2

The Simple Linear Regression Model

In this chapter, we consider the modeling between the dependent and an independent variable. When there is only one independent variable in the linear regression model, the model is generally termed as simple linear regression model. When there are more than one independent variables, then the linear model is termed as the multiple linear regression model which is the subject matter of the next chapter 3. The contents of this chapter will help the reader in better understanding of further chapters.

2.1 The Linear Model

Consider a simple linear regression model

$$y = \beta_0 + \beta_1 X + e \tag{2.1}$$

where y is the dependent (or study) variable and X is an independent (or explanatory) variable. The parameters β_0 and β_1 are the intercept term and slope parameter, respectively, which are usually called as regression coefficients. The unobservable error component e accounts for the failure of data to lie on the straight line and represents the difference between the true and observed realizations of y. For the purpose of statistical inferences, we assume that e is observed as independent and identically distributed random variable with mean zero and constant variance σ^2.

The independent variable is viewed as controlled by the experimenter, so it is considered as non-stochastic whereas y is viewed as a random variable with

$$\mathrm{E}(y) = \beta_0 + \beta_1 X$$

and

$$\mathrm{var}(y) = \sigma^2 \ .$$

Sometimes X can also be a random variable. Such an aspect is explained later in Section 2.15. In such a case, instead of simple mean and simple variance of y, we consider the conditional mean of y given $X = x$ as

$$\mathrm{E}(y|x) = \beta_0 + \beta_1 x$$

and the conditional variance of y given $X = x$ as

$$\mathrm{var}(y|x) = \sigma^2 \ .$$

The parameters β_0, β_1 and σ^2 are generally unknown and e is unobserved. The determination of the statistical model (2.1) depends on the determination (*i.e.*, estimation) of β_0, β_1 and σ^2.

Only T pairs of observations (x_t, y_t) $(t = 1, \ldots, T)$ on (X, y) are observed which are used to determine the unknown parameters.

Different methods of estimation can be used to determine the estimates of the parameters. Among them, the least squares and maximum likelihood principles are the most popular methods of estimation.

2.2 Least Squares Estimation

We observe a sample of T sets of observations (x_t, y_t) $(t = 1, \ldots, T)$ and in view of (2.1), we can write

$$y_t = \beta_0 + \beta_1 x_t + e_t \quad (t = 1, \ldots, T) \ . \tag{2.2}$$

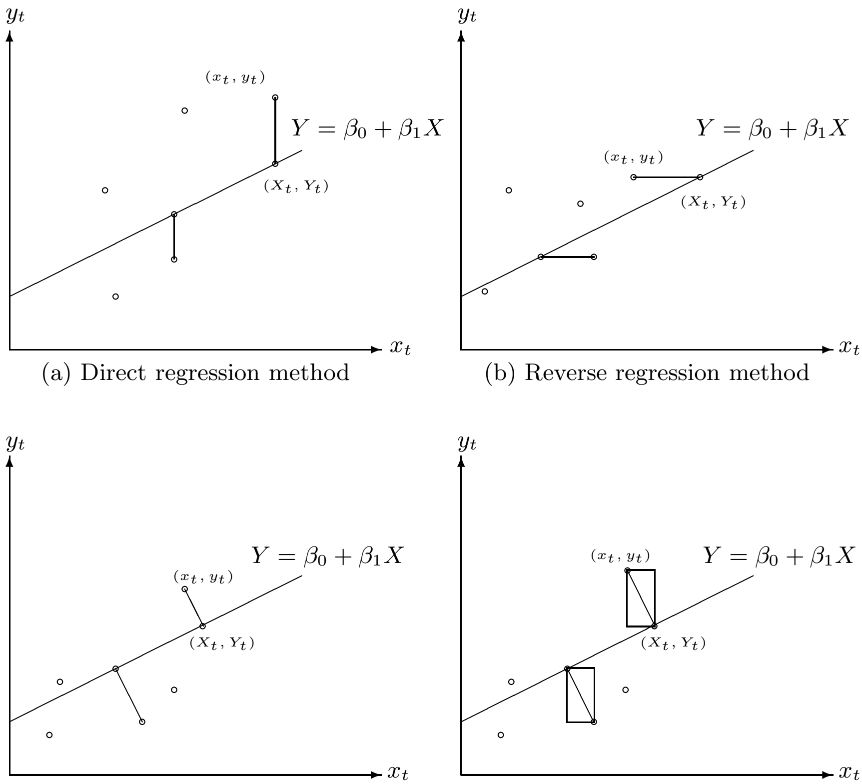
The principle of least squares aims at estimating β_0 and β_1 so that the sum of squares of difference between the observations and the line in the scatter diagram is minimum. Such an idea is viewed from different perspectives. When the vertical difference between the observations and the line in the scatter diagram (see Fig. 2.1(a)) is considered and its sum of squares is minimized to obtain the estimates of β_0 and β_1, the method is known as *direct regression*.

Another approach is to minimize the sum of squares of difference between the observations and the line in horizontal direction in the scatter diagram (see Fig. 2.1(b)) to obtain the estimates of β_0 and β_1. This is known as *reverse (or inverse) regression* method.

Alternatively, the sum of squares of perpendicular distance between the observations and the line in the scatter diagram (see Fig. 2.1(c)) is minimized to obtain the estimates of β_0 and β_1. This is known as *orthogonal regression* or *major axis regression* method.

The *least absolute deviation regression* method considers the sum of the absolute deviation of the observations from the line in the vertical direction in the scatter diagram (see Fig. 2.1(a)) to obtain the estimates of β_0 and β_1.

The *reduced major axis regression* method proposes to minimize the sum of the areas of rectangles defined between the observed data points and the nearest point on the line in the scatter diagram to obtain the estimates of the regression coefficients (see Fig. 2.1(d)).



(a) Direct regression method (b) Reverse regression method

(c) Major axis regression method (d) Reduced major axis method

FIGURE 2.1. Scatter diagrams of different methods of regression

One may note that the principle of least squares does not require any assumption about the form of probability distribution of e_t in deriving the least squares estimates. For the purpose of deriving the statistical inferences only, we assume that e_t's are observed as random variable ϵ_t with $\mathrm{E}(\epsilon_t) = 0$, $\mathrm{var}(\epsilon_t) = \sigma^2$ and $\mathrm{cov}(\epsilon_t, \epsilon_{t^*}) = 0$ for all $t \neq t^*$ $(t, t^* = 1, \ldots, T)$. This assumption is needed to find the mean and variance of the least squares estimates. The assumption that ϵ_t's are normally distributed is utilized while constructing the tests of hypotheses and confidence intervals of the parameters.

Based on these approaches, different estimates of β_0 and β_1 are obtained which have different statistical properties. Among them the direct regression approach is more popular. Generally, the direct regression estimates are referred as the least squares estimates. We will consider here the direct regression approach in more detail. Other approaches are also discussed.

2.3 Direct Regression Method

This method is also known as the *ordinary least squares estimation*. The regression models (2.1) and (2.2) can be viewed as the regression models for population and sample, respectively. The direct regression approach minimizes the sum of squares

$$S(\beta_0, \beta_1) = \sum_{t=1}^{T} e_t^2 = \sum_{t=1}^{T} (y_t - \beta_0 - \beta_1 x_t)^2 \tag{2.3}$$

with respect to β_0 and β_1.

The partial derivatives of (2.3) with respect to β_0 and β_1 are

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = -2 \sum_{t=1}^{T} (y_t - \beta_0 - \beta_1 x_t) \tag{2.4}$$

and

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = -2 \sum_{t=1}^{T} (y_t - \beta_0 - \beta_1 x_t) x_t, \tag{2.5}$$

respectively. The solution of β_0 and β_1 is obtained by setting (2.4) and (2.5) equal to zero. Thus obtained solutions are called the direct regression estimators, or usually called as the Ordinary Least Squares (OLS) estimators of β_0 and β_1.

This gives the ordinary least squares estimates b_0 of β_0 and b_1 of β_1 as

$$b_0 = \bar{y} - b_1 \bar{x} \tag{2.6}$$

$$b_1 = \frac{SXY}{SXX} \tag{2.7}$$

where

$$SXY = \frac{1}{T} \sum_{t=1}^{T} (x_t - \bar{x})(y_t - \bar{y}), \quad SXX = \frac{1}{T} \sum_{t=1}^{T} (x_t - \bar{x})^2, \quad \bar{x} = \frac{1}{T} \sum_{t=1}^{T} x_t$$

and $\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t$.

Further, using (2.4) and (2.5), we have

$$\begin{aligned}
\frac{\partial^2 S(\beta_0,\beta_1)}{\partial \beta_0^2} &= -2\sum_{t=1}^{T}(-1) = 2T\,,\\
\frac{\partial^2 S(\beta_0,\beta_1)}{\partial \beta_1^2} &= 2\sum_{t=1}^{T} x_t^2\,,\\
\frac{\partial^2 S(\beta_0,\beta_1)}{\partial \beta_0 \partial \beta_1} &= 2\sum_{t=1}^{T} x_t = 2T\bar{x}\,.
\end{aligned}$$

Thus we get the Hessian matrix which is the matrix of second order partial derivatives as

$$\begin{aligned}
H &= \begin{pmatrix} \frac{\partial^2 S(\beta_0,\beta_1)}{\partial \beta_0^2} & \frac{\partial^2 S(\beta_0,\beta_1)}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 S(\beta_0,\beta_1)}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 S(\beta_0,\beta_1)}{\partial \beta_1^2} \end{pmatrix}\\
&= 2\begin{pmatrix} T & T\bar{x} \\ T\bar{x} & \sum_{t=1}^{T} x_t^2 \end{pmatrix}\\
&= 2\begin{pmatrix} \mathbf{1}' \\ x' \end{pmatrix}(\mathbf{1}, x) \qquad (2.8)
\end{aligned}$$

where $\mathbf{1} = (1,\ldots,1)'$ is a T-vector of elements unity and and $x = (x_1,\ldots,x_T)'$ is a T-vector of observations on X. The matrix (2.8) is positive definite if its determinant and the element in the first row and column of H are positive. The determinant of H is

$$\begin{aligned}
|H| &= 2\left(T\sum_{t=1}^{T} x_t^2 - T^2\bar{x}^2\right)\\
&= 2T\sum_{t=1}^{n}(x_t-\bar{x})^2\\
&\geq 0. \qquad (2.9)
\end{aligned}$$

The case when $\sum_{t=1}^{T}(x_t-\bar{x})^2 = 0$ is not interesting because then all the observations are identical, *i.e.*, $x_t = c$ (some constant). In such a case there is no relationship between x and y in the context of regression analysis. Since $\sum_{t=1}^{T}(x_t-\bar{x})^2 > 0$, therefore $|H| > 0$. So H is positive definite for any (β_0, β_1); therefore $S(\beta_0,\beta_1)$ has a global minimum at (b_0, b_1).

The fitted line or the fitted linear regression model is

$$y = b_0 + b_1 X \qquad (2.10)$$

and the predicted values are

$$\hat{y}_t = b_0 + b_1 x_t \quad (t = 1,\ldots,T)\,. \qquad (2.11)$$

The difference between the observed value y_t and the fitted (or predicted) value $\hat{y}_t$ is called as a residual. The t^{th} residual is

$$\begin{aligned}\hat{\epsilon}_t &= y_t \sim \hat{y}_t \quad (t = 1, \ldots, T) \\ &= y_t - \hat{y}_t \\ &= y_t - (b_0 + b_1 x_t) \ . \end{aligned} \tag{2.12}$$

2.4 Properties of the Direct Regression Estimators

Note that b_0 and b_1 from (2.6) and (2.7) are the linear combinations of y_t $(t = 1, \ldots, T)$.

Therefore

$$b_1 = \sum_{t=1}^{T} k_t y_t$$

where $k_t = (x_t - \bar{x})/SXX$. Since $\sum_{t=1}^{T} k_t = 0$ and $\sum_{t=1}^{T} k_t x_t = 1$,

$$\begin{aligned}\mathrm{E}(b_1) &= \sum_{t=1}^{T} k_t \, \mathrm{E}(y_t) = \sum_{t=1}^{T} k_t(\beta_0 + \beta_1 x_t) \\ &= \beta_1 \ . \end{aligned} \tag{2.13}$$

Similarly,

$$\mathrm{E}(b_0) = \beta_0 \ . \tag{2.14}$$

Thus b_0 and b_1 are unbiased estimators of β_0 and β_1, respectively.

The variance of b_1 is

$$\begin{aligned}\mathrm{var}(b_1) &= \sum_{t=1}^{T} k_t^2 \, \mathrm{var}(y_t) + \sum_{t} \sum_{t^* \neq t} k_t k_{t^*} \, \mathrm{cov}(y_t, y_{t^*}) \\ &= \sigma^2 \sum_{t=1}^{T} k_t^2 \qquad \text{(since } y_1, \ldots, y_T \text{ are independent)} \\ &= \frac{\sigma^2}{SXX}. \end{aligned} \tag{2.15}$$

Similarly, the variance of b_0 is

$$\begin{aligned}\mathrm{var}(b_0) &= \mathrm{var}(\bar{y}) + \bar{x}^2 \, \mathrm{var}(b_1) - 2\bar{x} \, \mathrm{cov}(\bar{y}, b_1) \\ &= \sigma^2 \left(\frac{1}{T} + \frac{\bar{x}^2}{SXX} \right) \qquad \text{(since } \mathrm{cov}(\bar{y}, b_1) = 0\text{)} \end{aligned} \tag{2.16}$$

Finally, the covariance between b_0 and b_1 is

$$\begin{aligned}\mathrm{cov}(b_0, b_1) &= \mathrm{cov}(\bar{y}, b_1) - \bar{x} \, \mathrm{var}(b_1) \\ &= -\frac{\sigma^2 \bar{x}}{SXX} \ . \end{aligned} \tag{2.17}$$

It can further be shown that the ordinary least squares estimators b_0 and b_1 possess the minimum variance in the class of linear and unbiased estimators. So they are termed as the Best Linear Unbiased Estimators (BLUE). Such a property is known as the Gauss-Markov Theorem which is discussed in the context of multiple linear regression model in next chapter 3.

The estimate of σ^2 can be obtained from the sum of squares of residuals as

$$\begin{aligned} RSS &= \sum_{t=1}^{T} \hat{\epsilon}_t^2 \\ &= \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 \\ &= \sum_{t=1}^{T} y_t^2 - n\bar{y}^2 - b_1 SXY \\ &= \sum_{t=1}^{T} (y_t - \bar{y})^2 - b_1 SXY \\ &= SYY - b_1 SXY \ . \end{aligned} \tag{2.18}$$

Thus $\mathrm{E}(RSS) = (T-2)\sigma^2$, so an unbiased estimator of σ^2 is

$$s^2 = \frac{RSS}{T-2} \ . \tag{2.19}$$

Note that RSS has only $(T-2)$ degrees of freedom. The two degrees of freedom are lost due to estimation of b_0 and b_1.

The estimators of variances of b_0 and b_1 are obtained as

$$\widehat{\mathrm{var}(b_0)} = s^2 \left(\frac{1}{T} + \frac{\bar{x}^2}{SXX} \right) \tag{2.20}$$

and

$$\widehat{\mathrm{var}(b_1)} = \frac{s^2}{SXX}, \tag{2.21}$$

respectively.

It is observed that $\sum_{t=1}^{T} \hat{\epsilon}_t = 0$. In the light of this property, $\hat{\epsilon}_t$ can be regarded as an estimate of unknown ϵ_t $(t = 1, \ldots, T)$ and helps in verifying the different model assumptions in the given sample. The methods to verify the model assumptions are discussed in chapter 7.

Further, note that $\sum_{t=1}^{T} x_t \hat{\epsilon}_t = 0$, $\sum_{t=1}^{T} \hat{y}_t \hat{\epsilon}_t = 0$, $\sum_{t=1}^{T} y_t = \sum_{t=1}^{T} \hat{y}_t$ and the fitted line always passes through $(\bar{x}, \bar{y})$.

2.5 Centered Model

Sometimes when the observations on an independent variable X are measured around its mean, then the model (2.2) can be expressed as the centered version of (2.2),

$$\begin{aligned} y_t &= \beta_0 + \beta_1(x_t - \bar{x}) + \beta_1\bar{x} + e_t \quad (t = 1, \ldots, T) \\ &= \beta_0^* + \beta_1(x_t - \bar{x}) + e_t \end{aligned} \tag{2.22}$$

where $\beta_0^* = \beta_0 + \beta_1\bar{x}$ which relates the models (2.2) and (2.22). The first order partial derivatives of

$$S(\beta_0^*, \beta_1) = \sum_{t=1}^{T} e_t^2 = \sum_{t=1}^{T} [y_t - \beta_0^* - \beta_1(x_t - \bar{x})]^2 \qquad \text{(cf. (2.22))}$$

with respect to β_0^* and β_1, when equated to zero yield the direct regression least squares estimates of β_0^* and β_1 as

$$b_0^* = \bar{y} \tag{2.23}$$

and

$$b_1 = \frac{SXY}{SXX}, \tag{2.24}$$

respectively.

Thus the form of the estimate of slope parameter β_1 remains same as from model (2.2) whereas the form of the estimate of intercept term changes in the models (2.2) and (2.22).

Further, the Hessian matrix of the second order partial derivatives of $S(\beta_0^*, \beta_1)$ with respect to β_0^* and β_1 is positive definite at $\beta_0^* = b_0^*$ and $\beta_1 = b_1$ which ensures that $S(\beta_0^*, \beta_1)$ is minimized at $\beta_0^* = b_0^*$ and $\beta_1 = b_1$.

Considering the deviation e as random variable denoted by ϵ, we assume that $\mathrm{E}(\epsilon) = 0$ and $\mathrm{E}(\epsilon\epsilon') = \sigma^2 I$. It follows then

$$\begin{aligned} \mathrm{E}(b_0^*) = \beta_0^* \quad &, \quad \mathrm{E}(b_1) = \beta_1, \\ \mathrm{var}(b_0^*) = \frac{\sigma^2}{T} \quad &, \quad \mathrm{var}(b_1) = \frac{\sigma^2}{SXX}. \end{aligned}$$

In this case, the fitted model of (2.22) is

$$y = \bar{y} + b_1(x - \bar{x}), \tag{2.25}$$

and the predicted values are

$$\hat{y}_t = \bar{y} + b_1(x_t - \bar{x}) \quad (t = 1, \ldots, T). \tag{2.26}$$

Another worth noticing aspect in centered model is that

$$\mathrm{cov}(b_0^*, b_1) = 0. \tag{2.27}$$

2.6 No Intercept Term Model

A no-intercept model is

$$y_t = \beta_1 x_t + e_t \quad (t = 1, \ldots, T). \tag{2.28}$$

Such model arises when $x_t = 0$ implies $y_t = 0$ $(t = 1, \ldots, T)$. For example, in analyzing the relationship between the velocity (y) of a car and its acceleration (X), the velocity is zero when acceleration is zero.

Using the data (x_t, y_t), $t = 1, \ldots, T$, the direct regression least squares estimate of β_1 is obtained by minimizing $S(\beta_1) = \sum_{t=1}^T e_t^2 = \sum_{t=1}^T (y_t - \beta_1 x_t)^2$ as

$$b_1^* = \frac{\sum_{t=1}^T y_t x_t}{\sum_{t=1}^T x_t^2} \,. \tag{2.29}$$

Assuming that e is observed as random variable ϵ with $\mathrm{E}(\epsilon) = 0$ and $\mathrm{var}(\epsilon) = \sigma^2$, it is seen that for the estimator (2.29),

$$\begin{aligned} \mathrm{E}(b_1^*) &= \beta \\ \mathrm{var}(b_1^*) &= \frac{\sigma^2}{\sum_{t=1}^T x_t^2} \end{aligned}$$

and an unbiased estimator of σ^2 is

$$\frac{\sum_{t=1}^T y_t^2 - b_1 \sum_{t=1}^T y_t x_t}{T-1}.$$

The second order partial derivative of $S(\beta_1)$ with respect to β_1 at $\beta_1 = b_1$ is positive which insures that b_1 minimizes $S(\beta_1)$.

2.7 Maximum Likelihood Estimation

We assume that e_t's in (2.2) are observed as random variable ϵ_t's $(t = 1, \ldots, T)$ which are independent and identically distributed with $N(0, \sigma^2)$. Now we use the method of maximum likelihood to estimate the parameters of the linear regression model (2.1).

In the linear regression model

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t \quad (t = 1, \ldots, T),$$

the observations y_t $(t = 1, \ldots, T)$ are independently distributed with $N(\beta_0 + \beta_1 x_t, \sigma^2)$ for all $t = 1, \ldots, T$. The likelihood function of the given observations (x_t, y_t) and unknown parameters β_0, β_1, and σ^2 is

$$L(x_t, y_t; \beta_0, \beta_1, \sigma^2) = \prod_{t=1}^T \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} \exp\left[-\frac{1}{2\sigma^2}(y_t - \beta_0 - \beta_1 x_t)^2\right] \,.$$

The maximum likelihood estimates of β_0, β_1 and σ^2 can be obtained by maximizing $L(x_t, y_t; \beta_0, \beta_1, \sigma^2)$ or equivalently $\ln L(x_t, y_t; \beta_0, \beta_1, \sigma^2)$ where

$$\begin{aligned}\ln L(x_t, y_t; \beta_0, \beta_1, \sigma^2) &= -\left(\frac{T}{2}\right)\ln 2\pi - \left(\frac{T}{2}\right)\ln \sigma^2 \\ &\quad -\left(\frac{1}{2\sigma^2}\right)\sum_{t=1}^{T}(y_t - \beta_0 - \beta_1 x_t)^2 \ . \qquad (2.30)\end{aligned}$$

The normal equations are obtained by partial differentiation of (2.30) with respect to β_0, β_1 and σ^2 as

$$\frac{\partial \ln L(x_t, y_t; \beta_0, \beta_1, \sigma^2)}{\partial \beta_0} = -\frac{1}{\sigma^2}\sum_{t=1}^{T}(y_t - \beta_0 - \beta_1 x_t) \ , \qquad (2.31)$$

$$\frac{\partial \ln L(x_t, y_t; \beta_0, \beta_1, \sigma^2)}{\partial \beta_1} = -\frac{1}{\sigma^2}\sum_{t=1}^{T}(y_t - \beta_0 - \beta_1 x_t)x_t \qquad (2.32)$$

and

$$\frac{\partial \ln L(x_t, y_t; \beta_0, \beta_1, \sigma^2)}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}(y_t - \beta_0 - \beta_1 x_t)^2 \ . \qquad (2.33)$$

The normal equations (2.31), (2.32) and (2.33) are equated to zero and the solution is the maximum likelihood estimates of β_0, β_1 and σ^2 as

$$\tilde{b}_0 = \bar{y} - \tilde{b}_1 \bar{x} \ , \qquad (2.34)$$

$$\tilde{b}_1 = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{T}(x_t - \bar{x})^2} = \frac{SXY}{SXX} \qquad (2.35)$$

and

$$\tilde{s}^2 = \frac{\sum_{t=1}^{T}(y_t - \tilde{b}_0 - \tilde{b}_1 x_t)^2}{T}, \qquad (2.36)$$

respectively.

It can be verified that the Hessian matrix of second order partial derivative of $\ln L$ with respect to β_0, β_1, and σ^2 is negative definite at $\beta_0 = \tilde{b}_0$, $\beta_1 = \tilde{b}_1$, and $\sigma^2 = \tilde{s}^2$ which ensures that the likelihood function is maximized at these values.

It is noted here that the least squares and maximum likelihood estimates of β_0 and β_1 are identical but estimates of σ^2 differ. In fact,

$$\tilde{s}^2 = \left(\frac{T-2}{T}\right)s^2 \qquad \text{(cf. (2.19))}. \qquad (2.37)$$

Thus $\tilde{b}_0$ and $\tilde{b}_1$ are unbiased estimators of β_0 and β_1 whereas $\tilde{s}^2$ is a biased estimate of σ^2, but it is asymptotically unbiased. The variances of $\tilde{b}_0$ and $\tilde{b}_1$ are same as of b_0 and b_1 respectively but $\mathrm{var}(\tilde{s}^2) < \mathrm{var}(s^2)$.

2.8 Testing of Hypotheses and Confidence Interval Estimation

Consider the simple linear regression model (2.2) under the assumption that ϵ_t's are independent and identically distributed with $N(0, \sigma^2)$.

First we develop a test for the null hypothesis related to the slope parameter

$$H_0 : \beta_1 = \beta_{10}$$

where β_{10} is some given constant.

Assuming σ^2 to be known,

$$b_1 \sim N\left(\beta, \frac{\sigma^2}{SXX}\right), \tag{2.38}$$

and so the statistic

$$Z_{01} = \frac{b_1 - \beta_{10}}{\sqrt{\frac{\sigma^2}{SXX}}} \sim N(0, 1), \tag{2.39}$$

when H_0 is true.

The $100(1 - \alpha)\%$ confidence interval for β_1 can be obtained from (2.39) as

$$b_1 \pm z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{SXX}} \tag{2.40}$$

where $z_{1-\alpha/2}$ is the (1- $\alpha/2$) percentage point of the $N(0, 1)$ distribution.

When σ^2 is unknown in (2.38), then we proceed as follows. We know

$$\mathrm{E}\left(\frac{RSS}{T-2}\right) = \sigma^2$$

and

$$\frac{RSS}{\sigma^2} \sim \chi^2_{T-2}.$$

Further, RSS/σ^2 and b_1 are independently distributed, so the statistic

$$\begin{aligned} t_{01} &= \frac{b_1 - \beta_{10}}{\sqrt{\frac{\hat{\sigma}^2}{SXX}}} \\ &= \frac{b_1 - \beta_{10}}{\sqrt{\frac{RSS}{(T-2)SXX}}} \sim t_{T-2}, \end{aligned} \tag{2.41}$$

when H_0 is true.

The corresponding $100(1-\alpha)\%$ confidence interval of β_1 can be obtained from (2.41) as

$$b_1 \pm t_{T-2,1-\alpha/2}\sqrt{\frac{RSS}{(T-2)SXX}} \ . \tag{2.42}$$

A decision rule to test $H_1 : \beta_1 \neq \beta_{10}$ can be framed from (2.39) or (2.41) under the condition when σ^2 is known or unknown, respectively. For example, when σ^2 is unknown, the decision rule is to reject H_0 if

$$|t_{01}| > t_{T-2,1-\alpha/2}$$

where $t_{T-2,1-\alpha/2}$ is the $(1-\alpha/2)$ percentage point of the t-distribution with $(T-2)$ degrees of freedom. Similarly, the decision rule for one sided alternative hypothesis can also be framed.

A similar test statistic and test procedure can be developed for testing the hypothesis related to the intercept term

$$H_0 : \beta_0 = \beta_{00} \ ,$$

where β_{00} is some given constant.

When σ^2 is known, then the statistic

$$Z_{00} = \frac{b_0 - \beta_{00}}{\sqrt{\sigma^2\left(\frac{1}{T} + \frac{\bar{x}^2}{SXX}\right)}} \ \sim \ N(0,1) \ , \tag{2.43}$$

when H_0 is true.

When σ^2 is unknown, then the statistic

$$t_{00} = \frac{b_0 - \beta_{00}}{\sqrt{\frac{RSS}{T-2}\left(\frac{1}{T} + \frac{\bar{x}^2}{SXX}\right)}} \ \sim \ t_{T-2} \ , \tag{2.44}$$

when H_0 is true.

The $100(1-\alpha)\%$ confidence intervals for β_0 can be derived from (2.43) and (2.44) as

$$b_0 \pm z_{1-\alpha/2}\sqrt{\sigma^2\left(\frac{1}{T} + \frac{\bar{x}^2}{SXX}\right)} \tag{2.45}$$

when σ^2 is known and

$$b_0 \pm t_{T-2,1-\alpha/2}\sqrt{\frac{RSS}{T-2}\left(\frac{1}{T} + \frac{\bar{x}^2}{SXX}\right)} \tag{2.46}$$

when σ^2 is unknown.

A confidence interval for σ^2 can also be derived as follows. Since $RSS/\sigma^2 \sim \chi^2_{T-2}$, thus

$$P\left[\chi^2_{T-2,\alpha/2} \leq \frac{RSS}{\sigma^2} \leq \chi^2_{T-2,1-\alpha/2}\right] = 1 - \alpha \ . \tag{2.47}$$

The corresponding $100(1-\alpha)\%$ confidence interval for σ^2 is

$$\frac{RSS}{\chi^2_{T-2,1-\alpha/2}} \leq \sigma^2 \leq \frac{RSS}{\chi^2_{T-2,\alpha/2}} \ . \tag{2.48}$$

A joint confidence region for β_0 and β_1 can also be found. Such region will provide a $100(1-\alpha)\%$ confidence that both the estimates of β_0 and β_1 are correct. Consider the centered version of the linear regression model (2.22),

$$y_t = \beta_0^* + \beta_1(x_t - \bar{x}) + \epsilon_t$$

where $\beta_0^* = \beta_0 + \beta_1\bar{x}$. The least squares estimators of β_0^* and β_1 from (2.23) and (2.24) are

$$b_0^* = \bar{y} \quad \text{and} \quad b_1 = \frac{SXY}{SXX},$$

respectively. Here

$$\begin{aligned} \mathrm{E}(b_0^*) &= \beta_0^* \quad , \quad \mathrm{E}(b_1) = \beta_1 \quad , \\ \mathrm{var}(b_0^*) &= \frac{\sigma^2}{T} \quad , \quad \mathrm{var}(b_1) = \frac{\sigma^2}{SXX}. \end{aligned}$$

When σ^2 is known, then the statistics

$$\frac{b_0^* - \beta_0^*}{\sqrt{\frac{\sigma^2}{T}}} \sim N(0,1) \tag{2.49}$$

and

$$\frac{b_1 - \beta_1}{\sqrt{\frac{\sigma^2}{SXX}}} \sim N(0,1) \tag{2.50}$$

are independently distributed. Thus

$$\left(\frac{b_0^* - \beta_0^*}{\sqrt{\frac{\sigma^2}{T}}}\right)^2 \sim \chi^2_1 \tag{2.51}$$

and

$$\left(\frac{b_1 - \beta_1}{\sqrt{\frac{\sigma^2}{SXX}}}\right)^2 \sim \chi^2_1 \tag{2.52}$$

are also independently distributed because b_0^* and b_1 are independently distributed. Consequently sum of (2.51) and (2.52),

$$\frac{T(b_0^* - \beta_0^*)^2}{\sigma^2} + \frac{SXX(b_1 - \beta_1)^2}{\sigma^2} \sim \chi^2_2 \ . \tag{2.53}$$

Since

$$\frac{SSE}{\sigma^2} \sim \chi^2_{T-2}$$

and SSE is independently distributed of b_0^* and b_1, so the ratio

$$\frac{\left(\frac{T(b_0^*-\beta_0^*)^2}{\sigma^2} + \frac{SXX(b_1-\beta_1)^2}{\sigma^2}\right)/2}{\left(\frac{RSS}{\sigma^2}\right)/(T-2)} \sim F_{2,T-2} \,. \tag{2.54}$$

Substituting $b_0^* = b_0 + b_1\bar{x}$ and $\beta_0^* = \beta_0 + \beta_1\bar{x}$ in (2.54), we get

$$\left(\frac{T-2}{2}\right)\left[\frac{Q_f}{RSS}\right]$$

where

$$Q_f = T(b_0-\beta_0)^2 + 2\sum_{t=1}^{T} x_t(b_0-\beta_0)(b_1-\beta_1) + \sum_{t=1}^{T} x_t^2(b_1-\beta_1)^2 \,. \tag{2.55}$$

Since

$$P\left[\left(\frac{T-2}{2}\right)\frac{Q_f}{RSS} \le F_{2,T-2}\right] = 1-\alpha \tag{2.56}$$

holds true for all values of β_0 and β_1, so the $100(1-\alpha)\%$ confidence region for β_0 and β_1 is

$$\left(\frac{T-2}{2}\right)\frac{Q_f}{RSS} \le F_{2,T-2;1-\alpha} \,. \tag{2.57}$$

The confidence region given by (2.57) is an ellipse which gives the $100(1-\alpha)\%$ probability that β_0 and β_1 are contained simultaneously in this ellipse.

2.9 Analysis of Variance

A test statistic for testing $H_0 : \beta_1 = 0$ can also be formulated using the analysis of variance technique as follows.

On the basis of the identity

$$y_t - \bar{y}_t = (y_t - \bar{y}) - (\hat{y}_t - \bar{y}) \,,$$

the sum of squared residuals is

$$\begin{aligned} S(b) &= \sum_{t=1}^{T}(y_t-\hat{y}_t)^2 \\ &= \sum_{t=1}^{T}(y_t-\bar{y})^2 + \sum_{t=1}^{T}(\hat{y}_t-\bar{y})^2 - 2\sum_{t=1}^{T}(y_t-\bar{y})(\hat{y}_t-\bar{y}) \,. \end{aligned} \tag{2.58}$$

Further derivation yields

$$\begin{aligned}\sum_{t=1}^{T}(y_t-\bar{y})(\hat{y}_t-\bar{y}) &= \sum_{t=1}^{T}(y_t-\bar{y})b_1(x_t-\bar{x}) \\ &= b_1^2\sum_{t=1}^{T}(x_t-\bar{x})^2 \\ &= \sum_{t=1}^{T}(\hat{y}_t-\bar{y})^2 \ .\end{aligned}$$

Thus we have

$$\sum_{t=1}^{T}(y_t-\bar{y})^2 = \sum_{t=1}^{T}(y_t-\hat{y}_t)^2 + \sum_{t=1}^{T}(\hat{y}_t-\bar{y})^2 \ . \tag{2.59}$$

The left hand side of (2.59) is called the sum of squares about the mean or corrected sum of squares of y (*i.e.*, SS corrected) or SYY. The first term on right hand side describes the deviations: observation minus predicted value, *viz.*, the residual sum of squares, *i.e.*:

$$\text{SS Residual:} \ \ RSS = \sum_{t=1}^{T}(y_t-\hat{y}_t)^2 \ . \tag{2.60}$$

whereas the second term describes the proportion of variability explained by regression,

$$\text{SS Regression:} \ \ SS_{Reg} = \sum_{t=1}^{T}(\hat{y}_t-\bar{y})^2 \ . \tag{2.61}$$

If all observations y_t are located on a straight line, we have obviously $\sum_{t=1}^{T}(y_t-\hat{y}_t)^2 = 0$ and thus $SS_{corrected} = SS_{Reg}$.

Note that SS_{Reg} has only one degree of freedom as it is completely determined by b_1, SYY has $(T-1)$ degrees of freedom due to constraint $\sum_{t=1}^{T}(y_t-\bar{y}) = 0$ and SSE has $(T-2)$ degrees of freedom as it depends on b_0 and b_1.

All sums of squares are mutually independent and distributed as χ^2_{df} with *df* degrees of freedom if the errors are normally distributed.

The mean square due to regression is

$$MS_{Reg} = \frac{SS_{REG}}{1} \tag{2.62}$$

and mean square due to residuals is

$$MSE = \frac{RSS}{T-2} \ . \tag{2.63}$$

The test statistic for testing $H_0 : \beta_1 = 0$ is

$$F_0 = \frac{MS_{Reg}}{MSE} \; . \tag{2.64}$$

If $H_0 : \beta_1 = 0$ is true, then MS_{Reg} and MSE are independently distributed, and

$$F_0 \sim F_{1,T-2} \; .$$

The decision rule for $H_1 : \beta_1 \neq 0$ is to reject H_0 if

$$F_0 > F_{1,T-2;1-\alpha}$$

at $(1-\alpha)\%$ level of significance. The test procedure can be described in an Analysis of Variance table.

TABLE 2.1. Analysis of Variance for testing $H_0 : \beta_1 = 0$

Source of variation	Sum of squares	Degrees of freedom	Mean Square
Regression	SS_{Reg}	1	MS_{Reg}
Residual	SSE	$T-2$	MSE
Total	SYY	$T-1$	

The sample correlation coefficient then may be written as

$$r_{xy} = \frac{SXY}{\sqrt{SXX}\sqrt{SYY}} \; .$$

Moreover, we have

$$b_1 = \frac{SXY}{SXX} = r_{xy}\frac{SYY}{SXX} \; . \tag{2.65}$$

The estimator of σ^2 in this case may be expressed as

$$s^2 = \frac{1}{T-2}\sum \hat{\epsilon}_t^2 = \frac{1}{T-2}RSS \; . \tag{2.66}$$

Various alternative formulations for RSS are in use as well:

$$\begin{aligned} RSS &= \sum_{t=1}^{T}\left[y_t - (b_0 + b_1 x_t)\right]^2 \\ &= \sum_{t=1}^{T}\left[(y_t - \bar{y}) - b_1(x_t - \bar{x})\right]^2 \\ &= SYY + b_1^2 SXX - 2b_1 SXY \\ &= SYY - b_1^2 SXX \\ &= SYY - \frac{(SXY)^2}{SXX} \; . \end{aligned} \tag{2.67}$$

Further relations become immediately apparent:

$$SS_{corrected} = SYY$$

and

$$\begin{aligned} SS_{Reg} &= SYY - RSS \\ &= \frac{(SXY)^2}{SXX} \\ &= b_1^2 SXX \\ &= b_1 SXY \ . \end{aligned}$$

2.10 Goodness of Fit of Regression

It can be noted that a good fitted model is obtained when residuals are small. So a measure of quality of fitted model can be obtained with RSS. For the model (2.2), when intercept term is present in the model, a measure of goodness of fit is given by

$$\begin{aligned} R^2 &= 1 - \frac{RSS}{SYY} \\ &= \frac{SS_{Reg}}{SYY} . \end{aligned} \tag{2.68}$$

This is known as the coefficient of determination. The ratio SS_{Reg}/SYY describes the proportion of variability that is explained by regression in relation to the total variability of y. The ratio RSS/SYY describes the proportion of variability that is not covered by the regression.

It can be seen that

$$R^2 = r_{xy}^2 .$$

Clearly $0 \leq R^2 \leq 1$, so a value of R^2 closer to one indicates the better the fit and value of R^2 closer to zero indicates the poorer the fit.

It may be noted that when the regression line passes through origin, *i.e.*, $\beta_0 = 0$, then R^2 can not be used to judge the goodness of fitted model. In such a case, a possible measure of goodness of fit can be defined as

$$R_0^2 = 1 - \frac{\sum_{t=1}^{T} \hat{\epsilon}_t^2}{\sum_{t=1}^{T} y_t^2} \ . \tag{2.69}$$

The mean sum of squares due to residuals, SSE/df can also be used as a basis of comparison between the regression models with and without intercept terms.

2.11 Reverse Regression Method

The reverse (or inverse) regression approach minimizes the sum of squares of horizontal distances between the observed data points and the line in the scatter diagram (see Fig. 2.1(b)) to obtain the estimates of regression parameters. The reverse regression estimates $\hat{\beta}_{0R}$ of β_0 and $\hat{\beta}_{1R}$ of β_1 for the model (2.2) are obtained as

$$\hat{\beta}_{0R} = \bar{y} - \hat{\beta}_{1R}\bar{x} \tag{2.70}$$

and

$$\hat{\beta}_{1R} = \frac{SYY}{SXY}. \tag{2.71}$$

See Maddala (1992) for the derivation of (2.70) and (2.71). An important application of reverse regression method is in solving the calibration problem, see, e.g., Toutenburg and Shalabh (2001a), Shalabh and Toutenburg (2006).

2.12 Orthogonal Regression Method

Generally when uncertainties are involved in dependent and independent variables both, then orthogonal regression is more appropriate. The least squares principle in orthogonal regression minimizes the squared perpendicular distance between the observed data points and the line in the scatter diagram to obtain the estimates of regression coefficients. This is also known as major axis regression method. The estimates obtained are called as orthogonal regression estimates or major axis regression estimates of regression coefficients.

The squared perpendicular distance of observed data (x_t, y_t) $(t = 1, \ldots, T)$ from the line (see Fig. 2.1(c)) is

$$d_t^2 = (X_t - x_t)^2 + (Y_t - y_t)^2 \tag{2.72}$$

where (X_t, Y_t) denotes the t^{th} pair of observation without any error which lie on the line.

The objective is to minimize $\sum_{t=1}^{T} d_t^2$ to obtain the estimates of β_0 and β_1.

The observations (x_t, y_t) $(t = 1, \ldots, T)$ are expected to lie on the line

$$Y_t = \beta_0 + \beta_1 X_t$$

and define

$$\mathrm{E}_t = Y_t - \beta_0 - \beta_1 X_t = 0 \,. \tag{2.73}$$

The regression coefficients are to be obtained by minimizing (2.72) under the constraint (2.73) using the Lagrangian's multiplier method. The

Lagrangian function is

$$L_o = \sum_{t=1}^{T} d_t^2 - 2\sum_{t=1}^{T} \lambda_t \, \mathrm{E}_t \tag{2.74}$$

where $\lambda_1, \ldots, \lambda_T$ are the Lagrangian multipliers. The set of equations are obtained by setting

$$\frac{\partial L_o}{\partial X_t} = 0, \ \frac{\partial L_o}{\partial Y_t} = 0, \ \frac{\partial L_o}{\partial \beta_0} = 0 \ \text{ and } \ \frac{\partial L_o}{\partial \beta_1} = 0 \ \ (t = 1, \ldots, T) \ .$$

Thus

$$\frac{\partial L_o}{\partial X_t} = (X_t - x_t) + \lambda_t \beta_1 = 0 \tag{2.75}$$

$$\frac{\partial L_o}{\partial Y_t} = (Y_t - y_t) - \lambda_t = 0 \tag{2.76}$$

$$\frac{\partial L_o}{\partial \beta_0} = \sum_{t=1}^{T} \lambda_t = 0 \tag{2.77}$$

$$\frac{\partial L_o}{\partial \beta_1} = \sum_{t=1}^{T} \lambda_t X_t = 0 \ . \tag{2.78}$$

Since

$$\begin{aligned} X_t &= x_t - \lambda_t \beta_1 \quad (\text{cf. } (2.75)) \\ Y_t &= y_t + \lambda_t \quad (\text{cf. } (2.76)) \ , \end{aligned}$$

so

$$\mathrm{E}_t = (y_t + \lambda_t) - \beta_0 - \beta_1(x_t - \lambda_t \beta_1) = 0 \quad (\text{cf. } (2.73))$$

or

$$\lambda_t = \frac{\beta_0 + \beta_1 x_t - y_t}{1 + \beta_1^2} \ . \tag{2.79}$$

Also

$$\frac{\sum_{t=1}^{T} (\beta_0 + \beta_1 x_t - y_t)}{1 + \beta_1^2} = 0 \quad (\text{cf. } (2.77)) \tag{2.80}$$

and

$$\sum_{t=1}^{T} \lambda_t (x_t - \lambda_t \beta_1) = 0 \quad (\text{cf. } (2.75) \text{ and cf. } (2.78)).$$

Substituting λ_t from (2.79), we have

$$\frac{\sum_{t=1}^{T} (\beta_0 x_t + \beta_1 x_t^2 - y_t x_t)}{(1 + \beta_1^2)} - \frac{\beta_1 (\beta_0 + \beta_1 x_t - y_t)^2}{(1 + \beta_1^2)^2} = 0 \ . \tag{2.81}$$

Solving (2.80), we obtain an orthogonal regression estimate of β_0 as

$$\hat{\beta}_{0OR} = \bar{y} - \hat{\beta}_{1OR}\bar{x} \tag{2.82}$$

where $\hat{\beta}_{1OR}$ is an orthogonal regression estimate of β_1.

Now, using (2.82) in (2.81), we solve for β_1 as

$$\sum_{t=1}^{T}(1+\beta_1^2)\left[\bar{y}x_t - \beta_1\bar{x}x_t + \beta_1 x_t^2 - x_t y_t\right] - \beta_1\sum_{t=1}^{T}(\bar{y} - \beta_1\bar{x} + \beta_1 x_t - y_t)^2 = 0$$

or

$$(1+\beta_1^2)\sum_{t=1}^{T} x_t\left[y_t - \bar{y} - \beta_1(x_t - \bar{x})\right] + \beta_1\sum_{t=1}^{T}\left[-(y_t - \bar{y}) + \beta_1(x_t - \bar{x})\right]^2 = 0$$

or

$$(1+\beta_1^2)\sum_{t=1}^{T}(u_t + \bar{x})(v_t - \beta_1 u_t) + \beta_1\sum_{t=1}^{T}(-v_t + \beta_1 u_t)^2 = 0 \tag{2.83}$$

where $u_t = x_t - \bar{x}$ and $v_t = y_t - \bar{y}$.

Since $\sum_{t=1}^{T} u_t = \sum_{t=1}^{T} v_t = 0$, so (2.83) reduces to

$$\sum_{t=1}^{T}\left[\beta_1^2 u_t v_t + \beta_1(u_t^2 - v_t^2) - u_t v_t\right] = 0$$

or

$$\beta_1^2 SXY + \beta_1(SXX - SYY) - SXY = 0\ . \tag{2.84}$$

Solving (2.84) gives the orthogonal regression estimate of β_1 as

$$\hat{\beta}_{1OR} = \frac{(SYY - SXX) + \operatorname{sgn}(SXY)\sqrt{(SXX - SYY)^2 + 4SXY}}{2SXY} \tag{2.85}$$

where $\operatorname{sgn}(SXY)$ is the sign of SXY. Notice that (2.85) gives two solutions for $\hat{\beta}_{1OR}$. We choose the solution which minimizes $\sum_{t=1}^{T} d_t^2$. The other solution maximizes $\sum_{t=1}^{T} d_t^2$ and is in the direction perpendicular to the optimal solution.

The optimal solution can be chosen with the sign of SXY. If $SXY > 0$, then $\operatorname{sgn}(SXY) = 1$ and if $SXY < 0$, then $\operatorname{sgn}(SXY) = -1$.

2.13 Reduced Major Axis Regression Method

The least squares approaches generally minimize distances between the observed data points and the line in the scatter diagram. Alternatively, instead of distances, the area of rectangles defined between corresponding observed data points and nearest point on the line in the scatter diagram can also be minimized. Such an approach is more appropriate when the uncertainties are present in study and explanatory variables both and is called as reduced major axis regression.

The area of rectangle extended between the t^{th} observed data point and the line (see Fig. 2.1(d)) is

$$A_t = (X_t \sim x_t)(Y_t \sim y_t) \quad (t = 1, \ldots, T) \ .$$

where (X_t, Y_t) denotes the t^{th} pair of observation without any error which lie on the line.

The total area extended by T data points is

$$\sum_{t=1}^{T} A_t = \sum_{t=1}^{T} (X_t \sim x_t)(Y_t \sim y_t) \ . \tag{2.86}$$

All observed data points (x_t, y_t) $(t = 1, \ldots, T)$ are expected to lie on the line

$$Y_t = \beta_0 + \beta_1 X_t$$

and define

$$\mathrm{E}_t^* = Y_t - \beta_0 + \beta_1 X_t = 0.$$

So the sum of areas in (2.86) is to be minimized under the constraints E_t^* to obtain the reduced major axis estimates of regression coefficients. Using the Lagrangian multipliers method, the Lagrangian function is

$$\begin{aligned} L_R &= \sum_{t=1}^{T} A_t - \sum_{t=1}^{T} \mu_t \, \mathrm{E}_t^* \\ &= \sum_{t=1}^{T} (X_t - x_t)(Y_t - y_t) - \sum_{t=1}^{T} \mu_t \, \mathrm{E}_t^* \end{aligned} \tag{2.87}$$

where $\mu_1, \ldots, \mu_T$ are the Lagrangian multipliers. The set of equations are obtained by setting

$$\frac{\partial L_R}{\partial X_t} = 0, \ \frac{\partial L_R}{\partial Y_t} = 0, \ \frac{\partial L_R}{\partial \beta_0} = 0, \ \frac{\partial L_R}{\partial \beta_1} = 0 \quad (t = 1, \ldots, T) \ . \tag{2.88}$$

Thus

$$\frac{\partial L_R}{\partial X_t} = (Y_t - y_t) + \beta_1 \mu_t = 0 \tag{2.89}$$

$$\frac{\partial L_R}{\partial Y_t} = (X_t - x_t) + \mu_t = 0 \tag{2.90}$$

$$\frac{\partial L_R}{\partial \beta_0} = \sum_{t=1}^{T} \mu_t = 0 \tag{2.91}$$

$$\frac{\partial L_R}{\partial \beta_1} = \sum_{t=1}^{T} \mu_t X_t = 0 \; . \tag{2.92}$$

Now

$$\begin{aligned} X_t &= x_t + \mu_t \quad \text{(cf. (2.90))} \\ Y_t &= y_t - \beta_1 \mu_t \quad \text{(cf. (2.89))} \end{aligned} \tag{2.93}$$

$$\begin{aligned} \beta_0 + \beta_1 X_t &= y_t - \beta_1 \mu_t \\ \beta_0 + \beta_1 (x_t + \mu_t) &= y_t - \beta_1 \mu_t \quad \text{(cf. (2.93))} \end{aligned} \tag{2.94}$$

$$\mu_t = \frac{y_t - \beta_0 - \beta_1 x_t}{2\beta_1} \quad \text{(cf. (2.94))} \; . \tag{2.95}$$

Substituting μ_t in (2.91), we get the reduced major axis regression estimate of β_0 as

$$\hat{\beta}_{0RM} = \bar{y} - \hat{\beta}_{1RM} \bar{x} \tag{2.96}$$

where $\hat{\beta}_{1RM}$ is the reduced major axis regression estimate of β_1. Using (2.90), (2.95) and (2.96) in (2.92), we get

$$\sum_{t=1}^{T} \left(\frac{y_t - \bar{y} + \beta_1 \bar{x} - \beta_1 x_t}{2\beta_1} \right) \left(x_t - \frac{y_t - \bar{y} + \beta_1 \bar{x} - \beta_1 x_t}{2\beta_1} \right) = 0 \; . \tag{2.97}$$

Let $u_t = x_t - \bar{x}$ and $v_t = y_t - \bar{y}$, then (2.97) becomes

$$\sum_{t=1}^{T} (v_t - \beta_1 u_t)(v_t + \beta_1 u_t + 2\beta_1 \bar{x}) = 0 \; . \tag{2.98}$$

Using $\sum_{t=1}^{T} u_t = \sum_{t=1}^{T} v_t = 0$ in (2.98), we get

$$\sum_{t=1}^{T} v_t^2 - \beta_1^2 \sum_{t=1}^{T} u_t^2 = 0$$

which gives the reduced major axis regression estimate of β_1 as

$$\hat{\beta}_{1RM} = \operatorname{sgn}(SXY) \sqrt{\frac{SYY}{SXX}} \tag{2.99}$$

where $\text{sgn}(SXY) = 1$ if $SXY > 0$ and $\text{sgn}(SXY) = -1$ if $SXY < 0$. Note that (2.63) gives two solutions for $\hat{\beta}_{1\,RM}$. We choose the one which has same sign as of SXY.

2.14 Least Absolute Deviation Regression Method

In the method of least squares, the estimates of the parameters β_0 and β_1 in the model (2.2) are chosen such that the sum of squares of deviations $\sum_{t=1}^{T} e_t^2$ is minimum. In the method of least absolute deviation (LAD) regression, the parameters β_0 and β_1 in model (2.2) are estimated such that the sum of absolute deviations $\sum_{t=1}^{T} |e_t|$ is minimum. The LAD estimates $\hat{\beta}_{0L}$ and $\hat{\beta}_{1L}$ are the values β_0 and β_1, respectively which minimize

$$LAD(\beta_0, \beta_1) = \sum_{t=1}^{T} |y_t - \beta_0 - \beta_1 x_t|$$

for the given observations (x_t, y_t) $(t = 1, \ldots, T)$.

Conceptually, LAD procedure is simpler than OLS procedure because $|\hat{\epsilon}|$ (absolute residuals) is a more straightforward measure of the size of the residual than $\hat{\epsilon}^2$ (squared residuals). But no closed form of the LAD regression estimates of β_0 and β_1 is available. Only algorithm based LAD estimates can be obtained numerically. The non-uniqueness and degeneracy concepts are used in algorithms to judge the quality of the estimates. The concept of non-uniqueness relates to that more than one best lines pass through a data point. The degeneracy concept describes that the best line through a data point also passes through more than one other data points. The algorithm for finding the estimators generally proceeds in steps. At each step, the best line is found that passes through a given data point. The best line always passes through another data point, and this data point is used in the next step. When there is nonuniqueness, then there is more than one best line. When there is degeneracy, then the best line passes through more than one other data point. When either of the problem is present, then there is more than one choice for the data point to be used in the next step and the algorithm may go around in circles or make a wrong choice of the LAD regression line. The exact tests of hypothesis and confidence intervals for the LAD regression estimates can not be derived analytically. Instead they are derived analogous to the tests of hypothesis and confidence intervals related to ordinary least squares estimates. The LAD regression is discussed in Section 9.2.

More details about the theory and computations for LAD regression can be found in Bloomfield and Steiger (1983), Birkes and Dodge (1993), Dodge (1987a; 1987b; 1992; 2002).

2.15 Estimation of Parameters when X Is Stochastic

Suppose both dependent and independent variables are stochastic in the simple linear regression model

$$y = \beta_0 + \beta_1 X + \epsilon \tag{2.100}$$

where the deviations are observed as random variable ϵ. The observations (x_t, y_t), $t = 1, \ldots, T$ are assumed to be jointly distributed. Then the statistical inferences, which are conditional on X, can be drawn in such cases.

Assume the joint distribution of X and y to be bivariate normal $N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ where μ_x and μ_y are the means of X and y; σ_x^2 and σ_y^2 are the variances of X and y; and ρ is the correlation coefficient between X and y. Then the conditional distribution of y given $X = x$ is univariate normal with conditional mean

$$\mathrm{E}(y|X = x) = \beta_0 + \beta_1 x$$

and conditional variance

$$\mathrm{var}(y|X = x) = \sigma_{y \cdot x}^2 = \sigma_y^2(1 - \rho^2)$$

where

$$\beta_0 \quad = \quad \mu_y - \mu_x \beta_1 \tag{2.101}$$

and

$$\beta_1 \quad = \quad \frac{\sigma_y}{\sigma_x}\rho\ . \tag{2.102}$$

Thus the problem of estimation of parameters, when both X and y are stochastic, can be reformulated as the problem of estimation of parameters when T observations on conditional random variable $y|X = x$ are obtained as $y_t|x_t$ $(t = 1, \ldots, T)$ which are independent and normally distributed with mean $(\beta_0 + \beta_1 x_t)$ and variance $\sigma_{y \cdot x}^2$ with nonstochastic X.

The method of maximum likelihood yields the same estimates of β_0 and β_1 as earlier in the case of nonstochastic X as

$$\tilde{b}_0 = \bar{y} - \tilde{b}_1 \bar{x} \tag{2.103}$$

and

$$\tilde{b}_1 = \frac{SXY}{SXX}\ , \tag{2.104}$$

respectively.

Moreover, the correlation coefficient

$$\rho = \frac{\mathrm{E}(y - \mu_y)(X - \mu_x)}{\sigma_y \sigma_x} \tag{2.105}$$

can be estimated by the sample correlation coefficient

$$\begin{aligned}
\hat{\rho} &= \frac{\sum_{t=1}^{T}(y_t-\bar{y})(x_t-\bar{x})}{\sqrt{\sum_{t=1}^{T}(x_t-\bar{x})^2}\sqrt{\sum_{t=1}^{T}(y_t-\bar{y})^2}} \\
&= \frac{SXY}{\sqrt{SXX}\sqrt{SYY}} \\
&= \tilde{b}_1\sqrt{\frac{SXX}{SYY}}. \qquad (2.106)
\end{aligned}$$

Thus

$$\begin{aligned}
\hat{\rho}^2 &= \tilde{b}_1^2\,\frac{SXX}{SYY} \\
&= \tilde{b}_1\frac{SXY}{SYY} \\
&= \frac{SYY-\sum_{t=1}^{T}\hat{\epsilon}_t^2}{SYY} \\
&= R^2 \qquad (2.107)
\end{aligned}$$

which is same as the coefficient of determination mentioned as in Section 2.10. Thus R^2 has the same expression as in the case when X is fixed. Thus R^2 again measures the goodness of fitted model even when X is stochastic.