Spring 2020 | Quantitative Analysis
Chung-Ming Kuan

Problem Set 2: Solution

Part One: Hand-Written Exercise

1. Since $\hat{r}_{i,1}$ is the OLS residual of regressing x_1 on the constant one and $x_2, ..., x_k$, we can write:

$$x_{i,1} = \hat{c}_0 + \hat{c}_2 x_{i,2} + ... + \hat{c}_k x_{i,k} + \hat{r}_{i,1},$$

where $\hat{c}_j$ are the OLS estimates. We now continue with the proof:

$$\begin{aligned}\sum \hat{r}_{i,1}^2 &= \sum \hat{r}_{i,1} \left(x_{i,1} - \hat{c}_0 - \hat{c}_2 x_{i,2} - ... - \hat{c}_k x_{i,k}\right) \\ &= \sum \hat{r}_{i,1} x_{i,1} - \hat{c}_0 \sum \hat{r}_{i,1} - \hat{c}_2 \sum \hat{r}_{i,1} x_{i,2} - ... - \hat{c}_k \sum \hat{r}_{i,1} x_{i,k} \\ &= \sum \hat{r}_{i,1} x_{i,1},\end{aligned}$$

by the fact that $\sum \hat{r}_{i,1} = \sum \hat{r}_{i,1} x_{i,2} = ... = \sum \hat{r}_{i,1} x_{i,k} = 0$, and this completes the proof.

2. (a) $\hat{\beta}_1$ is still unbiased, since the homoskedasticity assumption played no role in showing that the OLS estimators are unbiased.

   (b) $\hat{\beta}_1$ **could be biased** (but can still be unbiased), since omitting an important variable can cause bias, but this is true only when x_3 is correlated with x_1 or x_2.

   (c) $\hat{\beta}_1$ is still unbiased, since we only include an irrelevant variables.

3. We can not calculate the OLS estimators because there is exact multicollinearity among regressors, that is, the 2^{nd} column equals the 1^{st} column times 2. Therefore, $(\mathbf{X'X})^{-\mathbf{1}}$ doesn't exist.

4. The least-squares(LS) criterion function is

$$\mathrm{Q} := \textstyle\sum_{i=1}^n u_i^2 = \sum_{i=1}^n (y_i - \beta_1 x_{i1} - ... - \beta_k x_{ik})^2$$

The first order conditions(FOCs) are

$$\frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^n (y_i - \beta_1 x_{i1} - ... - \beta_k x_{ik}) x_{i1} = 0 \tag{1}$$

$$\vdots$$

$$\frac{\partial Q}{\partial \beta_k} = -2 \sum_{i=1}^n (y_i - \beta_1 x_{i1} - ... - \beta_k x_{ik}) x_{ik} = 0 \tag{2}$$

From (1)...(2), we can get $\sum_{i=1}^{n} x_{ij}\hat{u}_i = 0, \forall 1 \leq j \leq k$.(Note that $\sum_{i=1}^{n} \hat{u}_i \neq 0$)
And

$$\begin{aligned}
SST &= \sum_{i=1}^{n}(y_i - \bar{y})^2 \\
&= \sum_{i=1}^{n}(y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \\
&= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) \\
&= SSR + SSE + 2\sum_{i=1}^{n}\hat{u}_i\hat{y}_i - 2\bar{y}\sum_{i=1}^{n}\hat{u}_i \\
&= SSR + SSE - 2\bar{y}\sum_{i=1}^{n}\hat{u}_i \quad (\text{By} \ \ (1)...(2), \ \ \sum_{i=1}^{n}\hat{u}_i\hat{y}_i = 0) \\
&\neq SSR + SSE
\end{aligned}$$

Thus, SSR is not necessarily less than SST, which means R^2 may be negative.