Interaction Variables with Panel Data

- Continuous and Categorical Interactions

Limited Dependent Variables

- Review of Linear Probability
- Probit/Logit Models
- Tobit Models
- Count Models
Handling Interactions in Stata

- So far, when interacting two variables, we have done so manually before running a regression
  - Sometimes there can be many interactions, so doing so manually is not the best use of time.

- For example, consider the following:
  \[
  \log(wage_{it}) = \beta_0 + \beta_t manu_{it} + \alpha_i + \alpha_t + u_{it}
  \]

- Within individuals and years, the effects of being in a manufacturing industry can vary by year.
  - \(\beta_t\) is year specific.
  - This can be estimated in a number of ways, so let’s write them out.
Precisely, defining $d_t$ as a dummy variable identifying each year $t$, the equation is written as:

$$\log(wage_{it}) = \beta_0 + \sum_t \beta^m_t \cdot d_t \cdot manu_{it} + \sum_t \beta_t \cdot d_t + \alpha_i + u_{it}$$

To estimate this equation, use the interaction operator, #.

- `xtreg lwage manu#year, fe`
- `xtreg lwage manu##year, fe`

The first produces average effects of each manu-year pair relative to an outside group (the constant)

- Eg. Within individuals, working in manufacturing in 1986 earns X in relative to non-manufacturing in 1980.

The second technique produces something very similar to the written specification above.

- Test whether there are yearly differences between manufacturing and non-manufacturing.
Handling Interactions in Stata

- Interactions can also involve a continuous variable,
  \[
  \log(wage_{it}) = \beta_0 + \beta_1 manu_{it} + \beta_2 year_t + \beta_3 manu_{it} \cdot year_t + \alpha_i + u_{it}
  \]

- This is reasonably easy to program manually, but try the interaction operators to see the differences in Stata conventions when using a continuous variable.

- Must impose that a variable is continuous using 'c.varname'

  - xtreg lwage manu#c.year, fe
  - xtreg lwage manu##c.year, fe

- The first is an *incomplete interaction* specification.

  - This is also called "wrong".

  - Must instead run xtreg lwage manu manu#c.year, fe

- The second technique automatically includes the full interaction
Handling Interactions in Stata

- Continuous interactions are useful when trying to account for time effects at a very detailed level.

- Consider the following within-individual regression with individual-specific time trends.

  \[ \log(wage_{it}) = \beta_0 + \beta_1 manu_{it} + \beta_i \cdot year_t + \alpha_i + u_{it} \]

- Writing out in dummy variable form:

  \[ \log(wage_{it}) = \beta_0 + \beta_1 manu_{it} + \sum_i \beta_i d_i + \sum_i \beta^y_i d_i \cdot year_t + u_{it} \]

- Why should this regression be considered in the within-individual wage regression?

  - Individuals that are moving to manufacturing and earning higher wages might be doing so for some unobserved reason within the individual.

- `xtreg lwage manu nr#c.year, fe`
Recall the linear probability model:

\[ y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u \]

where \( y \) either equals 1 or 0.

Note that:

\[ \Pr(y = 1|x) = E(y|x) \]

The estimates:

\[ \Pr(y = 1|x) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k \]

Are there conditions on \( \hat{y} \)?

→ No restrictions on \( \hat{y} \).
Discrete Dependent Variables

- Best way does not use OLS
  \[
  \Pr(y = 1|x) = G (\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)
  \]
- \(G\) is a symmetric distribution of the unobservables:
  \[
  G(a) = \Pr(u < a)
  \]
- Framework derived from a "latent variable model"
  \[
  y^* = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u
  \]
  \[
  y = I(y^* > 0)
  \]
- Use latent model to derive \(\Pr(y = 1|x)\):
  \[
  \Pr(y = 1|x) = \Pr(y^* > 0)
  \]
  \[
  = \Pr(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u > 0)
  \]
  \[
  = \Pr(u > -(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k))
  \]
Probit-Logit

- Using the definition of $G(a)$, and that $G$ is symmetric:
  \[
  \Pr(y = 1|x) = 1 - G\left(-(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)\right)
  = G\left(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k\right)
  \]

- **Logit** uses the Logistic Distribution
  \[
  G(z) = \frac{\exp(z)}{1 + \exp(z)}
  \]

- **Probit** uses the standard normal distribution.
  \[
  G(z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v^2}{2}\right) dv
  \]

- These models require "maximum likelihood estimation".

- However, they bound the dependent variable, so produce theoretically sensible estimates.
When we estimate the $\beta$’s in the probit and logit models, this is not the marginal effect on the outcome, $y$.

It is the marginal effect on the latent variable.

To see this, differentiate $\Pr(y = 1|x)$ with respect to $x_1$:

$$\frac{d \Pr(y = 1|x)}{dx} = g \left( \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \right) \beta_1$$

Since we are bounding the dependent variable using a non-linear function, the marginal effect depends on all the estimates and their values.

It is typical to evaluate the marginal effect at the sample means:

$$\frac{d \Pr(y = 1|\bar{x})}{dx} = g \left( \beta_0 + \beta_1 \bar{x}_1 + \cdots + \beta_k \bar{x}_k \right) \beta_1$$
Probit-Logit
Smoking habits of expectant mothers

- Model:
  \[
  Pr(\text{Smoke}=1 \mid x) = G(\beta_0 + \beta_1 \text{faminc} + \beta_2 \text{motheduc} + \beta_3 \text{fatheduc})
  \]

- \text{Smoke}=1 \text{ if Cigs} > 0, \text{ otherwise 0}

- Estimate \( \beta_1, \beta_2, \beta_3 \), using 'probit' and 'logit'.

- Differentiate to find "marginal effects"
  \[
  \frac{\partial Pr(\text{Smoke}=1 \mid x)}{\partial \text{faminc}} = g \left( \beta_0 + \beta_1 \text{faminc} + \beta_2 \text{motheduc} + \beta_3 \text{fatheduc} \right) \beta_1
  \]

- To calculate marginal effects at the mean, use the command 'mfx' directly after 'probit' or 'logit'.
  \[
  \frac{\partial Pr(\text{Smoke}=1 \mid x)}{\partial \text{faminc}} = g \left( \beta_0 + \beta_1 \text{faminc} + \beta_2 \text{motheduc} + \beta_3 \text{fatheduc} \right) \beta_1
  \]
In a earlier example, we evaluated the effect of cigarette prices on cigarettes consumed using:

\[ Cigs = \beta_0 + \beta_1 \text{price} + \beta_2 \text{faminc} + u \]

There are two big problems with this:

- We’re regressing quantities on prices, where prices could be endogenous.
- \( Cigs \) is not bounded, but predictions could be negative.

We have a technique to deal with the latter: the Tobit model

The Tobit model is again built on a latent variable framework, but differs in that the outcome is not discrete.

\[ y^* = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u \]

\[ y = \max(y^*, 0) \quad , \quad u \sim \text{Normal} \left( 0, \sigma^2 \right) \]

Estimation is again by maximum likelihood.
Tobit

Though the details are left for advanced reading, the estimated $\beta$’s are for the latent variable specification, and not the expected value.

The marginal effect of a variable on the observed outcome can be written as:

$$\frac{\partial E(y|x)}{\partial x_i} = \beta_1 \Phi \left( \frac{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}{\sigma} \right)$$

- This is the coefficient estimate scaled by $\Pr(y > 0)$

Estimate the following model by OLS and Tobit

$$Cigs = \beta_0 + \beta_1 price + \beta_2 faminc + u$$

Use ’tobit’ command to compute parameters

Use ’mfx, pred(e(a, b))’, to compute marginal effects, where

$$\frac{\partial E(y|x, a < y < b)}{\partial x_k}$$
In the smoking example, Cigs is actually a count variable.

The normal distribution is not appropriate for count data, since it is continuous and the data are not.

A common count data model is the "Poisson Regression", which uses the (discrete) Poisson Distribution as it’s base:

\[
Pr (y = h) = \exp (-\lambda) \cdot \frac{\exp (\lambda)^h}{h!}
\]

where \( \lambda = E(y) \)

The distribution is completely specified by its mean, \( \lambda \).

It is common to use the exponential distribution to estimate the mean.

\[
E \left( y \mid x_1, \ldots, x_k \right) = \exp \left( \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \right)
\]

Does not give negative values, but in the limit includes zero.
Poisson Regression

- Estimate this equation using the 'poisson' command

\[ E(y|x_1, ..., x_k) = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) \]

- 100\( \beta_1 \) is approximately \( \% \Delta E(y|x_1, ..., x_k) \) given \( \Delta x_1 = 1 \).

- If you’re feeling adventurous, defining \( x\beta = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \), you can calculate the probability of each count:

\[ \Pr(y = h|x) = \exp(-\exp(x\beta)) \cdot \frac{\exp(x\beta)^h}{h!} \]