

# Models for Time Series Data

**Professor V.M. Auken, PhD**



## Standardized regression coefficients

Usually it is difficult to compare the regression coefficients because the magnitude of  $\hat{\beta}_j$  reflects the units of measurement of  $j^{\text{th}}$  explanatory variable  $X_j$ . For example, in the following fitted regression model

$$\hat{y} = 5 + X_1 + 1000X_2,$$

$y$  is measured in liters,  $X_1$  is milliliters and  $X_2$  in liters. Although  $\hat{\beta}_2 \gg \hat{\beta}_1$  but effect of both explanatory variables is identical. One liter change in either  $X_1$  and  $X_2$  when other variable is held fixed produces the same change in  $\hat{y}$ .

Sometimes it is helpful to work with scaled explanatory and study variables that produces dimensionless regression coefficients.

These dimensionless regression coefficients are called as **standardized regression coefficients**.

There are two popular approaches for scaling which gives standardized regression coefficients.

We discuss them as follows:

## 1. Unit normal scaling

Employ unit normal scaling to each explanatory variable and study variable .

So define

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, k$$
$$y_i^* = \frac{y_i - \bar{y}}{s_y}$$

where

$$s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \quad \text{and} \quad s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

are the sample variances of  $j^{\text{th}}$  explanatory variable and study variable, respectively.

All scaled explanatory variables and the scaled study variable have mean zero and sample variance unity, i.e., using these new variables, the regression model becomes

$$y_i^* = \gamma_1 z_{i1} + \gamma_2 z_{i2} + \dots + \gamma_k z_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

Such centering removes the intercept term from the model. The least squares estimate of  $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_k)'$  is

$$\hat{\gamma} = (Z'Z)^{-1} Z' y^*.$$

This scaling has a similarity to standardizing a normal random variable, i.e., observation minus its mean and divided by its standard deviation. So it is called as a unit normal scaling.

## 2. Unit length scaling

In unit length scaling, define

$$\omega_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j^{1/2}}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, k$$

$$y_i^0 = \frac{y_i - \bar{y}}{SS_T^{1/2}}$$

where

$$S_j = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$$

is the corrected sum of squares for  $j^{\text{th}}$  explanatory variables  $X_j$  and

$$S_T = SS_T = \sum_{i=1}^n (y_i - \bar{y})^2$$

is the total sum of squares.

In this scaling, each new explanatory variable  $W_j$  has mean  $\bar{\omega}_j = \frac{1}{n} \sum_{i=1}^n \omega_{ij} = 0$  and length

$$\sqrt{\sum_{i=1}^n (\omega_{ij} - \bar{\omega}_j)^2} = 1.$$

In terms of these variables, the regression model is

$$y_i^o = \delta_1 \omega_{i1} + \delta_2 \omega_{i2} + \dots + \delta_k \omega_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

The least squares estimate of regression coefficient  $\delta = (\delta_1, \delta_2, \dots, \delta_k)'$  is

$$\hat{\delta} = (W'W)^{-1}W'y^o.$$

In such a case, the matrix  $W'W$  is in the form of correlation matrix, i.e.,

$$W'W = \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1k} \\ r_{12} & 1 & r_{23} & \cdots & r_{2k} \\ r_{13} & r_{23} & 1 & \cdots & r_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{1k} & r_{2k} & r_{3k} & \cdots & 1 \end{pmatrix}$$

where

$$\begin{aligned} r_{ij} &= \frac{\sum_{u=1}^n (x_{ui} - \bar{x}_i)(x_{uj} - \bar{x}_j)}{(S_u S_j)^{1/2}} \\ &= \frac{S_{ij}}{(S_u S_j)^{1/2}} \end{aligned}$$

is the simple correlation coefficient between the explanatory variables  $X_i$  and  $X_j$ .

Similarly,

$$W' y^o = (r_{1y}, r_{2y}, \dots, r_{ky})'$$

where

$$r_{jy} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(y_i - \bar{y})}{(S_{jj}SS_T)^{1/2}} = \frac{S_{jy}}{(S_{jj}SS_T)^{1/2}}$$

is the simple correlation coefficient between  $j^{\text{th}}$  explanatory variable  $X_j$  and study variable  $y$ .

Note that it is customary to refer  $r_{ij}$  and  $r_{jy}$  as correlation coefficient though  $X_j$ 's are not random variable.

If unit normal scaling is used, then

$$Z'Z = (n-1)W'W.$$

So the estimates of regression coefficient in unit normal scaling (i.e.,  $\hat{\gamma}$ ), and unit length scaling (i.e.,  $\hat{\delta}$ ), are identical.

So it does not matter which scaling is used, so

$$\hat{\gamma} = \hat{\delta}.$$

The regression coefficients obtained after such scaling, viz.,  $\hat{\gamma}$  or  $\hat{\delta}$  are usually called standardized regression coefficients.

The relationship between the original and standardized regression coefficients is

$$b_j = \hat{\delta}_j \left( \frac{SS_T}{S_{jj}} \right)^{1/2}, \quad j = 1, 2, \dots, k$$

and

$$b_0 = \bar{y} - \sum_{j=1}^k b_j \bar{x}_j$$

where  $b_0$  is the OLSE of intercept term and  $b_j$  are the OLSE of slope parameters  $\beta_j$ .

## The model in deviation form

The multiple linear regression model can also be expressed in the deviation form.

First all the data is expressed in terms of deviations from sample mean.

The estimation of regression parameters is performed in two steps:

**First step:** Estimate the slope parameters.

**Second step :** Estimate the intercept term.

The multiple linear regression model in deviation form is expressed as follows:

Let

$$A = I - \frac{1}{n} \ell \ell'$$

where  $\ell = (1, 1, \dots, 1)'$  is a  $n \times 1$  vector of each element unity.



So

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} - \frac{1}{n} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Then

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} (1, 1, \dots, 1) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$= \frac{1}{n} \ell' y$$

$$Ay = y - \ell \bar{y} = (y_1 - \bar{y}, y_2 - \bar{y}, \dots, y_n - \bar{y})'.$$

Thus pre-multiplication of any column vector by  $A$  produces a vector showing those observations in deviation form.

Note that

$$\begin{aligned} A\ell &= \ell - \frac{1}{n}\ell\ell'\ell \\ &= \ell - \frac{1}{n}\ell.n \\ &= \ell - \ell \\ &= 0 \end{aligned}$$

and  $A$  is symmetric and idempotent matrix.

In the model

$$y = X\beta + \varepsilon,$$

the OLSE of  $\beta$  is

$$b = (X'X)^{-1} X'y$$

and residual vector is

$$e = y - Xb.$$

Note that  $Ae = e$ .

Let the  $n \times k$  matrix  $X$  is partitioned as

$$X = [X_1 \quad X_2^*]$$

where  $X_1 = (1, 1, \dots, 1)'$  is  $n \times 1$  vector with all elements unity representing the intercept term,  $X_2^*$  is  $n \times (k-1)$  matrix of observations on  $(k-1)$  explanatory variables  $X_2, X_3, \dots, X_k$  and OLSE  $b = (b_1, b_2^*)'$  is suitably partitioned with OLSE of intercept term  $\beta_1$  as  $b_1$  and  $b_2^*$  as a  $(k-1) \times 1$  vector of OLSEs associated with  $\beta_2, \beta_3, \dots, \beta_k$ .

Then

$$y = X_1 b_1 + X_2^* b_2^* + e.$$

Premultiplication by  $A$  gives

$$\begin{aligned} Ay &= AX_1 b_1 + AX_2^* b_2^* + Ae \\ &= AX_2^* b_2^* + e. \end{aligned}$$

Further, premultiplication by  $X_2^{*'}$  gives

$$\begin{aligned} X_2^{*'} Ay &= X_2^{*'} AX_2^* b_2^* + X_2^{*'} e \\ &= X_2^{*'} AX_2^* b_2^*. \end{aligned}$$



VectorStock®

VectorStock.com/25671800

Since  $A$  is symmetric and idempotent, so

$$(AX_2^*)'(Ay) = (AX_2^*)'(AX_2^*)b_2^* .$$

This equation can be compared with the normal equations  $X'y = X'Xb$  in the model  $y = X\beta + \varepsilon$ . Such a comparison yields the following conclusions:

- $b_2^*$  is the sub-vector of OLSE.
- $Ay$  is the study variables vector in deviation form.
- $AX_2^*$  is the explanatory variable matrix in deviation form.
- This is normal equation in terms of deviations. Its solution gives OLS of slope coefficients as

$$b_2^* = \left[ (AX_2^*)'(AX_2^*) \right]^{-1} (AX_2^*)'(Ay).$$

The estimate of intercept term is obtained in the second step as follows:

Premultiplying  $y = Xb + e$  by  $\frac{1}{n}\ell'$  gives

$$\frac{1}{n}\ell'y = \frac{1}{n}\ell'Xb + \frac{1}{n}\ell'e$$

$$\bar{y} = [1 \ \bar{X}_2 \ \bar{X}_3 \ \dots \ \bar{X}_k] \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix} + 0$$

$$\Rightarrow b_1 = \bar{y} - b_2\bar{X}_2 - b_3\bar{X}_3 - \dots - b_k\bar{X}_k.$$

Now we explain the various sums of squares in terms of this model.

The total sum of squares is

$$TSS = y' Ay.$$

Since  $Ay = AX_2^*b_2^* + e$

$$\begin{aligned}y' Ay &= y' AX_2^*b_2^* + y' e \\&= (Xb + e)' AX_2^*b_2^* + y' e \\&= (X_1b_1 + X_2^*b_2^* + e)' AX_2^*b_2^* + (X_1b_1 + X_2^*b_2^* + e)' e \\&= b_2^{*'} X_2^{*'} AX_2^*b_2^* + e' e,\end{aligned}$$

so

$$TSS = SS_{reg} + SS_{res}$$

where the **sum of squares due to regression** is

$$SS_{reg} = b_2^{*'} X_2^{*'} AX_2^*b_2^*$$

and the **sum of squares due to residual** is

$$SS_{res} = e' e.$$

# QUESTIONS!

