Using Binary Graph

Michael Tierney*

December 5, 2012

Purpose

I wrote the program binary_graph because I was running probit models with interaction terms and or I was squaring a continuous variable of interest. I started out using CLARIFY to estimate marginal effects of the variables at their means, but I ended up using the code from Matt Golder's website (https://files.nyu.edu/mrg217/public/interaction.html) associated with Brambor, Clark, and Golder (2006) to graph the marginal effects.

While the Golder code is great, it involved considerable work because the marginal effects of non-linear models (such as logit and probit) depend on the value of the other variables in the model. For example, say I was interested in the following model of civil war onset with the following interaction term:

$$Pr(Union = 1) = F(\beta_0 + \beta_1 * age + \beta_2 * grade + \beta_3 * age * grade)$$

I could enter the appropriate values into the Golder code and get a nice graph. The problem is that if I wanted to add in a different control, say the following model:

$$Pr(Union = 1) = F(\beta_0 + \beta_1 * age + \beta_2 * grade + \beta_3 * age * grade + \beta_4 * not\_smsa)$$

I would have to go into the Golder code and enter in information for my new variable. Given that social scientists are often running many regressions with many different controls, I wanted an easier way, hence this program. With this program, one can simply type 'binary_graph interaction ci_95' after a probit or logic model to produce a graph of the marginal effect of age as grade changes. Any additional control variables in the model are held at their medians.[1]

In order to use the program, simply put binary_graph.ado in the appropriate ado folder.[2] Run your logit or probit with the syntax from the following example:

*michael.tiernay@nyu.edu

[1] Common statistical convention is to hold variables at their mean. The problem with doing so is if the control variables are binary or categorical. For example, if one is using binary variable in the model, calculating marginal effect at the mean of the control variables means calculating the marginal effect for an observation with a level of .45 (or whatever the mean happens to be), which doesn't make sense because we are using a binary variable. Thus, I choose to calculate the marginal effect at the medians of the controls, which will be the same as the means if the distribution of the control variables is normal.

[2] For mac users, go to Applications ¿ Stata ¿ ado ¿ updates ¿ b

For models with an interaction term

```
webuse union, clear

*Generate the interaction term
gen age_grade = age*grade

*Run the probit
probit union age grade age_grade not_smsa

*Run the binary_graph command
binary_graph interaction ci_95
```

The preceding lines of code will produce a graph of the marginal effect of age as grade changes along with a 95% confidence interval, see Figure 1 below. ‘ci_90’ can be specified instead of ‘ci_95’ if the analyst wishes to use a 90% confidence interval. The order of the variables in the probit model is crucial to producing the correct graph. In the example ‘union’ is the dependent variable, ‘age’ is our variable of interest, ‘grade’ is our modifying variable, and ‘age_grade’ is the interaction of the two, followed by all the control variables. The order: dependent variable, variable of interest, modifying variable, and interaction term must be kept. The order of the control variables after this does not matter.

For models with squared variable

```
webuse union, clear

*Generate the interaction term
gen age_2 = age*2

*Run the probit
probit union age age_2 grade not_smsa

*Run the binary_graph command
binary_graph square ci_90
```

The preceding lines of code will produce a graph of the marginal effect of age as age changes along with a 90% confidence interval, see Figure 2 below. The order of the variables in the probit model is crucial to producing the correct graph. In the example ‘union’ is the dependent variable, ‘age’ is our variable of interest, ‘age_2’ is its square, followed by all the control variables. The order: dependent variable, variable of interest, square of the variable of interest must be kept. The order of the control variables after this does not matter.

What is Being Estimated

See Greene (6th Edition) p. 775 for the following information.

The marginal effect of a probit model is:

$$\frac{\partial E[y|x]}{\partial x} = \phi(x'\beta)\beta$$

where ϕ is the standard normal density. For the examples above, we are estimating:

$$\phi(x'\beta)(\beta_1 + \beta_3 * grade)$$

for the model with an interaction term and:

$$\phi(x'\beta)(\beta_1 + \beta_2 * age)$$

for the model with age squared.

Everything is the same with the logit model, except that $\phi(x'\beta)$ is substituted with

$$\Delta(x'\beta) * (1 - \Delta(x'\beta))$$

where

$$\Delta(x'\beta) = \frac{e^{x'\beta}}{1 + e^{x'\beta}}$$

Finally, as discussed above $(x'\beta)$ is calculated using the medians of the independent variables.

Figure 1: Marginal Effect with an Interaction Term

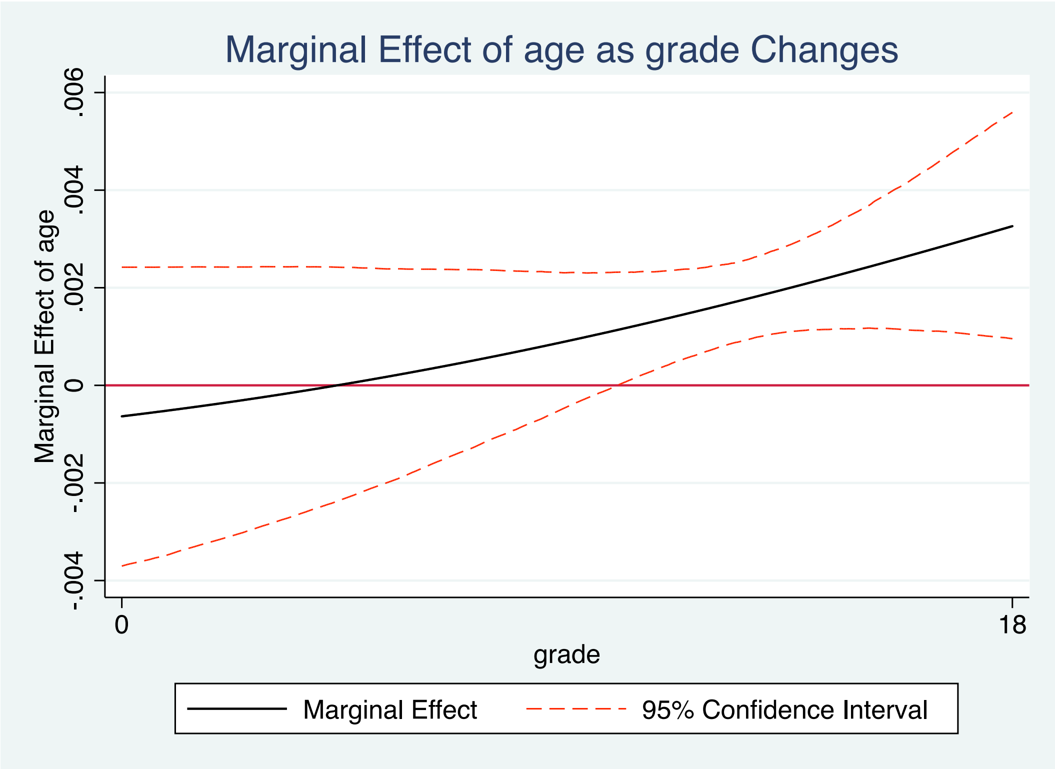

Figure 2: Marginal Effect with Squared Variable