Detailed Explanation of **Stata** Code for a Marginal Effect Plot for X

Below, I go through the **Stata** code for creating a marginal effect plot for X for an interaction model with the following basic form:

$$Y = \beta_0 + \beta_1 X + \beta_2 W + \beta_3 Z$$
$$+ \beta_4 XW + \beta_5 XZ + \beta_6 WZ$$
$$+ \beta_7 XWZ + \epsilon. \tag{1}$$

```
version 11.0
#delimit ;
log using "C:\matt\publications\jop2\webpage\interaction2.log", replace;
set more off;
```

The first line simply indicates the version of **Stata** that is being used. The second line tells **Stata** that all command lines are finished with a semi-colon. The third line opens a log file and indicates where it will be saved. The fourth line tells **Stata** not to pause when running the do-file.

```
*       ************************************************************ *;
*       ************************************************************ *;
*       File-Name:      interaction2.do                            *;
*       Date:           1/7/2013                                   *;
*       Author:         MG                                         *;
*       Purpose:        Do-file to produce a marginal effect plot for *;
*                       X when the interaction model is:            *;
*                       Y = b0 + b1X + b2W + b3Z + b4XW + b5XZ + b6WZ *;
*                       b7XZW + e                                   *;
*       Input File:     interaction2.dta                           *;
*       Output File:    interaction2.log                           *;
*       Data Output:    None                                       *;
*       Previous file:  None                                       *;
*       Machine:        laptop                                     *;
*       ************************************************************ *;
*       ************************************************************ *;
```

This part of the code is how I like to start all do-files and is not strictly necessary to create the sample marginal effect plot. It indicates the name of the do-file, the date on which it was written, the author, the purpose of the do-file, the data used with the do-file, the name of the associated log file, the name of any new data set that is created by the do-file, the name of any previous do-file or data set, and where the do-file was run.

```
use "C:\matt\publications\jop2\webpage\interaction2.dta", clear;
```

This tells **Stata** what data set to open and where to find it. It also clears any existing data in **Stata** 's memory.

```
set obs 1000;
```

This command sets the number of observations (rows) in the data set to 1,000. There is nothing special about this number. You will want to calculate the marginal effect of X at multiple values of W and Z. If you want to calculate the marginal effect at more values than you actually have in your original data set, then you will need to increase the number of observations in the data set. This is precisely what this line of code does.

```
regress  legislative_parties fragmentation concentration logmagnitude
         fragmentation_concentration logmagnitude_fragmentation
         logmag_concentration frag_conc_logmag
         proximity presidential_candidate proximity_presidential_candidate, robust;
```

This line estimates the chosen model. In this case, legislative_parties is the dependent variable, Y, fragmentation is our X, concentration is our W, logmagnitude is our Z, fragmentation_concentration is our XW, logmagnitude_fragmentation is our XZ, logmag_concentration is our WZ, frag_conc_logmag is our XWZ, proximity, presidential_candidate, proximity_presidential_candidate are control variables, and robust indicates the use of robust standard errors.

```
matrix b=e(b);
matrix V=e(V);

scalar b1=b[1,1];
scalar b2=b[1,2];
scalar b3=b[1,3];
scalar b4=b[1,4];
scalar b5=b[1,5];
scalar b6=b[1,6];
scalar b7=b[1,7];

scalar varb1=V[1,1];
scalar varb2=V[2,2];
scalar varb3=V[3,3];
scalar varb4=V[4,4];
scalar varb5=V[5,5];
scalar varb6=V[6,6];
scalar varb7=V[7,7];

scalar covb1b4=V[1,4];
scalar covb1b5=V[1,5];
scalar covb1b7=V[1,7];
scalar covb4b5=V[4,5];
scalar covb4b7=V[4,7];
scalar covb5b7=V[5,7];

scalar list b1 b2 b3 b4 b5 b6 b7 varb1 varb2 varb3 varb4 varb5 varb6 varb7
            covb1b4 covb1b5 covb1b7 covb4b5 covb4b7 covb5b7;
```

The first two lines grab the elements of the estimated coefficient vector and the variance-covariance matrix that are necessary to calculate the marginal effects and indicate statistical significance in the marginal effect plot. The estimated coefficient vector, b, is $1 \times k$, where k is the number of covariates including the constant term. The first element in b, b[1,1], is β_1, the second element, b[1,2] is β_2, and so on. The last element in b is the coefficient on the constant term, β_0. The estimated variance-covariance matrix is V. The diagonal elements indicate the variances of the coefficients, while the off-diagonal elements indicate the covariances

of the coefficients. For example, the element `V[1,1]` indicates the variance of β_1 and the element `V[1,4]` indicates the covariance between β_1 and β_4. The remaining lines of code create (and then list) scalars that are equal to the coefficients, variances, and covariances that are needed to calculate the desired marginal effects and standard errors. You can check the listed scalars with the regression output to make sure that you have done everything correctly.

```
*        ***************************************************************   *;
*           Calculate data necessary for top marginal effect plot.         *;
*        ***************************************************************   *;

generate MVZ=((_n-1)/10);
replace MVZ=. if _n>30;
```

The first line generates a variable `MVZ` that takes on all of the values of the modifying variable Z, `logmagnitude`, for which you want to calculate the marginal effect of X; the more values, the smoother your marginal effect line will be. `_n` in **Stata** indicates the observation number starting at 1. If your modifying variable starts at 1 and increases in increments of 1, then you could simply type generate `generate MVZ=_n;`. If your modifying variable starts at 0 and increases in increments of 1, then you would need to start at `_n-1`, and so you would type generate `MVZ=_n-1;`. If you have a continuous modifying variable that starts at 0 and you want to calculate the marginal effect of X at each 0.1 increment in the modifying variable Z, then you would type generate `MVZ=((_n-1)/10);`. The modifying variable Z, `logmagnitude`, in this specific example starts at 0, and I would like to calculate the marginal effect of X at each increment of 0.1 in the modifying variable Z: generate `MVZ=((_n-1)/10)`.

The second line determines the last value of the modifying variable Z for which you want to calculate the marginal effect of X. `replace MVZ=. if _n>30` indicates that the highest value of Z for which I wish to calculate the marginal effect of X is 3.

```
*        ***************************************************************   *;
*           We want to calculate the marginal effect of X on Y for all      *;
*           values of the modifying variable Z. We also want to calculate    *;
*           this marginal effect as Z changes for specific values of the     *;
*           second modifying variable W.  In the code below, we calculate    *;
*           the marginal effect of X on Y for all values of Z when W=0,      *;
*           when W=1, when W=2, and when W=3.                                 *;
*        ***************************************************************   *;

gen conb0=b1+b4*0+b5*MVZ+b7*(0*MVZ) if _n<30;
gen conb1=b1+b4*1+b5*MVZ+b7*(1*MVZ) if _n<30;
gen conb2=b1+b4*2+b5*MVZ+b7*(2*MVZ) if _n<30;
gen conb3=b1+b4*3+b5*MVZ+b7*(3*MVZ) if _n<30;
gen conb4=b1+b4*4+b5*MVZ+b7*(4*MVZ) if _n<30;
```

These lines calculate the marginal effect of X on Y for the desired values of the modifying variables W and Z. The lines calculate the marginal effect of X for $W = \{0, 1, 2, 3\}$ and for all values of Z. The marginal effect of X for an interactive model like that shown in Eq. 1 is:

$$\frac{\partial Y}{\partial X} = \beta_1 + \beta_4 W + \beta_5 Z + \beta_7 ZW. \tag{2}$$

```
*       ***************************************************************   *;
*       Calculate the standard errors for the marginal effect of X on Y *;
*       ***************************************************************   *;

gen conse0=sqrt(varb1
                + varb4*(0^2) + varb5*MVZ^2 + varb7*MVZ^2*(0^2)
                + 2*0*covb1b4 + 2*MVZ*covb1b5 + 2*0*MVZ*covb1b7+2*0*MVZ*covb4b5
                + 2*(0^2)*MVZ*covb4b7 + 2*0*(MVZ^2)*covb5b7)  if _n<30;

gen conse1=sqrt(varb1
                + varb4*(1^2) + varb5*MVZ^2 + varb7*MVZ^2*(1^2)
                + 2*1*covb1b4 + 2*MVZ*covb1b5 + 2*1*MVZ*covb1b7+2*1*MVZ*covb4b5
                + 2*(1^2)*MVZ*covb4b7 + 2*1*(MVZ^2)*covb5b7)  if _n<30;

gen conse2=sqrt(varb1
                + varb4*(2^2) + varb5*MVZ^2 + varb7*MVZ^2*(2^2)
                + 2*2*covb1b4 + 2*MVZ*covb1b5 + 2*2*MVZ*covb1b7+2*2*MVZ*covb4b5
                + 2*(2^2)*MVZ*covb4b7 + 2*2*(MVZ^2)*covb5b7)  if _n<30;

gen conse3=sqrt(varb1
                + varb4*(3^2) + varb5*MVZ^2 + varb7*MVZ^2*(3^2)
                + 2*3*covb1b4 + 2*MVZ*covb1b5 + 2*3*MVZ*covb1b7+2*3*MVZ*covb4b5
                + 2*(3^2)*MVZ*covb4b7 + 2*3*(MVZ^2)*covb5b7)  if _n<30;

gen conse4=sqrt(varb1
                + varb4*(4^2) + varb5*MVZ^2 + varb7*MVZ^2*(4^2)
                + 2*4*covb1b4 + 2*MVZ*covb1b5 + 2*4*MVZ*covb1b7+2*4*MVZ*covb4b5
                + 2*(4^2)*MVZ*covb4b7 + 2*4*(MVZ^2)*covb5b7)  if _n<30;
```

These lines calculate the standard errors for the marginal effect of X on Y for the desired values of the modifying variables Z and W. The standard error for the marginal effect of X for an interactive model like that shown in Eq. 1 is the square root of:

$$\begin{aligned}
\mathrm{var}(\beta_1) &+ W^2\mathrm{var}(\beta_4) + Z^2\mathrm{var}(\beta_5) + Z^2W^2\mathrm{var}(\beta_7) \\
&+ 2W\mathrm{cov}(\beta_1\beta_4) + 2Z\mathrm{cov}(\beta_1\beta_5) + 2WZ\mathrm{cov}(\beta_1\beta_7) + 2WZ\mathrm{cov}(\beta_4\beta_5) \\
&+ 2W^2Z\mathrm{cov}(\beta_4\beta_7) + 2WZ^2\mathrm{cov}(\beta_5\beta_7).
\end{aligned} \tag{3}$$

```
*       ***************************************************************   *;
*                       Create t statistics                              *;
*       ***************************************************************   *;

gen t0=conb0/conse0;
gen t1=conb1/conse1;
gen t2=conb2/conse2;
gen t3=conb3/conse3;
gen t4=conb4/conse4;
```

These lines calculate t-statistics for our marginal effects. We will use these to determine when the marginal effect is statistically significant at the desired level of significance.

```
*       ***************************************************************   *;
*       Generate a 'shadow' variable that is equal to the marginal       *;
*       effects that have already been calculated.  These will be        *;
*       helpful in placing the stars in the figure as you will see.      *;
```

```
*       ****************************************************************   *;
gen consb0=conb0;
gen consb1=conb1;
gen consb2=conb2;
gen consb3=conb3;
gen consb4=conb4;
```

We are going to place stars when the marginal effect is statistically significant. These lines just create what "shadow" marginal effects. Their use will become obvious in a moment.

```
*       ****************************************************************   *;
*          Make the 'shadow' variable be missing if the t score is not    *;
*          larger than the critical level of significance that you want.   *;
*       ****************************************************************   *;

replace consb0 = . if abs(t0)<2.01;
replace consb1 = . if abs(t1)<2.01;
replace consb2 = . if abs(t2)<2.01;
replace consb3 = . if abs(t3)<2.01;
replace consb4 = . if abs(t4)<2.01;
```

These lines replace "shadow" marginal effects as missing if they are not statistically significant at the 95% level. In other words, make them missing if the t-statistic is not larger than the critical value for statistical significance. Given the sample size and degrees of freedom in this particular example, the critical value is 2.01 for a test with level $\alpha = 0.05$. You will have to change this critical value depending on your degrees of freedom and desired level of α. Now we are left with only statistically significant "shadow" marginal effects.

```
*       ****************************************************************   *;
*          Generate a string variable called str1 that is designated with *;
*          a star.                                                         *;
*       ****************************************************************   *;

generate str1 txt="*";

set textsize 100;
```

Since I want to use a star to indicate statistical significance (to mirror how we typically represent statistical significance in traditional tables of results), I need to create a string variable that is a star. This is what the first line does. The second line simply specifies the text size.

```
*       ****************************************************************   *;
*          Graph the effect of fragmentation on legislative parties       *;
*          conditional on logmag                                          *;
*       ****************************************************************   *;

graph twoway line conb0 MVZ, clpattern(solid) clwidth(thin)
        || scatter consb0 MVZ, mlabel(txt) msymbol(i) mlabsize(vsmall) mlabgap(-1.0) mlabposition(11)
        || line conb1 MVZ, clpattern(solid) clwidth(thin)
        || scatter consb1 MVZ, mlabel(txt) msymbol(i) mlabsize(vsmall) mlabgap(-1.0) mlabposition(11)
        || line conb2 MVZ, clpattern(solid) clwidth(thin)
        || scatter consb2 MVZ, mlabel(txt) msymbol(i) mlabsize(vsmall) mlabgap(-1.0) mlabposition(11)
```

```
|| line conb3 MVZ, clpattern(solid) clwidth(thin)
|| scatter consb3 MVZ, mlabel(txt) msymbol(i) mlabsize(vsmall) mlabgap(-1.0) mlabposition(11)
|| ,
xlabel(0 0.5 1 1.5  2 2.5 3, nogrid labsize(2.5))
ylabel(-0.4 -0.2 0 0.2 0.4 0.6 0.8, nogrid labsize(2.5))
yscale(noline)  xscale(noline)  legend(off)
yline(0, lcolor(black))
xtitle("Logged Average District Magnitude")
ytitle("Marginal Effect of Fragmentation")
xsca(titlegap(4)) ysca(titlegap(4))
scheme(s2mono) graphregion(fcolor(white) ilcolor(white) lcolor(white));
```

The first eight lines of code plot the marginal effect of X on Y for $0 \leq Z \leq 3$ when $W = \{0, 1, 2, 3\}$. These lines also overlay stars to indicate when these effects are statistically significant. The remaining lines help to format the marginal effect plot.