ECON 253: Introduction to Econometrics
Bryn Mawr College
Interaction Term Coefficient Explanation

TOPIC: Understanding the interpretation of the coefficient on the interaction term

BACKGROUND:

Consider the model we've seen before,

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i + u_i, \qquad Eq. (1)$$

where all notation is consistent with what we use in class. Also, consistent with class, the expectation of the error term conditional on values of X, Z, and $X_i Z_i$ is zero.

As we learned in class, β_3 can be interpreted as the average effect on Y from the interaction of a one-unit change in X and a one-unit change in Z, separate from a one-unit change in X and a one-unit change in Z.

This is pretty abstract. It might be helpful to think about this in the context of an example as we've done in class. Suppose Y is wages, X is years of education and Z is years of work experience. What β_1 will give you is the average effect of another year of education on wage. β_2 will give you the average effect of another year of work experience on wage. Now, don't you think that more education and more experience will combine together in some way that is distinct to affect wages? Maybe employers see education *and* experience as being particularly attractive above and beyond more education and more experience separately. β_3 will give us this effect, in other words, it is the effect of the interaction of one more year of education *and* one more year of work experience on wages *separate* from the isolated effects of one more year of education and one more year of experience on their own.

But how can we show this mathematically?

DERIVATION:

In math, I'll define this effect as,

$$\beta_3 = \Delta_1^{XZ} - \Delta_1^{X} - \Delta_1^{Z} \qquad Eq. (2)$$

where Δ_1^{Z} is a one-unit change in Z, Δ_1^{X} is a one-unit change in X, and Δ_1^{XZ} is a one-unit change in the interaction of X and Z. This is simply the statement made above, "β_3 can be interpreted as the average effect on Y from the interaction of a one-unit change in X and a one-unit change in Z, separate from a one-unit change in X and a one-unit change in Z." in mathematical notation.

Let's show this. We'll start by defining two values, ϕ_1 and ϕ_2, that can take on any real number. Formally, $\phi_1, \phi_2 \in \mathbb{R}$.

Now, let's look at the following conditional expectations of Y from $Eq.(1)$ above given values of X, Z, XZ and one-unit changes in X and Z,

$$E[Y|X = \phi_1, Z = \phi_2, XZ = \phi_1\phi_2] = \beta_0 + \beta_1\phi_1 + \beta_2\phi_2 + \beta_3\phi_1\phi_2, \qquad Eq.(3)$$

$$\begin{aligned} &E[Y|X = \phi_1 + 1, Z = \phi_2, XZ = (\phi_1 + 1)\phi_2] \\ &\qquad = \beta_0 + \beta_1(\phi_1 + 1) + \beta_2\phi_2 + \beta_3(\phi_1 + 1)\phi_2, \end{aligned} \qquad Eq.(4)$$

$$\begin{aligned} &E[Y|X = \phi_1, Z = \phi_2 + 1, XZ = \phi_1(\phi_2 + 1)] \\ &\qquad = \beta_0 + \beta_1\phi_1 + \beta_2(\phi_2 + 1) + \beta_3\phi_1(\phi_2 + 1), and \end{aligned} \qquad Eq.(5)$$

$$\begin{aligned} &E[Y|X = \phi_1 + 1, Z = \phi_2 + 1, XZ = (\phi_1 + 1)(\phi_2 + 1)] \\ &\qquad = \beta_0 + \beta_1(\phi_1 + 1) + \beta_2(\phi_2 + 1) + \beta_3(\phi_1 + 1)(\phi_2 + 1). \end{aligned} \qquad Eq.(6)$$

Now, define, one-unit changes as follows,

$$\Delta_1^X = Eq.(4) - Eq.(3) = \beta_1 + \beta_3\phi_2, \qquad Eq.(7)$$

$$\Delta_1^Z = Eq.(5) - Eq.(3) = \beta_2 + \beta_3\phi_1, and \qquad Eq.(8)$$

$$\Delta_1^{XZ} = Eq.(6) - Eq.(3) = \beta_1 + \beta_2 + \beta_3(\phi_1 + \phi_2 + 1). \qquad Eq.(9)$$

The last few expressions are what you get after some straightforward simplification.

Note that $Eq.(7)$ implies $\beta_1 = \Delta_1^X - \beta_3\phi_2$ and $Eq.(8)$ implies $\beta_2 = \Delta_1^Z - \beta_3\phi_1$. Substituting these expressions into $Eq.(9)$ results in,

$$\Delta_1^{XZ} = \Delta_1^X - \beta_3\phi_2 + \Delta_1^Z - \beta_3\phi_1 + \beta_3(\phi_1 + \phi_2 + 1). \qquad Eq.(10)$$

Simplifying the right-hand side and isolating β_3 results in,

$$\beta_3 = \Delta_1^{XZ} - \Delta_1^X - \Delta_1^Z, \qquad Eq.(11)$$

which is what we wanted to show.