Simple Regression (Appendix)

Recall that

$$b_2 = \frac{\sum_{t=1}^{T}(X_t - \overline{X})(Y_t - \overline{Y})}{\sum_{t=1}^{T}(X_t - \overline{X})^2} = \frac{\sum_{t=1}^{T}(X_t - \overline{X})Y_t}{\sum_{t=1}^{T}(X_t - \overline{X})^2} \qquad \text{(A.1)}$$

since $\sum_{t=1}^{T}(X_t - \overline{X}) = 0$. Let

$$w_t = \frac{(X_t - \overline{X})}{\sum_{t=1}^{T}(X_t - \overline{X})^2}. \qquad \text{(A.2)}$$

Each w_t is a constant, since *Xs* are fixed. Substituting into the equation (A.1), we have

$$b_2 = \sum_{t=1}^{T} w_t Y_t \qquad \text{(A.3)}$$

which expresses the estimated parameter as a weighted sum of the observations on the dependent variables. It is obvious that $\sum_{t=1}^{T} w_t = 0$. According to the definition of w_t,

$$\begin{aligned}\sum_{t=1}^{T} w_t X_t &\underbrace{=}_{\text{since } \sum w_t \overline{X} = 0} \sum_{t=1}^{T} w_t X_t + \sum_{t=1}^{T} w_t \overline{X} = \sum_{t=1}^{T} w_t (X_t - \overline{X}) \\ &= \sum_{t=1}^{T}\left(\frac{X_t - \overline{X}}{\sum_{t=1}^{T}(X_t - \overline{X})^2}\right)(X_t - \overline{X}) = \frac{\sum_{t=1}^{T}(X_t - \overline{X})^2}{\sum_{t=1}^{T}(X_t - \overline{X})^2}. \\ &= 1\end{aligned}$$

Result 1: $E(b_2) = \beta_2$.

Proof: Since $Y_t = \beta_1 + \beta_2 X_t + e_t$,

$$\begin{aligned} b_2 &= \sum_{t=1}^{T} w_t(\beta_1 + \beta_2 X_t + e_t) \\ &= \beta_2 + \sum_{t=1}^{T} w_t e_t \end{aligned} \qquad \text{(A.4)}$$

from the facts that $\sum_{t=1}^{T} w_t = 0$ and $\sum_{t=1}^{T} w_t X_t = 1$. Therefore,

$$E(b_2) = \beta_2 + \sum_{t=1}^{T} w_t E(e_t) = \beta_2$$

since $E(e_t) = 0$. ■

Result 2: $Var(b_2) = \dfrac{\sigma^2}{\sum_{t=1}^{T}(X_t - \overline{X})}$.

Proof: $Var(b_2) = Cov(b_2, b_2) = Cov\left(\beta_2 + \sum_{j=1}^{T} w_j e_j,\ \beta_2 + \sum_{t=1}^{T} w_t e_t\right)$

$$= Cov\left(\sum_{j=1}^{T} w_j e_j,\ \sum_{t=1}^{T} w_t e_t\right) \qquad \text{since } \beta_2 \text{ is constant}$$

$$= \sum_{j=1}^{T}\sum_{t=1}^{T} w_j w_t Cov\left(e_j,\ e_t\right)$$

$$= \sum_{t=1}^{T} w_t^2 Cov(e_t, e_t) \qquad \text{since } Cov\left(e_j, e_t\right) = 0 \text{ for } j \neq t$$

$$= \sum_{t=1}^{T} w_t^2 Var\left(e_t\right) = \sum_{t=1}^{T} w_t^2 \sigma^2 = \sigma^2 \frac{\sum_{t=1}^{T}\left(X - \overline{X}\right)^2}{\left(\sum_{t=}^{T}\left(X_t - \overline{X}\right)^2\right)^2}$$

$$= \frac{\sigma^2}{\sum_{t=}^{T}\left(X_t - \overline{X}\right)^2}. \quad \blacksquare$$

Result 3: $E(b_1) = \beta_1$.

Proof: $b_1 = \overline{Y} - b_2\overline{X} = \left(\beta_1 + \beta_2\overline{X} + \sum_{t=1}^{T}\frac{1}{T}e_t\right) - b_2\overline{X} = \beta_1 - (b_2 - \beta_2)\overline{X} + \sum_{t=1}^{T}\frac{1}{T}e_t$.

Hence, $E\left(b_1\right) = \beta_1 - \left(E\left(b_2\right) - \beta_2\right) + \sum_{t=1}^{T}\frac{1}{T}E\left(e_t\right) = \beta_1$ since $E(b_2) = \beta_2$ and $E\left(e_t\right) = 0$. $\blacksquare$

Result 4: $Cov(\overline{Y}, b_2) = 0$.

Proof: $Cov(\overline{Y}, b_2) = Cov(\sum_{j=1}^{T}\frac{1}{T}Y_j,\ \sum_{t=1}^{T} w_t Y_t) = \sum_{j=1}^{T}\sum_{t=1}^{T}\frac{1}{T}w_t Cov(Y_j, Y_t)$

$$= \sum_{t=1}^{T}\frac{1}{T} w_t Cov\left(Y_t,\ Y_t\right) \qquad \text{since } Cov\left(Y_j, Y_t\right) = 0 \text{ for } j \neq t$$

$$= \sum_{t=1}^{T}\frac{1}{T} w_t Var\left(Y_t\right) = \sum_{t=1}^{T}\frac{1}{T} w_t \sigma^2 = 0 \qquad \text{since } \sum_{t=1}^{T} w_t = 0. \quad \blacksquare$$

Result 5: $Var(b_1) = \dfrac{\sigma^2 \sum_{t=1}^{T} X_t^2}{T\sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2}$.

Proof: $Var\left(b_1\right) = Cov\left(b_1,\ b_1\right) = Cov\left(\overline{Y} - b_2\overline{X},\ \overline{Y} - b_2\overline{X}\right)$

$$= Cov\left(\overline{Y},\ \overline{Y}\right) + Cov\left(-b_2\overline{X},\ -b_2\overline{X}\right) + Cov\left(\overline{Y},\ -b_2\overline{X}\right) + Cov\left(-b_2\overline{X},\ \overline{Y}\right)$$

$$= Cov\left(\overline{Y},\ \overline{Y}\right) + \overline{X}^2 Cov\left(b_2,\ b_2\right) - \overline{X}Cov\left(\overline{Y},\ b_2\right) - \overline{X}Cov\left(b_2,\ \overline{Y}\right)$$

$$= Var\left(\overline{Y}\right) + \overline{X}^2 Var\left(b_2\right) \qquad \text{since } Cov(\overline{Y}, b_2) = 0$$

$$= \frac{\sigma^2}{T} + \frac{\sigma^2 \overline{X}^2}{\sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2} = \frac{\sigma^2 \sum_{t=1}^{T} X_t^2}{T\sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2}. \quad \blacksquare$$

Result 6: $Cov(b_1,\ b_2) = \dfrac{-\sigma^2 \overline{X}}{\sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2}$.

Proof: $Cov(b_1,\ b_2) = Cov(\overline{Y} - b_2\overline{X},\ b_2) = Cov(\overline{Y},\ b_2) - \overline{X}\ Cov(b_2,\ b_2) = -\overline{X}\ Var(b_2)$

$$= \frac{-\sigma^2 \overline{X}}{\sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2}. \quad \blacksquare$$

The Gauss-Markov Theorem (The OLS estimators are BLUEs)

Proof: We only prove the slope term b_2. It is also true for the intercept term b_1. Consider a general **<u>linear</u>** combination of the *Ys* that takes the form $\tilde{\beta}_2 = \sum_{t=1}^{T} d_t Y_t$, where d_t is nonrandom. The best linear unbiased estimator (BLUE) has the two properties: (1) $\tilde{\beta}_2$ is unbiased and (2) $Var(\tilde{\beta}_2)$ is the smallest within the class of linear and unbiased estimators. Define $a_t = d_t - w_t$. We have

$$\tilde{\beta}_2 = \sum_{t=1}^{T}(w_t + a_t)Y_t = \sum_{t=1}^{T} w_t Y_t + \sum_{t=1}^{T} a_t Y_t = b_2 + \sum_{t=1}^{T} a_t(\beta_1 + \beta_2 X_t + e_t)$$

$$= b_2 + \beta_1 \sum_{t=1}^{T} a_t + \beta_2 \sum_{t=1}^{T} a_t X_t + \sum_{t=1}^{T} a_t e_t$$

$$E(\tilde{\beta}_2) = E(b_2) + \beta_1 \sum_{t=1}^{T} a_t + \beta_2 \sum_{t=1}^{T} a_t X_t + \sum_{t=1}^{T} a_t E(e_t)$$

$$= \beta_2 + \beta_1 \sum_{t=1}^{T} a_t + \beta_2 \sum_{t=1}^{T} a_t X_t$$

For $\tilde{\beta}_2$ to be unbiased, we need this to be β_2, which can happen if and only if

$$\sum_{t=1}^{T} a_t = 0 \qquad \text{and} \qquad \sum_{t=1}^{T} a_t X_t = 0$$

$$Var(\tilde{\beta}_2) = Cov(\tilde{\beta}_2,\ \tilde{\beta}_2) = Cov\left(\sum_{j=1}^{T}(w_j + a_j)Y_j,\ \sum_{t=1}^{T}(w_t + a_t)Y_t\right)$$

$$= Cov\left(\sum_{j=1}^{T} w_j Y_j,\ \sum_{t=1}^{T} w_t Y_t\right) + Cov\left(\sum_{j=1}^{T} a_j Y_j,\ \sum_{t=1}^{T} a_t Y_t\right) + 2Cov\left(\sum_{j=1}^{T} w_j Y_j,\ \sum_{t=1}^{T} a_t Y_t\right)$$

$$= Cov(b_2,\ b_2) + \sum_{t=1}^{T} a_t^2 Var(Y_t) + 2\sum_{t=1}^{T} w_t a_t Var(Y_t) \qquad \text{since } Cov(Y_j, Y_t) = 0 \text{ for } j \neq t$$

$$= Var(b_2) + \sum_{t=1}^{T} a_t^2 \sigma^2 + 2\sum_{t=1}^{T} w_t a_t \sigma^2$$

The third term is zero since $\sum_{t=1}^{T} w_t a_t = \frac{\sum_{t=1}^{T}(X_t - \overline{X})a_t}{\sum_{t=1}^{T}(X_t - \overline{X})^2} = \frac{\sum_{t=1}^{T} X_t a_t - \overline{X}\sum_{t=1}^{T} a_t}{\sum_{t=1}^{T}(X_t - \overline{X})^2} = 0$.

Because $\sum_{t=1}^{T} a_t^2 \sigma^2 \geq 0$, we have proved the Gauss-Markov theorem. That is,

$$Var(\tilde{\beta}_2) \geq Var(b_2). \quad \blacksquare$$

Result 7: $\sum_{t=1}^{T} \hat{e}_t = 0$.

Proof: $\sum_{t=1}^{T} \hat{e}_t = \sum_{t=1}^{T}(Y_t - \hat{Y}_t) = \sum_{t=1}^{T}(Y_t - b_1 - b_2 X_t) = \sum_{t=1}^{T} Y_t - \sum_{t=1}^{T} b_1 - \sum_{t=1}^{T} b_2 X_t$

$$= T\overline{Y} - \sum_{t=1}^{T}(\overline{Y} - b_2\overline{X}) - b_2\sum_{t=1}^{T} X_t = T\overline{Y} - (T\overline{Y} - b_2 T\overline{X}) - b_2 T\overline{X} = 0. \quad \blacksquare$$

Result 8: $\sum_{t=1}^{T} \hat{e}_t X_t = 0$.

Proof: $\sum_{t=1}^{T} \hat{e}_t X_t = \sum_{t=1}^{T}(Y_t - b_1 - b_2 X_t)X_t = \sum_{t=1}^{T}(Y_t - (\overline{Y} - b_2\overline{X}) - b_2 X_t)X_t$

$$= \sum_{t=1}^{T}(Y_t - \overline{Y})X_t - b_2\sum_{t=1}^{T}(X_t - \overline{X})X_t$$

$$= \sum_{t=1}^{T}(Y_t - \overline{Y})(X_t - \overline{X}) - b_2\sum_{t=1}^{T}(X_t - \overline{X})(X_t - \overline{X})$$

since $\sum_{t=1}^{T}(Y_t - \overline{Y})\overline{X} = \sum_{t=1}^{T}(X_t - \overline{X})\overline{X} = 0$

$= 0. \quad \blacksquare$ since $b_2 = \frac{\sum_{t=1}^{T}(Y_t - \overline{Y})(X_t - \overline{X})}{\sum_{t=1}^{T}(X_t - \overline{X})^2}$

Result 9: $E(\hat{\sigma}^2) = E\left[\sum_{t=1}^{T} \hat{e}_t^2 / (T-2)\right] = \sigma^2$.

Proof: $\sum_{t=1}^{T} \hat{e}_t^2 = \sum_{t=1}^{T}(Y_t - b_1 - b_2 X_t)^2 = \sum_{t=1}^{T}\left[(\beta_1 + \beta_2 X_t + e_t) - (\overline{Y} - b_2\overline{X}) - b_2 X_t\right]^2$

$= \sum_{t=1}^{T}\left[(\beta_1 + \beta_2 X_t + e_t) - (\beta_1 + \beta_2\overline{X} - \overline{e} - b_2\overline{X}) - b_2 X_t\right]^2$ $\qquad \overline{e} = \sum_{t=1}^{T} e_t / T$

$$= \sum_{t=1}^{T}\left[\beta_2(X_t - \overline{X}) + (e_t - \overline{e}) - b_2(X_t - \overline{X})\right]^2$$

$$= \sum_{t=1}^{T}\left[-(b_2 - \beta_2)(X_t - \overline{X}) + (e_t - \overline{e})\right]^2$$

$$= (b_2 - \beta_2)^2\sum_{t=1}^{T}(X_t - \overline{X})^2 + \sum_{t=1}^{T}(e_t - \overline{e})^2 - 2(b_2 - \beta_2)\sum_{t=1}^{T}(X_t - \overline{X})(e_t - \overline{e})$$

$$= (b_2 - \beta_2)^2\sum_{t=1}^{T}(X_t - \overline{X})^2 + \sum_{t=1}^{T}(e_t - \overline{e})^2 - 2(b_2 - \beta_2)\sum_{t=1}^{T}(X_t - \overline{X})e_t$$

$$= (b_2 - \beta_2)^2 \sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2 + \sum_{t=1}^{T}(e_t - \overline{e})^2 - 2(b_2 - \beta_2)^2 \sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2$$

since $b_2 = \beta_2 + \sum\left(X_t - \overline{X}\right)e_t \Big/ \sum\left(X_t - \overline{X}\right)^2$

$$= -(b_2 - \beta_2)^2 \sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2 + \sum_{t=1}^{T}(e_t - \overline{e})^2$$

Take expectation on both sides to have

$$E\left(\sum_{t=1}^{T}\hat{e}_t^2\right) = -E\left[(b_2 - \beta_2)^2\right] \sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2 + E\left[\sum_{t=1}^{T}(e_t - \overline{e})^2\right]$$

$$= -Var(b_2) \sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2 + E\left[\sum_{t=1}^{T}(e_t - \overline{e})^2\right]$$

$$= \frac{-\sigma^2}{\sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2} \sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2 + (T-1)\sigma^2$$

$$= (T-2)\sigma^2$$

Hence, $E\left[\sum_{t=1}^{T}\hat{e}_t^2 \Big/ (T-2)\right] = \sigma^2$. ■

Given a value of the explanatory variable, X_{T+1}, we would like to predict a value of the dependent variable, Y_{T+1}. The least squares **predictor** is:

$$\hat{Y}_{T+1} = b_1 + b_2 X_{T+1}.$$

The corresponding prediction error is defined as:

$$f = \hat{Y}_{T+1} - Y_{T+1} = (b_1 - \beta_1) + (b_2 - \beta_2)X_{T+1} + e_{T+1}.$$

The least squares estimator of **mean response**, μ_{T+1}, when $X = X_{T+1}$ is

$$\hat{\mu}_{T+1} = b_1 + b_2 X_{T+1}$$

Its estimation error is given by

$$\hat{\mu}_{T+1} - E(Y_{T+1}) = (b_1 - \beta_1) + (b_2 - \beta_2)X_{T+1}.$$

Result 10: $E(f) = 0$ and $Var(f) = \sigma^2\left(1 + \frac{1}{T} + \frac{\left(X_{T+1} - \overline{X}\right)^2}{\sum_{t=1}^{T}\left(X_t - \overline{X}\right)^2}\right)$.

Proof: $E(f) = E(\hat{Y}_{T+1} - Y_{T+1}) = \left[E(b_1) - \beta_1\right] + \left[E(b_2) - \beta_2\right]X_{T+1} + E(e_{T+1}) = 0$.

$$Var(f) = E\left(f^2\right) = E\left[(b_1 - \beta_1) + (b_2 - \beta_2)X_{T+1} + e_{T+1}\right]^2$$

$$= E\left[(b_1-\beta_1)^2\right] + E\left[(b_2-\beta_2)^2\right]X_{T+1}^2 + E\left[e_{T+1}^2\right] + 2E[(b_1-\beta_1)(b_2-\beta_2)]X_{T+1}$$

Notice that all the cross-product terms involving estimated parameters and e_{T+1} become zero when expected values are taken, since $(b_1-\beta_1)$ and $(b_2-\beta_2)$ are linear combinations of $e_1, e_2, \ldots, e_T$, all of which are uncorrelated with e_{T+1}.

$$Var(f) = Var(b_1) + X_{T+1}^2 Var(b_2) + Var(e_{T+1}) + 2X_{T+1}Cov(b_1, b_2)$$

$$= \sigma^2\left[\left(\frac{1}{T} + \frac{\overline{X}^2}{\sum_{t=1}^T (X_t-\overline{X})^2}\right) + \frac{X_{T+1}^2}{\sum_{t=1}^T (X_t-\overline{X})^2} + 1 + \frac{-2\overline{X}X_{t+1}}{\sum_{t=1}^T (X_t-\overline{X})^2}\right]$$

$$= \sigma^2\left[1 + \frac{1}{T} + \frac{X_{T+1}^2 - 2X_{T+1}\overline{X} + \overline{X}^2}{\sum_{t=1}^T (X_t-\overline{X})^2}\right] = \sigma^2\left[1 + \frac{1}{T} + \frac{(X_{T+1}-\overline{X})^2}{\sum_{t=1}^T (X_t-\overline{X})^2}\right]. \quad \blacksquare$$

Result 11: $E\left(\hat{\mu}_{T+1} - E(\hat{Y}_{T+1})\right) = 0$ and $Var(\hat{\mu}_{T+1}) = \sigma^2\left(\frac{1}{T} + \frac{(X_{T+1}-\overline{X})^2}{\sum_{t=1}^T (X_t-\overline{X})^2}\right)$.

Proof: It is the same as the Result 10 except the random error term e_{T+1}. $\blacksquare$

Result 12: $\sum_{t=1}^T (Y_t-\overline{Y})^2 = \sum_{t=1}^T (\hat{Y}_t-\overline{Y})^2 + \sum_{t=1}^T (Y_t-\hat{Y}_t)^2$.

That is, SST = SSR + SSE.

Proof: $\sum_{t=1}^T (Y_t-\overline{Y})^2 = \sum_{t=1}^T (Y_t-\hat{Y}_t+\hat{Y}_t-\overline{Y})^2 = \sum_{t=1}^T \left[(\hat{Y}_t-\overline{Y}) + (Y_t-\hat{Y}_t)\right]^2$

$$= \sum_{t=1}^T (\hat{Y}_t-\overline{Y})^2 + \sum_{t=1}^T (Y_t-\hat{Y}_t)^2 + 2\sum_{t=1}^T (\hat{Y}_t-\overline{Y})(Y_t-\hat{Y}_t)$$

$$= \sum_{t=1}^T (\hat{Y}_t-\overline{Y})^2 + \sum_{t=1}^T (Y_t-\hat{Y}_t)^2$$

It is true because the cross-product term drops to zero:

$$\sum_{t=1}^T (\hat{Y}_t-\overline{Y})(Y_t-\hat{Y}_t) = \sum_{t=1}^T (b_1+b_2X_t-\overline{Y})\hat{e}_t$$

$$= \sum_{t=1}^T (b_1-\overline{Y})\hat{e}_t + \sum_{t=1}^T b_2X_t\hat{e}_t$$

$$= (b_1-\overline{Y})\sum_{t=1}^T \hat{e}_t + b_2\sum_{t=1}^T X_t\hat{e}_t \quad \text{since } \sum_{t=1}^T \hat{e}_t = \sum_{t=1}^T \hat{e}_tX_t = 0$$

$$= 0. \quad \blacksquare$$

The Matrix Form of Simple Linear Regression Models

Consider the simple regression model with N observations

$$y_n = \beta_1 + \beta_2 x_n + \varepsilon_n, \qquad n = 1, 2, \ldots, N \tag{B1}$$

This can be written as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \tag{B2}$$

or

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{B3}$$

where $\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$ $\quad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}$ $\quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ $\quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$

Assumptions of the linear Regression model:

(1) $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

(2) $E(\varepsilon) = \begin{bmatrix} E(\varepsilon_1) \\ \vdots \\ E(\varepsilon_N) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}$

(3) $E(\varepsilon\,\varepsilon') = \begin{bmatrix} E(\varepsilon_1^2) & \cdots & E(\varepsilon_1\varepsilon_N) \\ \vdots & \ddots & \vdots \\ E(\varepsilon_N\varepsilon_1) & \cdots & E(\varepsilon_N^2) \end{bmatrix} = \begin{bmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_N.$

(4) $\mathbf{X}$ is an $N \times 2$ matrix with $\det(\mathbf{X}'\,\mathbf{X}) \neq 0$.

(5) (Optional) $\quad \varepsilon \sim N(\mathbf{0},\ \sigma^2 \mathbf{I}_N)$.

Under the assumptions discussed before, the best (minimum variance) linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ is obtained by minimizing the error sum of squares

$$\mathbf{S}(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \tag{B4}$$

This is known as the Gauss-Markov theorem.

Using the formulas for vector differentiation, we have

$$\frac{\partial S(\beta)}{\partial \beta} = -2X'Y + 2X'X\beta \,. \quad \text{(B5)}$$

Setting the equation (B6) to be zero gives the normal equation

$$(\mathbf{X}'\ \mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\ \mathbf{Y}. \quad \text{(B6)}$$

Since the square matrix $(\mathbf{X}'\ \mathbf{X})$ is non-singular, the OLS estimator $\hat{\boldsymbol{\beta}}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad \text{(B7)}$$

Substituting equation (B3) into equation (B7), we get

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\ \mathbf{X}\beta + \varepsilon) = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$$

Since $E(\boldsymbol{\varepsilon}) = 0$, we have $E(\hat{\beta}) = \beta$. Thus $\hat{\beta}$ is an unbiased estimator. Also,

$$\mathbf{Var}(\hat{\beta}) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1}$$

$$= \sigma^2(X'X)^{-1} \quad \text{since } E(\varepsilon\varepsilon') = \sigma^2 I_N .$$

The $\hat{\beta}$ is unbiased and has a covariance matrix $\sigma^2(X'X)^{-1}$.

According to equation (B7), the vectors of fitted values $\hat{Y}$ and the least squares residuals $\hat{\varepsilon}$ are

$$\hat{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{P}_\mathbf{X}\mathbf{Y} \quad \text{(B8)}$$

$$\hat{\varepsilon} = \mathbf{Y} - \hat{Y} = \mathbf{Y} - \mathbf{P}_\mathbf{X}\mathbf{Y} = (\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{Y} = \mathbf{M}_\mathbf{X}\mathbf{Y} \quad \text{(B9)}$$

where $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{M}_\mathbf{X} = (\mathbf{I}_N - \mathbf{P}_\mathbf{X})$. **Note** that it can be shown that $\mathbf{P}_\mathbf{X}\,\mathbf{P}_\mathbf{X} = \mathbf{P}_\mathbf{X}$ and $\mathbf{M}_\mathbf{X}\,\mathbf{M}_\mathbf{X} = \mathbf{M}_\mathbf{X}$ (P_X and M_X are called **idempotent** matrices).

The total variation of the dependent variable is the sum of squared deviations from its mean (SST):

$$\mathbf{SST} = \sum_{i=1}^{N}(y_i - \bar{y})^2 = \mathbf{Y}'\ \mathbf{Y} - n\bar{y}^2 .$$

It can be shown that total sum of squares = regression sum of squares + error sum of squares (**SST** = **SSR** + **SSE**), where

$$\mathbf{SSR} = \mathbf{Y}'\mathbf{P}_\mathbf{X}\mathbf{Y} - n\bar{y}^2 = \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - n\bar{y}^2 = \hat{\boldsymbol{\beta}}'X'Y - n\bar{y}^2$$

$$\mathbf{SSE} = Y'(I_N - P_X)Y = Y'M_XY = Y'M_X'M_XY = \hat{\varepsilon}'\hat{\varepsilon}$$

The unbiased estimator of σ^2 is

$$\hat{\boldsymbol{\sigma}}^2 = \frac{\boldsymbol{SSE}}{\boldsymbol{N}-2} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{N-2} = \frac{\boldsymbol{Y}'(\boldsymbol{I_N}-\boldsymbol{P_X})\boldsymbol{Y}}{N-2}.$$

We now calculate elements of some matrices discussed above in order to verify some results derived in earlier classes:

$$(\mathbf{X}'\mathbf{X}) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} = \begin{bmatrix} N & \sum_{n=1}^N x_n \\ \sum_{n=1}^N x_n & \sum_{n=1}^N x_n^2 \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{N\sum_{n=1}^N x_n^2 - \left(\sum_{n=1}^N x_n\right)^2} \begin{bmatrix} \sum_{n=1}^N x_n^2 & -\sum_{n=1}^N x_n \\ -\sum_{n=1}^N x_n & N \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{Y}) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^N y_n \\ \sum_{n=1}^N x_n y_n \end{bmatrix}.$$

Base on the above results, we have

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$= \frac{1}{N\sum_{n=1}^N x_n^2 - \left(\sum_{n=1}^N x_n\right)^2} \begin{bmatrix} \sum_{n=1}^N x_n^2 & -\sum_{n=1}^N x_n \\ -\sum_{n=1}^N x_n & N \end{bmatrix} \begin{bmatrix} \sum_{n=1}^N y_n \\ \sum_{n=1}^N x_n y_n \end{bmatrix}$$

$$= \frac{1}{N\sum_{n=1}^N x_n^2 - \left(\sum_{n=1}^N x_n\right)^2} \begin{bmatrix} \sum_{n=1}^N x_n^2 \sum_{n=1}^N y_n - \sum_{n=1}^N x_n \sum_{n=1}^N x_n y_n \\ N\sum_{n=1}^N x_n y_n - \sum_{n=1}^N x_n \sum_{n=1}^N y_n \end{bmatrix}$$

The covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1} = \frac{\sigma^2}{N\sum_{n=1}^N x_n^2 - \left(\sum_{n=1}^N x_n\right)^2} \begin{bmatrix} \sum_{n=1}^N x_n^2 & -\sum_{n=1}^N x_n \\ -\sum_{n=1}^N x_n & N \end{bmatrix}$$

The Matrix Form of Multiple Linear Regression Models

Suppose that we have the following N observations:

$$\begin{array}{ccccccccccccc} y_1 & = & \beta_1 & + & \beta_2 x_{12} & + & \beta_3 x_{13} & + & \cdots & + & \beta_K x_{1K} & + & \varepsilon_1 \\ y_2 & = & \beta_1 & + & \beta_2 x_{22} & + & \beta_3 x_{23} & + & \cdots & + & \beta_K x_{2K} & + & \varepsilon_2 \\ \vdots & & \vdots & & \vdots & & \vdots & & \cdots & & \vdots & & \vdots \\ y_N & = & \beta_1 & + & \beta_2 x_{N2} & + & \beta_3 x_{K3} & + & \cdots & + & \beta_K x_{NK} & + & \varepsilon_N \end{array}$$

The matrix form is

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

where

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} 1 & x_{12} & \cdots & x_{1K} \\ 1 & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N2} & \cdots & x_{NK} \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

The assumptions of the multiple linear Regression model are the same as the simple linear regression model *except* $\mathbf{X}$ being an $N \times K$ matrix. Other matrix algebras are exactly the same as the simple linear regression model