MA 575 Linear Models:

Cedric E. Ginestet, Boston University

Adding Covariates and OLS Variance Estimator

Week 8, Lecture 1

1 Adding Covariates to a Simple Regression

In simple linear regression, we have seen that the t-statistic for the slope coefficient of the variable $\mathbf{x}_1 := [x_1, \ldots, x_n]^T$, is given by

$$t_1(\mathbf{x}_1) := \frac{\mathrm{SXY} \,/\, \mathrm{SXX}}{\widehat{\sigma}/\sqrt{\mathrm{SXX}}} = \frac{\mathrm{SXY}}{\widehat{\sigma}\sqrt{\mathrm{SXX}}}.$$

If we were to introduce a new variable, $\mathbf{x}_2 := [x_{12}, \ldots, x_{n2}]^T$, in this model, the new t-statistic for the slope of this particular variable would take the form

$$t_1(\mathbf{x}_1, \mathbf{x}_2) = \frac{\widehat{\beta}_1}{\mathrm{se}(\widehat{\beta}_1)} = \frac{[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}]_1}{\widehat{\sigma}\sqrt{(\mathbf{X}^T\mathbf{X})^{-1}_{11}}}.$$

In this setting, it can be seen that the relationship between $t_1(\mathbf{x}_1)$ and $t_1(\mathbf{x}_1, \mathbf{x}_2)$ is somewhat complex. Many factors enter in the computation of both of these quantities, and in some cases, the introduction of $\mathbf{x}_2$ may actually *increase* the value of the t-statistic for $\mathbf{x}_1$.

In multiple regression, recall that we have the following relationship for $\widehat{\boldsymbol{\beta}}^*$, which is defined as $\widehat{\boldsymbol{\beta}}$ without the OLS estimator of the y-intercept,

$$\widehat{\boldsymbol{\beta}}^* = (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T\mathcal{Y},$$

where observe that

$$\mathcal{X}^T\mathcal{Y} := \begin{vmatrix} s^2_{1y} \\ s^2_{2y} \end{vmatrix}.$$

If the model has only two variables, then we have

$$\mathbf{S} := \mathcal{X}^T\mathcal{X} = \begin{vmatrix} s^2_{11} & s^2_{12} \\ s^2_{21} & s^2_{22} \end{vmatrix},$$

such that the first entry in the vector $\widehat{\boldsymbol{\beta}}^*$ is

$$\widehat{\beta}^*_1 = \frac{1}{|\mathbf{S}|}\left(s^2_{22}s^2_{1y} - s^2_{12}s^2_{2y}\right).$$

Here, we may consider two different stylized scenarios, where we make special assumptions on the relationship between $\mathbf{x}_1$ and $\mathbf{x}_2$.

i. Clearly, if the two covariates are **uncorrelated**, then we have $s^2_{12} = 0$, and then we recover the slope coefficient of the simple regression model that solely contains $\mathbf{x}_1$,

$$\widehat{\beta}^*_1 = \frac{s^2_{22}s^2_{1y}}{s^2_{22}s^2_{11}} = \frac{\mathrm{SX_1Y}}{\mathrm{SX_1X_1}}.$$

ii. Moreover, if we had previously **standardized** the empirical variances of $\mathbf{x}_2$ and $\mathbf{x}_1$, we would have $s_{22}^2 = s_{11}^2 = 1$, and therefore $|\mathbf{S}| \in [0, 1]$, since $\mathbf{S}$ is positive definite, and its determinant must be greater than zero. In this case, it is easy to see that the slope coefficient of $\mathbf{x}_1$ would necessarily increase after introducing $\mathbf{x}_2$.

Crucially, however, little can be said about the general behavior of the values of the corresponding t-statistics. Various cases can be constructed such that the t-statistic for $\mathbf{x}_1$ may increase or decrease depending on the relationship between $\mathbf{x}_1$ and $\mathbf{x}_2$, and the variances of $\mathbf{y}$, $\mathbf{x}_1$, and $\mathbf{x}_2$.

It is especially difficult to make any general statement about the changes in the statistical significance of these statistics, because **correlation is not a transitive operator**. That is, for any realization of three random vectors denoted $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, if

$$\mathbb{C}\text{ov}(\mathbf{x}, \mathbf{y}) > 0, \text{ and } \mathbb{C}\text{ov}(\mathbf{y}, \mathbf{z}) > 0;$$

then it not true, in general, that

$$\mathbb{C}\text{ov}(\mathbf{x}, \mathbf{z}) > 0.$$

This can be easily verified by using a geometrical argument. However, note that in some circumstances, if the first two correlation coefficients are sufficiently large, then it follows that the third correlation will necessarily be positive.

2 Unbiasedness of OLS Variance

Recall that the OLS estimator for σ^2 is given by

$$\widehat{\sigma}^2 := \frac{1}{n - p^*} \text{RSS}(\widehat{\boldsymbol{\beta}}).$$

We wish to take the expectation of this quantity, such that

$$\mathbb{E}[\widehat{\sigma}^2 | \mathbf{X}] = \frac{1}{n - p^*} \mathbb{E}[\text{RSS}(\widehat{\boldsymbol{\beta}}) | \mathbf{X}].$$

Firstly, we can reformulate the OLS estimator of the RSS in the following way,

$$\text{RSS}(\widehat{\boldsymbol{\beta}}) = \widehat{\mathbf{e}}^T \widehat{\mathbf{e}} = \mathbf{y}^T (\mathbf{I} - \mathbf{H})^T (\mathbf{I} - \mathbf{H}) \mathbf{y} = \mathbf{y}^T (\mathbf{I} - \mathbf{H}) \mathbf{y}, \tag{1}$$

using the fact that $\mathbf{I} - \mathbf{H}$ is both *symmetry* and *idempotent*. Moreover, we have shown in a previous homework that for any random vector $\mathbf{y}$, we have

$$\mathbb{V}\text{ar}[\mathbf{y}] = \mathbb{E}[(\mathbf{y} - \mathbb{E}[\mathbf{y}])(\mathbf{y} - \mathbb{E}[\mathbf{y}])^T] = \mathbb{E}[\mathbf{y}\mathbf{y}^T] - \mathbb{E}[\mathbf{y}]\mathbb{E}[\mathbf{y}]^T,$$

which immediately gives

$$\mathbb{E}[\mathbf{y}\mathbf{y}^T] = \mathbb{V}\text{ar}[\mathbf{y}] + \mathbb{E}[\mathbf{y}]\mathbb{E}[\mathbf{y}]^T.$$

Finally, it is also true that for any quadratic form, $\mathbf{y}^T \mathbf{A} \mathbf{y}$, with a random vector $\mathbf{y} \in \mathbb{R}^n$, we have

$$\mathbb{E}[\mathbf{y}^T \mathbf{A} \mathbf{y}] = \mathbb{E}\left[\text{tr}(\mathbf{y}^T \mathbf{A} \mathbf{y})\right] = \mathbb{E}\left[\text{tr}(\mathbf{A} \mathbf{y} \mathbf{y}^T)\right] = \text{tr}(\mathbf{A} \mathbb{E}[\mathbf{y}\mathbf{y}^T]).$$

using the standard cycling property of the trace, $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$, and the fact that $\mathbf{A}$ is non-random.

These last two results can be put to good use in equation (1), in order to obtain

$$\begin{aligned}
\mathbb{E}[\mathrm{RSS}(\widehat{\boldsymbol{\beta}})|\mathbf{X}] &= \mathbb{E}[\mathbf{y}^T(\mathbf{I}-\mathbf{H})\mathbf{y}|\mathbf{X}] \\
&= \mathbb{E}[\mathrm{tr}(\mathbf{y}^T(\mathbf{I}-\mathbf{H})\mathbf{y})|\mathbf{X}] \\
&= \mathrm{tr}((\mathbf{I}-\mathbf{H})\mathbb{E}[\mathbf{y}\mathbf{y}^T|\mathbf{X}]) \\
&= \mathrm{tr}((\mathbf{I}-\mathbf{H})\,\mathbb{V}\mathrm{ar}[\mathbf{y}|\mathbf{X}]) + \mathrm{tr}((\mathbf{I}-\mathbf{H})\mathbb{E}[\mathbf{y}|\mathbf{X}]\mathbb{E}[\mathbf{y}|\mathbf{X}]^T).
\end{aligned}$$

We can use another cycling argument for the second trace in this equation. Moreover, because the resulting term, $\mathbb{E}[\mathbf{y}|\mathbf{X}]^T(\mathbf{I}-\mathbf{H})\mathbb{E}[\mathbf{y}|\mathbf{X}]$, is a quadratic form, we can also drop the trace for notational convenience. This gives

$$\begin{aligned}
\mathbb{E}[\widehat{\sigma}^2|\mathbf{X}] &= \mathrm{tr}((\mathbf{I}-\mathbf{H})\mathbf{I}\sigma^2) + \mathbb{E}[\mathbf{y}|\mathbf{X}]^T(\mathbf{I}-\mathbf{H})\mathbb{E}[\mathbf{y}|\mathbf{X}] \\
&= \mathrm{tr}((\mathbf{I}-\mathbf{H})\mathbf{I}\sigma^2) + \boldsymbol{\beta}^T\mathbf{X}^T(\mathbf{I}-\mathbf{H})\mathbf{X}\boldsymbol{\beta} \\
&= \sigma^2\mathrm{tr}(\mathbf{I}-\mathbf{H}) + \boldsymbol{\beta}^T\mathbf{X}^T(\mathbf{X}\boldsymbol{\beta}-\mathbf{H}\mathbf{X}\boldsymbol{\beta}) \\
&= \sigma^2\mathrm{tr}(\mathbf{I}-\mathbf{H}),
\end{aligned}$$

where the last equality can be easily verified as follows,

$$\mathbf{H}\mathbf{X} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X} = \mathbf{X},$$

or simply by observing that since $\mathbf{H}$ projects to the column space of $\mathbf{X}$, it necessarily follows that every column of $\mathbf{X}$ is unaffected by $\mathbf{H}$. That is, for any column vector of $\mathbf{X}$, denoted $\mathbf{x}_j := [x_{1j}, \ldots, x_{nj}]^T$, we have $\mathbf{H}\mathbf{x}_j = \mathbf{x}_j$ for every $j = 0, \ldots, p$.

Therefore, we have just shown that

$$\mathbb{E}[\mathrm{RSS}(\widehat{\boldsymbol{\beta}})|\mathbf{X}] = \sigma^2\mathrm{tr}(\mathbf{I}-\mathbf{H}) = \sigma^2(\mathrm{tr}(\mathbf{I}) - \mathrm{tr}(\mathbf{H})) = (n-p^*)\sigma^2,$$

since the trace of a projection matrix is simply the dimension of the space to which it projects. Alternatively, because the trace of a matrix is also the sum of its eigenvalues, we could also use the fact that an orthogonal projection only has ones and zeros as eigenvalues. We have thus demonstrated that the OLS estimator for σ^2 is unbiased. That is, for any value $\sigma^2 \in \mathbb{R}^+$, it is true that

$$\mathbb{E}[\widehat{\sigma}^2|\mathbf{X}] = \sigma^2.$$