Jared Yu
Module 9 Discussion

Suppose that we fit the full model $y = b_0 + b_1 x_1 + b_2 x_2 + e$ to n statistically independent observations and obtain the least-squares estimator of b_1.

x_1 and x_2 are fixed (that is, they are not random variables).

Then we fit the subset model $y = b_0 + b_1 x_1 + e$ to the same n observations and obtain another least-squares estimator of b_1.

Discuss which least-squares estimator of b_1 is more biased for the parameter b_1 and which has smaller variance. State assumptions underlying your discussion.

Ans: (*Note: This largely references pp. 329-330 from the Textbook.*)
A first assumption that we can add is that $n \geq K + 1$, where here $K = 2$ in the full model and $K = 1$ in the subset model. This is should not be an issue, since the requirement here is only that $n \geq 3$ or $n \geq 2$ which is quite minimal.

From previous chapters, we have learned that the (full) model can be written in matrix notation,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where the only difference here is that β and ε are used in place of b and e respectively. This just follows the traditional notation, rather than the notation of the problem. Equation (1) can be rewritten as follows,

$$\mathbf{y} = \mathbf{X}_p\boldsymbol{\beta}_p + \mathbf{X}_r\boldsymbol{\beta}_r + \boldsymbol{\varepsilon}, \tag{2}$$

where $\mathbf{X}_p\boldsymbol{\beta}_p$ corresponds with x_1 (and x_0 which are all 1's), b_0, and b_1, while $\mathbf{X}_r\boldsymbol{\beta}_r$ corresponds with x_2 and b_2.

For the full model we have the following estimates for $\boldsymbol{\beta}$ and σ^2,

$$\widehat{\boldsymbol{\beta}}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{3}$$

and

$$\hat{\sigma}^{2*} = \frac{\mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\beta}}^*\mathbf{X}'\mathbf{y}}{n - K - 1} = \frac{y'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}}{n - 1}. \tag{4}$$

The subset model, $\mathbf{y} = \mathbf{X}_p\boldsymbol{\beta}_p + \boldsymbol{\varepsilon}$, has the corresponding estimates,

$$\widehat{\boldsymbol{\beta}}_p = \left(\mathbf{X}_p'\mathbf{X}_p\right)^{-1}\mathbf{X}_p'\mathbf{y} \tag{5}$$

and

$$\hat{\sigma}^{2*} = \frac{\mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\beta}}_p\mathbf{X}_p'\mathbf{y}}{n - p} = \frac{y'\left[\mathbf{I} - \mathbf{X}_p\left(\mathbf{X}_p'\mathbf{X}_p\right)^{-1}\mathbf{X}_p'\right]\mathbf{y}}{n - 3}. \tag{6}$$

Regarding b_1, the expected valued of $\widehat{\boldsymbol{\beta}}_p$ is

$$E(\widehat{\boldsymbol{\beta}}_p) = \boldsymbol{\beta}_p + \left(\mathbf{X}_p'\mathbf{X}_p\right)^{-1}\mathbf{X}_p'\mathbf{X}_r\boldsymbol{\beta}_r = \boldsymbol{\beta}_p + \mathbf{A}\boldsymbol{\beta}_r \tag{7}$$

where $\mathbf{A} = \left(\mathbf{X}_p'\mathbf{X}_p\right)^{-1}\mathbf{X}_p'\mathbf{X}_r$ is also known as the alias matrix. Equation (7) indicates then that b_1 is biased in the case of the subset model, since $E(\widehat{\boldsymbol{\beta}}_p) \neq \boldsymbol{\beta}_p$. The exception would be in the case that $b_2 = 0$ or x_1 and x_2 are orthogonal, in which case $\mathbf{A}$ would be zero. On the other hand, we know from before that the multiple linear regression (MLR) model is unbiased. So, given Equation (3), we can see that within the full model, it is unbiased for b_1.

Regarding the variance of b_1, we have that for $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\boldsymbol{\beta}}_p$ that

$$Var(\widehat{\boldsymbol{\beta}}^*) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

and

$$Var(\widehat{\boldsymbol{\beta}}_p) = \sigma^2(\mathbf{X}_p'\mathbf{X}_p)^{-1}.$$

From this, the matrix $Var(\widehat{\boldsymbol{\beta}}^*) - Var(\widehat{\boldsymbol{\beta}}_p)$ is positive semidefinite. This implies that the variance of the estimates for the full model is greater than or equal to the variance of the estimate in the subset model.

For simplicity, we can denote $C = (\mathbf{X}'\mathbf{X})^{-1}$ and $\tilde{C} = (\mathbf{X}_p'\mathbf{X}_p)^{-1}$. The variance of b_1 then can be found by looking at $\sigma^2 C_{11}$ and $\sigma^2 \tilde{C}_{11}$. Since the resulting matrix $Var(\widehat{\boldsymbol{\beta}}^*) - Var(\widehat{\boldsymbol{\beta}}_p)$ is positive semidefinite, that implies that the variance for b_1 will also be the same if not larger.

Note: I am somewhat confused about one issue regarding my discussion at the end. When I discuss the resulting matrix that is calculated from $Var(\widehat{\boldsymbol{\beta}}^*) - Var(\widehat{\boldsymbol{\beta}}_p)$, I am directly sourcing this from the textbook. However, since these are matrices based on the inverse of the covariance matrix between the corresponding $\mathbf{X}$ and $\mathbf{X}_p$, my thinking is that their dimensions will not match, given that the former is larger than the latter. I think I am missing something either obvious or not stated explicitly, but in my current thinking it'd require these matrices to have matching dimensions before the difference can be taken. The idea however is simply that the variance of b_1, found by looking at the diagonal is either equal or larger in the full model.