**FIXED EFFECTS ACTIVITY GUIDANCE**

**Activity learning goals:**

* Use fixed effects models in situations with time-invariant unobserved heterogeneity.
* Estimate fixed effects models using first differences.
* Estimate fixed effects models using within transformations.

**Introducing the activity:**

Do you believe getting married makes people less likely to commit crimes? Why? In this exercise we develop a new method that can be used to test this hypothesis. Suppose you have data containing the number of crimes committed in the previous year and current marital status for 500 individuals. Additionally, suppose you have two observations per individual spaced four years apart. Data where you have multiple observations per individual spread across time is called panel data or longitudinal data.

**Guiding students during the activity:**

1. *Consider the following model:*

*Suggest at least two omitted variables that could induce bias in your estimate of β1.*

Students are very good at coming up with possible confounders here. We have had students suggest that violent tendencies, risk aversion, and ability to earn a market wage are all correlated with marital status and could be predictors of criminal behavior.

1. *Suppose all of the omitted variable bias comes from variables whose values do not change across time. Let ui in the following model represent the contribution of these variables. We will call this the “fixed effect.”*

*We cannot estimate this model directly with OLS because we do not observe , and the unobserved part of the equation () may be correlated with marital status. That said, this equation must hold in both time period 1 and 2:*

*How might you combine these equations to get an equation that* ***can*** *be estimated with OLS? Verify that each of the assumptions required by OLS holds and interpret in the context of your new model equation.*

Most students figure out that if they subtract one equation from the other, they get a new equation that does not contain the fixed effect. The key is for students to recognize that the error term in the new model () is mean zero and uncorrelated with the new explanatory variable ().

1. *Now suppose you had three time periods of data. Propose another method that uses all of your data to estimate .*

It is fairly unusual for students to come up with a within-difference model (i.e., one where they subtract the individual-specific mean values across time from each observation), and they more often difference the first two equations and the second and third equations.

**Wrapping up the activity:**

When we show them the first difference method, it usually looks very similar (if not identical) to what they’ve invented. The key is to point out that estimating this model requires regressing changes in criminal activity on changes in marital status. The model is identified by both marriages and marital dissolutions. That is, the model assumes that the effect of a marriage is exactly the negative of the effect of a divorce or widowhood. This is not always a reasonable assumption.

We also ask the class what it means that the differenced model does not have an intercept. We explain that this implies that the change across time (in this case during the 4-year period) will be on average zero if there is no change in marital status. In some situations this is realistic, and we discuss whether this is the case here. The answer hinges on whether we think an individual’s propensity to commit crime changes as they age. To address this possibility, we introduce a time fixed effect into the model.