Errata Correction

Unfortunately, I made a minor mistake in explaining the non-decreasing property of R^2 in class. I will provide a more thorough discussion here to clarify the corresponding points. Recall that we have data matrix $\boldsymbol{X}$, which is a $n \times (k+1)$ matrix as we have discussed in class. With the OLS estimation $\hat{\boldsymbol{\beta}}$, we have

$$\boldsymbol{Y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{e}. \tag{1}$$

If we consider adding an additional regressor X_0 $(n \times 1)$,

$$\boldsymbol{Y} = \boldsymbol{X}\tilde{\boldsymbol{\beta}} + X_0\tilde{\beta}_0 + \tilde{\boldsymbol{e}}. \tag{2}$$

From (1), we have $\boldsymbol{e}^\top \boldsymbol{X} = \boldsymbol{0}$; From (2), we have $\tilde{\boldsymbol{e}}^\top \boldsymbol{X} = \tilde{\boldsymbol{e}}^\top X_0 = \boldsymbol{0}$. Using these results, we can jointly derive that

$$\begin{aligned} \boldsymbol{e}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{e}^\top \boldsymbol{e} &= \boldsymbol{e}^\top \boldsymbol{X}\tilde{\boldsymbol{\beta}} + \boldsymbol{e}^\top X_0\tilde{\beta}_0 + \boldsymbol{e}^\top \tilde{\boldsymbol{e}}, \\ \tilde{\boldsymbol{e}}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} + \tilde{\boldsymbol{e}}^\top \boldsymbol{e} &= \tilde{\boldsymbol{e}}^\top \boldsymbol{X}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{e}}^\top X_0\tilde{\beta}_0 + \tilde{\boldsymbol{e}}^\top \tilde{\boldsymbol{e}}, \end{aligned}$$

and finally

$$\begin{aligned} \boldsymbol{e}^\top \boldsymbol{e} &= \boldsymbol{e}^\top X_0\tilde{\beta}_0 + \boldsymbol{e}^\top \tilde{\boldsymbol{e}}, & (3) \\ \tilde{\boldsymbol{e}}^\top \boldsymbol{e} &= \tilde{\boldsymbol{e}}^\top \tilde{\boldsymbol{e}}. & (4) \end{aligned}$$

By substituting (4) into (3), we can solve

$$\boldsymbol{e}^\top X_0\tilde{\beta}_0 = \boldsymbol{e}^\top \boldsymbol{e} - \tilde{\boldsymbol{e}}^\top \tilde{\boldsymbol{e}}. \tag{5}$$

Now we calculate $\tilde{\boldsymbol{e}}^\top \tilde{\boldsymbol{e}}$ as follows

$$\begin{aligned} \tilde{\boldsymbol{e}}^\top \tilde{\boldsymbol{e}} &= \left[\boldsymbol{X}\left(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\right) - X_0\tilde{\beta}_0 + \boldsymbol{e}\right]^\top \left[\boldsymbol{X}\left(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\right) - X_0\tilde{\beta}_0 + \boldsymbol{e}\right] \\ &= \boxed{\left[\boldsymbol{X}\left(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\right) - X_0\tilde{\beta}_0\right]^\top \left[\boldsymbol{X}\left(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\right) - X_0\tilde{\beta}_0\right]} + \boxed{2\boldsymbol{e}^\top \boldsymbol{X}\left(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\right)} - 2\boldsymbol{e}^\top X_0\tilde{\beta}_0 + \boldsymbol{e}^\top \boldsymbol{e} \\ &= \boxed{\left[\boldsymbol{X}\left(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\right) - X_0\tilde{\beta}_0\right]^\top \left[\boldsymbol{X}\left(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\right) - X_0\tilde{\beta}_0\right]} - 2\left(\boldsymbol{e}^\top \boldsymbol{e} - \tilde{\boldsymbol{e}}^\top \tilde{\boldsymbol{e}}\right) + \boldsymbol{e}^\top \boldsymbol{e}. \end{aligned}$$

By rewriting the third equation above, we obtain

$$\boldsymbol{e}^{\top}\boldsymbol{e} = \underbrace{\boxed{\left[\boldsymbol{X}\left(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\right) - X_0\tilde{\beta}_0\right]^{\top}\left[\boldsymbol{X}\left(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}\right) - X_0\tilde{\beta}_0\right]}}_{\geqslant 0} + \tilde{\boldsymbol{e}}^{\top}\tilde{\boldsymbol{e}} \tag{6}$$

which suggests that $\boldsymbol{e}^{\top}\boldsymbol{e} \geqslant \tilde{\boldsymbol{e}}^{\top}\tilde{\boldsymbol{e}}$. Note that in class I missed some terms in the red box of collected terms.