Math & Stat for MBA
Problem Set #2

Student ID :
Sungchan Park

Answer 4-1 There are explanatory variables, $x_1, \cdots, x_k$
Even though all the observations on x_i is multipled by λ,
other variables are not affected because each variable is independent.

Assump that x_1 is multipled by λ.
Other variables are not changed, so we can simply think them as constant.

$$\Rightarrow y = \beta_1 + \beta_2 X + \varepsilon$$

$$\hat{\beta} = \frac{1}{T\sum x_t^2 - (\sum x_t)^2} \begin{bmatrix} \sum x_t^2 \cdot \sum y_t - \sum x_t \sum x_t y_t \\ T\sum x_t y_t - \sum x_t \sum y_t \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

$\times \lambda$

$$\hat{\beta}' = \frac{1}{\lambda^2} \times \frac{1}{T\sum x_t^2 - (\sum x_t)^2} \begin{bmatrix} \lambda^2 (\sum x_t^2 \cdot \sum y_t - \sum x_t \sum x_t y_t) \\ \lambda (T\sum x_t y_t - \sum x_t \sum y_t) \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1' \\ \hat{\beta}_2' \end{bmatrix}$$

$\hat{\beta}_1' = \hat{\beta}_1$, $\hat{\beta}_2' = \frac{1}{\lambda} \cdot \hat{\beta}_2$ $\Rightarrow$ There is not any change on constant.
But, the corresponding regression coefficient is multipled by $1/\lambda$

From the point of view, if a particular explanatory variable is measured in thousands of kgs instead of millions of kgs, the corresponding coefficient is multipled by 1/1000 and other variables is not changed.

Answer 4-2 i) $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$, replace X_1 by $X_1 + \lambda$

$$Y = \beta_0 + \beta_1(X_1+\lambda) + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$
$$= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon + \beta_1\lambda, \quad \varepsilon' = \varepsilon + \beta_1 \lambda$$
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon'$$

There is the translation of X_1 axies by λ.
So, we add $\beta_1\lambda$ to ε to explain this change.

ii) $Y = A \cdot P^{\alpha} \cdot L^{\beta} \cdot M^{\gamma}$
$$\log Y = \log A + \alpha \log P + \beta \log L + \gamma \log M$$

Assump that P is measure in thousands of dollars instead of millions of dollars.
Now then, P is multipled by 1000.

$$\log Y = \log A + \alpha \log 1000P + \beta \log L + \gamma \log M$$
$$= \log A + \alpha \cdot \log 1000 + \alpha \log P + \beta \log L + \gamma \log M$$
$$= \log A' + \alpha \log P + \beta \log L + \gamma \log M$$

There is not any change in the corresponding coefficient.
We can say that, the corresponding coefficient is independent of the units in which the variable is measured.

Answer 5-1 There are two regression models

Model 1: $Y = \beta_0 + \beta_1 X + \varepsilon'$
Model 2: $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon$

$$RSS = \sum (Y_i - \hat{Y})^2 = \sum \hat{\varepsilon}_i^2$$

RSS for Model 1 : $\sum \hat{\varepsilon}'^2 = \sum (\hat{\beta}_2 X^2 + \hat{\beta}_3 X^3 + \hat{\varepsilon})^2$
RSS for Model 2 : $\sum \hat{\varepsilon}^2$

Most of cases, the RSS of Complicated model is lower than that of Simple model. In Simple model, we restrict the corresponding coefficient of other Complicated Variables as 0. With restriction, It is nearly impossible to find the minimum value of $\sum \hat{\varepsilon}^2$

In this case, we already know that the true relationship between X and Y is linear, $\hat{\beta}_2 = \hat{\beta}_3 = 0$. So, we would expect RSS for them to be the same.

$$\sum \hat{\varepsilon}'^2 \approx \sum \hat{\varepsilon}^2$$

Answer 5-2

Where X and Y is non-linear, $\hat{\beta}_2, \hat{\beta}_3$ might not be 0. we would expect RSS for Model 1 is higher than other.

$$\sum \hat{\varepsilon}'^2 = \sum (\hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\varepsilon})^2 \geq \sum \hat{\varepsilon}^2$$