(k)} \end{bmatrix}^{-1} D'\underline{Y} \text{ because } \underline{d}'_{(i)}\underline{d}_{(j)} = 0 \text{ if } i \neq j \\
&= \underline{Y}'D\begin{bmatrix} \{\underline{d}'_{(1)}\underline{d}_{(1)}\}^{-1} & 0 & 0 & \cdots & 0 \\ 0 & \{\underline{d}'_{(2)}\underline{d}_{(2)}\}^{-1} & 0 & \cdots & 0 \\ 0 & 0 & \{\underline{d}'_{(3)}\underline{d}_{(3)}\}^{-1} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \{\underline{d}'_{(k)}\underline{d}_{(k)}\}^{-1} \end{bmatrix} D'\underline{Y} \\
&= \underline{Y}'\left[\{\underline{d}'_{(1)}\underline{d}_{(1)}\}^{-1}\underline{d}_{(1)}, \{\underline{d}'_{(2)}\underline{d}_{(2)}\}^{-1}\underline{d}_{(2)}, \cdots, \{\underline{d}'_{(k)}\underline{d}_{(k)}\}^{-1}\underline{d}_{(k)}\right]\begin{bmatrix} \underline{d}'_{(1)}\underline{Y} \\ \underline{d}'_{(2)}\underline{Y} \\ \cdots \\ \underline{d}'_{(k)}\underline{Y} \end{bmatrix} \\
&= \left[\{\underline{d}'_{(1)}\underline{d}_{(1)}\}^{-1}\underline{Y}'\underline{d}_{(1)}, \{\underline{d}'_{(2)}\underline{d}_{(2)}\}^{-1}\underline{Y}'\underline{d}_{(2)}, \cdots, \{\underline{d}'_{(k)}\underline{d}_{(k)}\}^{-1}\underline{Y}'\underline{d}_{(k)}\right]\begin{bmatrix} \underline{d}'_{(1)}\underline{Y} \\ \underline{d}'_{(2)}\underline{Y} \\ \cdots \\ \underline{d}'_{(k)}\underline{Y} \end{bmatrix} \\
&= \sum_{i=1}^{k} \frac{\{\underline{d}'_{(i)}\underline{Y}\}^2}{\underline{d}'_{(i)}\underline{d}_{(i)}}
\end{aligned}$$

i.e. $Q = \sum_{i=1}^{k} Q_i$ where Q_i is the sum of squares of the linear function $\underline{d}'_{(i)}\underline{Y}$.

Corollary 3.7.2

It follows that $E(Q) = E(Q_1) + E(Q_2) + \ldots + E(Q_k)$

$$= \sigma^2 + \frac{\{\underline{d}'_{(1)} E(\underline{Y})\}^2}{\underline{d}'_{(1)}\underline{d}_{(1)}} + \sigma^2 + \frac{\{\underline{d}'_{(2)} E(\underline{Y})\}^2}{\underline{d}'_{(2)}\underline{d}_{(2)}} + \ldots + \sigma^2 + \frac{\{\underline{d}'_{(k)} E(\underline{Y})\}^2}{\underline{d}'_{(k)}\underline{d}_{(k)}}$$

$$= k\sigma^2 + Q_m$$

where Q_m *is obtained from* Q *by replacing* $\underline{Y}$ *with* $E(\underline{Y})$ *in* Q.

In particular, if the k linear functions belong to the error set, $E(Q) = k\sigma^2$.

Theorem 3.7.2

Let $L_1 = L(C_1'\underline{Y})$ *and* $L_2 = L(C_2'\underline{Y})$ *be two linear sets with* k_1 *and* k_2 *degrees of freedom (d. f.) respectively such that every linear function of* L_1 *is orthogonal to every linear function of* L_2. *Further, let* $L = L(C'\underline{Y})$ *be the linear set with* $(k_1 + k_2)$ *d. f. which contains* L_1 *and* L_2 *as subsets. If* Q_1 *and* Q_2 *are respectively the sums of squares of the linear sets* L_1 *and* L_2 *and* Q *is the sum of squares of the linear set* L, *then* $Q = Q_1 + Q_2$.

Proof

It is always possible to find k_i mutually orthogonal linear functions which form a basis for L_i (i = 1; 2). Let $\underline{c}'_{11}\underline{Y}, \underline{c}'_{12}\underline{Y}, \cdots, \underline{c}'_{1k_1}\underline{Y}$ be an orthogonal basis for L_1 and $\underline{c}'_{21}\underline{Y}, \underline{c}'_{22}\underline{Y}, \cdots, \underline{c}'_{2k_2}\underline{Y}$ be an orthogonal basis for L_2 which is orthogonal to L_1 . If L is a linear set with $(k_1 + k_2)$ d. f. and contains L_1 and L_2 as subsets, then the linear functions which form a basis for L, $\underline{c}'_{11}\underline{Y}, \underline{c}'_{12}\underline{Y}, \cdots, \underline{c}'_{1k_1}\underline{Y}, \underline{c}'_{21}\underline{Y}, \underline{c}'_{22}\underline{Y}, \cdots, \underline{c}'_{2k_2}\underline{Y}$ will be mutually orthogonal.

According to Theorem 3.7.1 Q, the sum of squares of the linear set L, is given by $Q = Q_{11} + \cdots + Q_{1k_1} + Q_{21} + \cdots + Q_{2k_2}$ where Q_{ij} is the sum of squares of the linear function $\underline{c}'_{ij}\underline{Y}$ (i = 1, 2; j = 1, 2, . . . , k_i). Therefore,

$$Q = (Q_{11} + \cdots + Q_{1k_1}) + (Q_{21} + \cdots + Q_{2k_2})$$

$$= Q_1 + Q_2$$

where $Q_1 = Q_{11} + \cdots + Q_{1k_1}$ is the sum of squares of the linear set L_1 and $Q_2 = Q_{21} + \cdots + Q_{2k_2}$ is the sum of squares of the linear set L_2.

Corollary 3.7.3

If a linear set with r *d.o.f. is subdivided into* t *mutually orthogonal sets of linear functions with* $r_1, r_2, \ldots, r_t$ *d.o.f. respectively and* $r = r_1 + r_2 + \ldots + r_t$, *then the sum of squares* Q *with* r *d.o.f. can be correspondingly subdivided into the component sum of squares* $Q_1, Q_2, \ldots, Q_t$ *with* $r_1, r_2, \ldots, r_t$ *d.o.f. respectively such that*

$Q = Q_1 + Q_2 + \ldots + Q_t$.

Corollary 3.7.4

The sum of squares of the linear set of all linear functions of the variates $Y_1, Y_2, \ldots, Y_n$ *is* $\underline{Y}'\underline{Y} = \sum_{i=1}^{n} Y_i^2$ *, with* n *d. f. This follows immediately from the previous paragraphs if it is noted that* $Y_1, Y_2, \ldots, Y_n$ *are* n *mutually orthogonal linear functions which form a basis for the linear set of all linear functions of* $Y_1, Y_2, \ldots, Y_n$. *(If* $C = I: n \times n$ *in* $L(C'\underline{Y})$, *then it is clear that* $Y_1, Y_2, \ldots, Y_n$ *is an orthogonal basis for the linear set.)*

Remark 3.7.1

If a set of linear functions of the parameters has k d. f., then the linear set of their best estimators, in correspondence with Theorem 3.5.5, will also have k d.f. Therefore, it can be said that the k d. f. belong to the linear functions of the parameters or to their best estimators. The corresponding sum of squares will always be associated with (and calculated for) the best estimators.

Theorem 3.7.3

If $L(A'\underline{Y})$ *is the linear set of all best estimators and has* $r = \text{rank}(A)$ *d. f. and if* Q_0 *is the corresponding sum of squares, then* $Q_0 = \hat{\underline{\Pi}}'A'\underline{Y}$, *where* $\hat{\underline{\Pi}}$ *is a solution of the normal equations* $A'A\hat{\underline{\Pi}} = A'\underline{Y}$.

Proof

In correspondence with equation (3.7.1), $Q_0 = \underline{Y}'A(A'A)^c A'\underline{Y}$.

If $\hat{\underline{\Pi}}$ satisfies the equations $A'A\underline{\pi} = A'\underline{Y}$, it follows that $\hat{\underline{\Pi}} = (A'A)^c A'\underline{Y}$

so that $Q_0 = \underline{Y}'A\hat{\underline{\Pi}} = \hat{\underline{\Pi}}'A'\underline{Y}$.

Remark 3.7.2

1. Because $L(A'Y)$ has r d. f. these r d. f. are associated with Q_0.
2. Because the matrix $A(A'A)^c A'$ is independent of the choice of a conditional inverse of $A'A$, Q_0 is unique, *i.e.* independent of the choice of a conditional inverse for A'A.
3. If $A'A$ is non-singular, then $Q = \underline{Y}'A(A'A)^{-1} A'\underline{Y}$.

Definition 3.7.3 (Sum of squares for errors)

The set of all linear functions which belong to the error set is the set $L(E'\underline{Y})$ *with* $n_e = n - r$ *d. f. where* $V(E)$ *is the vector space orthogonal to* $V(A)$. *The sum of squares of this set is called the sum of squares for errors and denoted by* Q_e .

Now let $\underline{e}'_1\underline{Y}, \underline{e}'_2\underline{Y}, \cdots, \underline{e}'_{n_e}\underline{Y}$ be an orthogonal basis for the error set ≡ L(E′$\underline{Y}$). Then the sum of squares for errors can be written as

$$Q_e = \sum_j \frac{(\underline{e}'_j\underline{Y})^2}{\underline{e}'_j\underline{e}_j} \quad \text{and further, } E(Q_e) = n_e\,\sigma^2 \ \ (\textit{Why?}).$$

The preceding expression for Q_e is not convenient for calculation purposes. A convenient form for Q_e can be found with the help of the following corollary:

Corollary 3.7.5

$$\underline{Y}'\underline{Y} = Q_0 + Q_e \tag{3.7.2}$$

i.e. the sum of squares for all linear functions of $Y_1, Y_2, \ldots, Y_n$ *is equal to the sum of squares for the best estimators plus the sum of squares for errors (compare with Theorem 3.7.2). The d. f. is split correspondingly as* $n = r + n_e$.

Further, it follows from equation (3.7.2) *that*

$$\begin{aligned} Q_e &= \underline{Y}'\underline{Y} - Q_0 \\ &= \underline{Y}'\underline{Y} - \underline{Y}'A\hat{\underline{\Pi}} = \underline{Y}'\underline{Y} - \hat{\underline{\Pi}}'A'\underline{Y}. \end{aligned}$$

Corollary3.7.6

It has already been proved that the minimum value of $Q = (\underline{Y}' - \underline{\pi}'A')(\underline{Y} - A\underline{\pi})$ *can be obtained by letting* $\underline{\pi} = \hat{\underline{\Pi}}$ *where* $\hat{\underline{\Pi}}$ *is a solution of the normal equations* $A'A\hat{\underline{\Pi}} = A'\underline{Y}$. *Therefore,*

$$\begin{aligned} Q_{min} &= (\underline{Y}' - \hat{\underline{\Pi}}'A')(\underline{Y} - A\hat{\underline{\Pi}}) \\ &= \underline{Y}'\underline{Y} - \underline{Y}'A\hat{\underline{\Pi}} - \hat{\underline{\Pi}}'A'\underline{Y} + \hat{\underline{\Pi}}'A'A\hat{\underline{\Pi}} \\ &= \underline{Y}'\underline{Y} - \underline{Y}'A\hat{\underline{\Pi}} \\ &= Q_e. \end{aligned}$$

Therefore, it follows that $Q_e = \sum_{i=1}^{n}(Y_i - \sum_{j=1}^{m} a_{ij}\hat{\Pi}_j)^2$ *where* A *is of size* $n \times m$.

Definition 3.7.4

The mean sum of squares is defined as the sum of squares divided by the d. f. associated with the sum of squares concerned and it is denoted by S^2**.**

Corollary 3.7.7

The mean sum of squares for errors with n_e *d.f. is an unbiased estimator for* σ^2 *because*

$$E(Q_e / n_e) = \frac{1}{n_e}E(Q_e) = \frac{1}{n_e}n_e\sigma^2 = \sigma^2.$$

The mean sum of squares for errors, namely Q_e / n_e *can be written as* S_e^2. *Therefore, it follows that* S_e^2 *is an unbiased estimator for* σ^2.

Remark 3.7.3

The observed sum of squares follows from the preceding as follows:

For the best estimators: $q_0 = \underline{y}'A\underline{\hat{\pi}} = \underline{\hat{\pi}}'A'\underline{y}$ (where $\underline{\hat{\pi}}$ is a solution of the normal equations $A'A\underline{\hat{\pi}} = A'\underline{y}$)

The d. f. associated with q_0 is rank(A) = r.

For errors: $q_e = \underline{y}'\underline{y} - \underline{\hat{\pi}}'A'\underline{y}$ with $n_e = n - r$ d. f.

Total: $q_t = \underline{y}'\underline{y}$ with n d. f.

3.8 RELATIONSHIP BETWEEN SUMS OF SQUARES AND QUADRATIC FORMS

Is any quadratic form also a sum of squares in the sense of Definition 3.7.2? In this section we provide the answer to this question. In correspondence with Definition 3.7.2 it follows that *the* ***sum of squares*** *associated with the linear set* $\mathscr{L}(A'\underline{y})$, *can be expressed as* $\underline{y}'A(A'A)^cA'\underline{y}$ *where the* $n \times n$ *matrix* $A(A'A)^cA'$ *is the projection matrix which projects the vector* $\underline{y}: n \times 1$ *onto the column space of the* $n \times m$ *matrix* A *of rank* r. *The* ***d. f.*** *of the sum of squares is defined as the rank of the matrix* A.

The following lemmas are needed:

Lemma 3.8.1

Let A: $n \times n$ *and* B: $n \times n$ *be idempotent matrices. The sum* A + B *is idempotent if and only if* AB = BA = 0.

Proof

Let AB = BA = 0.

Then it follows that $(A + B)(A + B) = AA + BB + AB + BA = A + B$

i.e. A + B is idempotent.

Conversely:

Let A + B be idempotent.

Then, $(A + B)(A + B) = A + B$

i.e. $AA + BB + AB + BA = A + B$

i.e. $A + B + AB + BA = A + B$

i.e. $AB + BA = 0$.

Therefore, it follows that $AB + ABA = 0$ and $ABA + BA = 0$

so that $AB - BA = 0$.

But, $AB + BA = 0$ so that it follows that $AB = BA = 0$. ♦

Lemma 3.8.2

For any matrix $A: n \times m$, *the null space of* A, *namely* $\mathcal{N}(A)$, *is equal to the column space of* $I_m - A^cA$, *namely* $\mathcal{C}(I_m - A^cA)$.

Proof

Let $\underline{x}: m \times 1 \in \mathcal{C}(I_m - A^cA)$.

Then, $\underline{x} = (I - A^cA)\underline{d}$ for some $\underline{d}: m \times 1$

i.e. $A\underline{x} = A(I - A^cA)\underline{d} = (A - A A^cA)\underline{d} = \underline{0}$

i.e. $\underline{x} \in \mathcal{N}(A)$.

Conversely:

Let $\underline{x}: m \times 1 \in \mathcal{N}(A)$

i.e. $A\underline{x} = \underline{0}$

so that $\underline{x} = \underline{x} - BA\underline{x}$ for any B of size $m \times n$.

Therefore, $\underline{x} = \underline{x} - A^cA\underline{x}$ for any conditional inverse A^c of A

i.e. $\underline{x} = (I_m - A^cA)\underline{x} \in \mathcal{C}(I_m - A^cA)$.

As a result, $\mathcal{N}(A) = \mathcal{C}(I_m - A^cA)$. ♦

Lemma 3.8.3

The matrix $A: n \times n$ *is idempotent if and only if the null space of A is equal to the column space of* $(I - A)$.

Proof

Let $\mathcal{N}(A) = \mathcal{C}(I - A)$.

Then, $(I - A)\underline{d} \in \mathcal{C}(I - A)$ for *any* vector $\underline{d}: n \times 1$

i.e. $(I - A)\underline{d} \in \mathcal{N}(A)$ for *any* vector $\underline{d}: n \times 1$

i.e. $A(I - A)\underline{d} = \underline{0}$ for *any* vector $\underline{d}: n \times 1$

i.e. $A\,\underline{d} = AA\underline{d}$ for *any* vector $\underline{d}: n \times 1$

i.e. A is idempotent.

Conversely:

Let A be idempotent.

Then it follows that $AAA = AA = A$,

i.e. A is also a conditional inverse.

But, according to Lemma 3.8.2, $\mathcal{N}(A) = \mathcal{V}(I_m - A^cA)$ for *any* $A: n \times m$ so that for idempotent $A: n \times n$ it follows that $\mathcal{N}(A) = \mathcal{V}(I_n - A^cA) = \mathcal{V}(I_n - AA) = \mathcal{V}(I_n - A)$. ♦

Lemma 3.8.4

The matrix $A: n \times n$ *is idempotent if and only if* $\text{Rank}(A) + \text{Rank}(I - A) = n$.

Proof

According to Lemma 3.8.3, $A: n \times n$ idempotent if and only if $\mathcal{N}(A) = \mathcal{V}(I - A)$.

It follows for arbitrary $\underline{x}: n \times 1 \in \mathcal{N}(A)$ that $A\underline{x} = \underline{0}$

so that $\underline{x} = (I - A)\underline{x} \in \mathcal{V}(I - A)$.

As a result, $\mathcal{N}(A) \subset \mathcal{V}(I - A)$

and $\mathcal{N}(A) = \mathcal{V}(I - A)$ if and only if $\text{Dim}\{\mathcal{N}(A)\} = \text{Dim}\{\mathcal{V}(I - A)\}$.

Therefore, it follows that $A: n \times n$ is idempotent if and only if $\text{Dim}\{\mathcal{N}(A)\} = \text{Dim}\{\mathcal{V}(I - A)\}$

i.e. if and only if $n - \text{Rank}(A) = \text{Rank}(I - A)$

i.e. if and only if $\text{Rank}(I - A) + \text{Rank}(A) = n$. ♦

Lemma 3.8.5

Let $A_1, A_2, \ldots, A_k$ *be symmetric* $n \times n$ *matrices such that* $A_1 + A_2 + \ldots + A_k = I$.

Then it follows that each one of the allegations below imply the remaining two:

1. $A_iA_j = 0: n \times n$ for $j \neq i = 1, 2, \ldots, k$.
2. A_i *is idempotent for* $i = 1, 2, \ldots, k$.
3. $\text{Rank}(A_1) + \text{Rank}(A_2) + \ldots + \text{Rank}(A_k) = n$.

Proof

Let Allegation 1 be true.

The requirement that $A_iA_j = 0: n \times n$ for $j \neq i = 1, 2, \ldots, k$

implies that $A_i = A_iI = A_i(A_1 + A_2 + \ldots + A_k) = A_iA_i$.

Therefore, Allegation 1 implies Allegation 2.

Now let Allegation 2 be true.

The requirement that A_i is idempotent for $i = 1, 2, \ldots, k$

implies that $\text{Rank}(A_i) = \text{tr}(A_i)$ for $i = 1, 2, \ldots, k$

so that $\text{Rank}(A_1) + \text{Rank}(A_2) + \ldots + \text{Rank}(A_k) = \text{tr}(A_1) + \text{tr}(A_2) + \ldots + \text{tr}(A_k)$

$$= \text{tr}(A_1 + A_2 + \ldots + A_k)$$
$$= \text{tr}(I_n)$$
$$= n.$$

Therefore, Allegation 2 implies Allegation 3.

Now let Allegation 3 be true.

If $\text{Rank}(A_1) + \text{Rank}(A_2) + \ldots + \text{Rank}(A_k) = n$

It follows that $\text{Rank}(I - A_i) = \text{Rank}\left(\sum_{\substack{j=1 \\ j\neq i}}^{k} A_j\right) \leq \sum_{\substack{j=1 \\ j\neq i}}^{k} \text{Rank}\left(A_j\right) = n - \text{Rank}(A_i)$

i.e. $\text{Rank}(A_i) + \text{Rank}(I - A_i) \leq n.$

But, $\text{Rank}(A_i) + \text{Rank}(I - A_i) \geq \text{Rank}(A_i + I - A_i) = n$

i.e. $\text{Rank}(A_i) + \text{Rank}(I - A_i) = n$

so that with the help of Lemma 3.8.4 it follows that A_i is idempotent for $i = 1, 2, \ldots, k$.

Further, $\text{Rank}(I - A_i - A_j) = \text{Rank}\left(\sum_{\substack{t=1 \\ t\neq i;\, t\neq j}}^{k} A_t\right) \leq \sum_{\substack{t=1 \\ t\neq i;\, t\neq j}}^{k} \text{Rank}\left(A_t\right) = n - \text{Rank}(A_i) - \text{Rank}(A_j)$

i.e. $\text{Rank}(I - A_i - A_j) \leq n - \text{Rank}(A_i + A_j)$ (because $\text{Rank}(A_i + A_j) \leq \text{Rank}(A_i) + \text{Rank}(A_j)$)

i.e. $\text{Rank}(I - A_i - A_j) + \text{Rank}(A_i + A_j) \leq n.$

But, $\text{Rank}(I - A_i - A_j) + \text{Rank}(A_i + A_j) \geq \text{Rank}(I - A_i - A_j + A_i + A_j) = n$

so that $\text{Rank}(I - A_i - A_j) + \text{Rank}(A_i + A_j) = n$

and in correspondence with Lemma 3.8.4 it follows that $(A_i + A_j)$ is idempotent.

Further, it follows now, in correspondence with Lemma 3.8.1, that $A_iA_j = 0$ for all $j \neq i = 1, 2, \ldots, k$.

As a result, Allegation 3 implies that Allegation 1 is true and therefore the proof of Lemma 3.8.5 is complete. ♦

Consider the quadratic form $q = \underline{y}'A\underline{y}$ where $A: n \times n$ is **symmetric and idempotent**.

For such a quadratic form it is true that $A'A(A'A)A'A = A'(AA)A = A'AA = A'A$

so that $A'A$ is a conditional inverse for $A'A$. Further, the projection matrix which is associated with the column space of A is $A(A'A)^c A'$. Because this matrix is unique, it follows that the projection matrix concerned can also be written as $A(A'A)A'$ and because of the symmetry and idempotency of A it follows that the projection matrix is given by $AAA = AA = A$. From this it follows that the quadratic form q = $\underline{y}'A\underline{y}$ where A: n × n is **symmetric and idempotent** can also be interpreted as the *sum of squares* associated with the linear set $\mathscr{L}(A'\underline{y})$. The **degrees of freedom** of the **sum of squares** q is the **dimension** of the column space of A or the **rank** of the matrix A.

3.9 TRANSFORMATION OF PARAMETERS

Let the parameters $\pi_1, \pi_2, \ldots, \pi_m$ be transformed to a new set of parameters $\beta_1, \beta_2, \ldots, \beta_m$ by $\underline{\pi} = D\underline{\beta}$ where D: m × m is a non-singular matrix. This gives a one-to-one correspondence between $\underline{\pi}$ and $\underline{\beta}$. The linear function $\underline{k}'\underline{\pi}$ is transformed to the linear function $\underline{k}'D\underline{\beta} = (D'\underline{k})'\underline{\beta}$. Conversely, the linear function $\underline{g}'\underline{\beta}$ of $\underline{\beta}$ is transformed to a linear function of $\underline{\pi}$ by using the relationship $\underline{\beta} = D^{-1}\underline{\pi}$ *i.e.* $\underline{g}'\underline{\beta}$ is transformed to

$$\underline{g}'D^{-1}\underline{\pi} = \{(D^{-1})'\underline{g}\}'\underline{\pi}.$$

Now, $E(\underline{Y}) = A\underline{\pi} = AD\underline{\beta}$. If $\underline{k}'\underline{\pi}$ is linearly estimable, then $\underline{k}'$ is of the form $\underline{c}'A$ and in this case $\underline{k}'D\underline{\beta} = \underline{c}'AD\underline{\beta}$ is also linearly estimable. The normal equations become $D'A'AD\hat{\underline{\beta}} = D'A'\underline{y}$. Because D′ is non-singular, the normal equations reduce from left multiplication by $(D')^{-1}$ to $A'AD\hat{\underline{\beta}} = A'\underline{y}$. If $\underline{b}$ is a solution of the before mentioned system, then $AD\underline{b}$ is uniquely determined and $A'AD\underline{b} = A'\underline{y}$. If $\hat{\underline{\pi}}$ is a solution for the normal equations $A'A\hat{\underline{\pi}} = A'\underline{y}$, then $A\hat{\underline{\pi}}$ $\underline{\beta}$ is uniquely determined. It follows that $A'AD\underline{b} = A'A\hat{\underline{\pi}}$ and that $AD\underline{b} = A\hat{\underline{\pi}}$.

Further, the best estimator($\underline{k}'\underline{\pi}$) = $\underline{k}'\hat{\underline{\Pi}} = \underline{c}'A\hat{\underline{\Pi}} = \underline{c}'AD\underline{B}$ = best estimator($\underline{k}'D\underline{\beta}$). The sum of squares for errors is $\underline{Y}'\underline{Y} - \underline{Y}'A\hat{\underline{\Pi}} = \underline{Y}'\underline{Y} - \underline{Y}'AD\underline{B}$.

The preceding results are stated in the following theorem:

Theorem 3.9.1

If the parameters π_1 , π_2 , . . . , π_m *are transformed to a new set of parameters* β_1 , β_2 , . . . , β_m *by using the transformation* $\underline{\pi} = D\underline{\beta}$ *where* D: $m \times m$ *is a non-singular matrix, then*

(i) *estimable functions are transformed into estimable functions,*

(ii) *the best estimators of a linear function are the same as the best estimators of its transformation,*

(iii) *the sum of squares for errors is invariant.*

****