Chapter 1

LINEAR MODELS AND QR DECOMPOSITION

1 INTRODUCTION

A common problem in statistics is that of estimating parameters of some assumed relationship between one or more variables. One such relationship is

\[ y = f(a_1, \ldots, a_n), \quad (1.1) \]

where \( y \) is the dependent (endogenous, explained) variable and \( a_1, \ldots, a_n \) are the independent (exogenous, explanatory) variables. Regression analysis estimates the form of the relationship \((1.1)\) by using the observed values of the variables. This attempt at describing how these variables are related to each other is known as model building.

Exact functional relationships such as \((1.1)\) are inadequate descriptions of statistical behavior. Thus, the specification of the relationship \((1.1)\) is explained as

\[ y = f(a_1, \ldots, a_n) + \epsilon, \quad (1.2) \]

where \( \epsilon \) is the disturbance term or error, whose specific value in any single observation cannot be predicted. The purpose of \( \epsilon \) is to characterize the discrepancies that emerge between the actual observed value of \( y \) and the values that would be assigned by an exact functional relationship. The difference between the observed and predicted value of \( y \) is called the residual.

2 LINEAR MODEL SPECIFICATION

A linear model is one in which \( y \), or some transformation of \( y \), can be expressed as a linear function of \( a_i \), or some transformation of \( a_i \) \((i = 1, \ldots, n)\). Here only linear models where endogenous and exogenous variables do not require any transformations will be considered. In this case, the relationship
(1.2) can be written as

\[ y = x_1a_1 + x_2a_2 + \cdots + x_na_n + \varepsilon, \tag{1.3} \]

where \( x_i \) \((i = 1, \ldots, n)\) are unknown constants.

If there are \( m \) \((m > n)\) sample observations, the linear model (1.3) gives rise to the following set of \( m \) equations

\[
\begin{align*}
Y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + \varepsilon_1 \\
Y_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + \varepsilon_2 \\
&\vdots \\
Y_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n + \varepsilon_m
\end{align*}
\]

or

\[
\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix}. \tag{1.4}
\]

In compact form the latter can be written as

\[ y = Ax + \varepsilon, \tag{1.5} \]

where \( y, \varepsilon \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n} \) and \( x \in \mathbb{R}^n \).

To complete the description of the linear model (1.5), characteristics of the error term \( \varepsilon \) and the matrix \( A \) must be specified. The first assumption is that the expected value of \( \varepsilon \) is zero, that is, \( E(\varepsilon) = 0 \). The second assumption is that the various values of \( \varepsilon \) are normally distributed. The final assumption is that \( A \) is a non–stochastic matrix, which implies \( E(A^T \varepsilon) = 0 \). In summary, the complete mathematical specification of the (general) linear model which is being considered is

\[ y = Ax + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 \Omega). \tag{1.6} \]

The notation \( \varepsilon \sim N(0, \sigma^2 \Omega) \) indicates that the error vector \( \varepsilon \) is assumed to come from a normal distribution with mean zero and variance–covariance (or dispersion) matrix \( \sigma^2 \Omega \), where \( \Omega \) is a symmetric non–negative definite matrix and \( \sigma \) is an unknown scalar [124].

### 2.1 THE ORDINARY LINEAR MODEL

Consider the Ordinary Linear Model (OLM):

\[ y = Ax + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_m). \tag{1.7} \]