PLSC 503: "Multivariate Analysis for Political Research"

Multiplicative Interaction Terms
March 28, 2017

Introduction

Let's start with our typical K-variable regression model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

where the coefficient vector $\mathbf{X}$ might contain some mixture of continuous and discrete (dummy) variables. So far, we have assumed that each of the variables enters only linearly – in other words, this regression assumes that the impact of any given X on Y is the same irrespective of the value of all the other $\mathbf{X}$s. In fact, there are many, many, many instances where this assumption is untenable, and where the impact of one element of $\mathbf{X}$ on $\mathbf{Y}$ is contingent on the value(s) of some other element(s) in $\mathbf{X}$. Consider just three of many examples.

First, as we discussed previously, a model that has two dichotomous covariates

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + u_i$$

can be thought of as one in which there are four expected values of $\mathbf{Y}$ (corresponding to each of the four possible covariate patterns on the two $\mathbf{X}$ variables), but with the restriction that the differences between those expectations when one such covariate is held constant are the same. That is, the model requires that

$$E(Y|D_1 = 1, D_2 = 0) - E(Y|D_1 = 0, D_2 = 0) = E(Y|D_1 = 1, D_2 = 1) - E(Y|D_1 = 0, D_2 = 1)[\equiv \beta_1]$$

and, similarly, that

$$E(Y|D_1 = 0, D_2 = 1) - E(Y|D_1 = 0, D_2 = 0) = E(Y|D_1 = 1, D_2 = 1) - E(Y|D_1 = 1, D_2 = 0)[\equiv \beta_2]$$

What if, however, the difference between the expected value of $Y|D_1 = 0, D_2 = 0$ and $Y|D_1 = 0, D_2 = 1$ is different than the difference between the expected values of $Y|D_1 = 1, D_2 = 0$ and $Y|D_1 = 1, D_2 = 1$?

- More concretely, what is the expected impact of (say) the Democratic Party endorsing a candidate in an open primary on the probability that a given voter will vote for him or her?
    - The answer is, "It depends on whether that voter is a Democrat or a Republican."
    - If the former, the effect is positive,

  - If the latter, the effect is negative.

  Or, think about the effect on refugee outflows (Y) of the presence or absence of a civil war in the "source" country. We would expect this effect to be positive; however, basic geography (e.g., Zipf's Law) suggests that the expected number of refugees fleeing to a second country will also be a function of whether or not that country is contiguous with the host country: civil wars will increase refugee flows between two countries, but the effect will be far greater if the host country is right next door.

The same is true for continuous variables: For example, what is the influence of the level of union membership on the degree of social welfare support in a country? The answer likely depends on how liberal or conservative the government in question is: more conservative governments need pay little attention to union demands for (e.g.) higher wages, so the effect will be small; in contrast, more left/liberal governments (who rely on unions for electoral support) will be far more responsive to unions' calls for greater levels of social protections for workers.

Finally, think about the relationship between the level of political repression in a country and the number of coup attempts it experiences.

- We might ask what the impact of a one-unit increase in repression on the likelihood of a coup is; however,
- That effect likely depends on the level of repression currently in the country.
  - In a highly repressive society, increasing repression likely has the effect of quashing coup planning and attempts; on the other hand,
  - In a non-repressive society, increasing repression might cause resentment, thereby increasing the risk of a coup.
- In other words, the relationship between coups and repression is *curvilinear*; increases in repression first increase, then decrease the risk of a coup.
- This can be thought of as a particular kind of interaction as well, in that a variable X's impact on Y depends on the level of X itself.

Models where the impact of a particular covariate on the response variable varies as a function of one or more other covariates can be thought of generally as *interactive models*. If considered properly, this can include a broad range of possible specifications; I'll talk first about a general case, and work through some of the math associated with interaction effects, and then get into some more specific examples, including different combinations of "types" of variables which can interact, higher-order polynomial effects, and so forth. On Tuesday , we'll get into the practical side of interaction terms.

A General Overview of Interaction Effects

Consider first a very simple interactive model with three variables:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + u_i \tag{1}$$

Here, $X_1 X_2$ is just that: X_1 multiplied by X_2. I'll generally (and somewhat incorrectly) refer to β_1 and β_2 as the *direct effects* of X_1 and X_2 on Y,[1] and to β_3 as the coefficient for the *interaction term.*

Why multiply the two variables together? To answer that, we need to think about the net impact of X_1 on Y in (1):

$$\begin{aligned} \mathrm{E}(\mathrm{Y}_i) &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} \\ &= \beta_0 + \beta_2 X_{2i} + (\beta_1 + \beta_3 X_{2i}) X_{1i} \end{aligned} \tag{2}$$

If we take the first derivative of this with respect to X_1, we get:

$$\frac{\partial \mathrm{E}(\mathrm{Y}_i)}{\partial X_1} = \beta_1 + \beta_3 X_{2i}. \tag{3}$$

In other words, the marginal impact of X_1 on Y now explicitly depends on the value of X_2. In this light, it is sometimes useful to think of (2) as:

$$\mathrm{E}(Y_i) = \beta_0 + \beta_2 X_{2i} + \psi_1 X_{1i} \tag{4}$$

where $\psi_1 = \beta_1 + \beta_3 X_{2i}$ is a "quasi-coefficient" for the marginal impact of X_1 on Y. We can write the same equation in terms of X_2:

$$\begin{aligned} \mathrm{E}(Y_i) &= \beta_0 + \beta_1 X_{1i} + (\beta_2 + \beta_3 X_{1i}) X_{2i} \\ &= \beta_0 + \beta_1 X_{1i} + \psi_2 X_{2i} \end{aligned} \tag{5}$$

In such an interactive model, *each interacted covariate's influence on Y is now conditional on the values of the other X(s) with which it is interacted.* That means that **the coefficients now have conditional – rather than marginal – interpretations**. In particular,

- The direct effects β_1 and β_2 represent the *conditional* impact of X_1 and X_2 respectively on Y, *conditional on the value of the other interacted variable being zero.*
    - Think about it: If (say) $X_2 = 0$, then (2) becomes

[1] Brambor et al. (2006) refer to these variables as "constitutive terms," but the meaning is the same.

$$
\begin{aligned}
\mathrm{E}(Y_i) &= \beta_0 + \beta_1 X_{1i} + \beta_2(0) + \beta_3 X_{1i}(0) \\
&= \beta_0 + \beta_1 X_{1i}
\end{aligned}
$$

And the same holds if $X_1 = 0$:

$$
\begin{aligned}
\mathrm{E}(Y_i) &= \beta_0 + \beta_1(0) + \beta_2 X_{2i} + \beta_3(0) X_{2i} \\
&= \beta_0 + \beta_2 X_{2i}
\end{aligned}
$$

- In that circumstance, $\psi_1 = \beta_1$ and $\psi_2 = \beta_2$.
- Thus, the signs of $\hat{\beta}_1$ and $\hat{\beta}_2$ are relevant, but only for a very specific circumstance – that where the value of the interacted variable is zero.

- The interaction coefficient β_3 indicates whether or not the effect of X_1 on Y is systematically (and monotonically/linearly) different over different values of X_2, and vice-versa.
  - A positive value of β_3 means that the impact of X_1 on Y grows more positive (/ less negative) at larger values of X_2, and less positive (/ more negative) at lower values of X_2.
  - Likewise, a positive value of β_3 means that the impact of X_2 on Y grows more positive (/ less negative) at larger values of X_1, and less positive (/ more negative) at lower values of X_1.
  - The reverse is true in both instances if $\beta_3 < 0$.
- In most instances, the quantities we now care about are not β_1 and β_2, but rather ψ_1 and ψ_2, which can be thought of as the *marginal* influence of X_1 and X_2 on Y, respectively.

The upshot of this discussion is that **in an interactive model, one can no longer talk about purely marginal effects; instead, *all* influences of interacted variables are now necessarily conditional on the value(s) of the variable(s) with which they are interacted.** I will be hammering this into your brains a lot over the next couple days.

Inference

The model in (1) is, despite the interacted covariates, still a standard linear regression model; as such, all of the formulae we learned for OLS regression will "work" just fine. That means that we can obtain estimates of the $\boldsymbol{\beta}$s in the standard way, and that we can get estimated standard errors for those coefficient estimates in the usual way, as the square roots of the diagonal elements of $\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$.

However, as noted above, those quantities are now no longer what we generally want to know. Instead, we're usually interested in the marginal effects; that is, on ψ_1 and/or ψ_2. In principle, we can combine values of $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ to get point estimates of ψ_1 and ψ_2; thus,

$$\hat{\psi}_1 = \hat{\beta}_1 + \hat{\beta}_3 X_2 \tag{6}$$

and

$$\hat{\psi}_2 = \hat{\beta}_2 + \hat{\beta}_3 X_1 \tag{7}$$

However, we would also usually like to conduct inference on these quantities, to assess whether (e.g.) the marginal effect of X_1 is statistically differentiable from some value of interest (like zero). To do so requires that we estimate the variance of $\hat{\psi}_1$ and $\hat{\psi}_2$; as both Freidrich (1982) and Brambor et al. (2006) note, we can do so by noting that, in general,

$$\text{Var}(a + bZ) = \text{Var}(a) + Z^2\text{Var}(b) + 2Z\text{Cov}(a, b)$$

So, if we are interested in $\text{Var}(\hat{\psi}_1)$, we can estimate this quantity as:

$$\widehat{\text{Var}(\hat{\psi}_1)} = \widehat{\text{Var}(\hat{\beta}_1)} + X_2^2\widehat{\text{Var}(\hat{\beta}_3)} + 2X_2\widehat{\text{Cov}(\hat{\beta}_1, \hat{\beta}_3)}. \tag{8}$$

We can use a similar equation for the estimated variance of $\hat{\psi}_2$:

$$\widehat{\text{Var}(\hat{\psi}_2)} = \widehat{\text{Var}(\hat{\beta}_2)} + X_1^2\widehat{\text{Var}(\hat{\beta}_3)} + 2X_1\widehat{\text{Cov}(\hat{\beta}_2, \hat{\beta}_3)}. \tag{9}$$

Notice a few things about the formulae in (8) and (9):

- Both depend on – but are not the same as – the estimated variances for the direct effects β_1 and β_2, and/or the interaction term β_3.

- Both also depend on the level/value of the interacted variable; that means that, just as the "slope" of X_1 on Y depends on X_2, so too does the estimated variability around that slope estimate also depend on X_2.

We'll discuss these issues in more detail below, and get into how, practically speaking, one estimates, presents, and discusses these effects on Thursday.

Types of Interactions: Conceptual Matters

For the rest of the class, we'll talk about models with interaction terms of various "kinds;" these will include interactions of covariates at various levels of measurement, as well as some other "special" cases (e.g., quadratic terms, higher-order interactive models, etc.).

Two Dichotomous Covariates

Consider a very simple model with two dichotomous covariates:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{1i} D_{2i} + u_i \tag{10}$$

In this model, the expected value of Y can take on four possible values:

$$\begin{aligned}
\mathrm{E}(Y|D_1 = 0, D_2 = 0) &= \beta_0 \\
\mathrm{E}(Y|D_1 = 1, D_2 = 0) &= \beta_0 + \beta_1 \\
\mathrm{E}(Y|D_1 = 0, D_2 = 1) &= \beta_0 + \beta_2 \\
\mathrm{E}(Y|D_1 = 1, D_2 = 1) &= \beta_0 + \beta_1 + \beta_2 + \beta_3
\end{aligned}$$

In each case, the expectation corresponds to the expected value (mean) of Y for the subset of observations defined by the combination of values on D_1 and D_2. As such, it is clear to see that the model in (10) is a less-restricted of the general two-variable model:

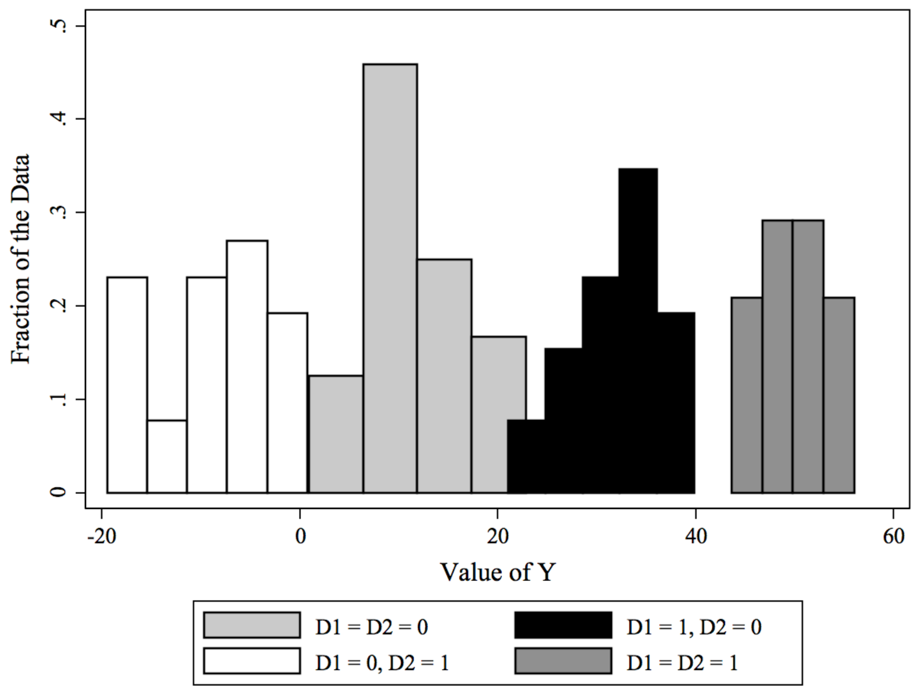
$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + u_i \tag{11}$$

where the influence of a change in D_1 (D_2) is allowed to vary depending on the value of D_2 (D_1). Graphically, we can see this in a couple ways; one is simply through a histogram of values of Y for the four possible values of D_1 and D_2, as in Figure 1:[2]

A second (better) alternative is to graph boxplots of the values of Y for each of the different covariate patterns:

[2]The data in these figures are simulated with $N = 100$; for the two dummy-variable examples, $Y_i = 10 + 20D_{1i} - 20D_{2i} + 40D_{1i}D_{2i} + u_i$, where $u_i \sim N(0, 5)$.

Figure 1: Values of Y for Various Combinations of Values of D_1 and D_2



This illustrates nicely that:

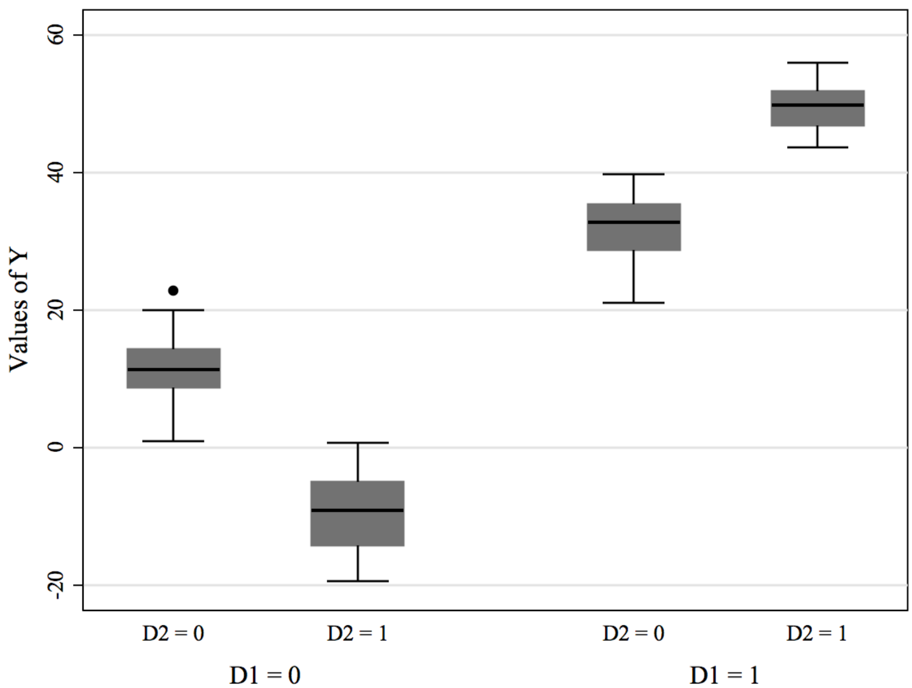
- The effect of a change in D_2 varies depending on the value of D_1:
  - That effect is negative when $D_1 = 0$, but
  - Positive when $D_1 = 1$.
- Similarly, the effect of a change in D_1 varies depending on the value of D_2: while both are positive, the effect is much larger when $D_2 = 1$ than when $D_2 = 0$.

Note two other related things about the two-dummy interactive model:

- Assuming you've gone with a dummy coding for the two Ds, the direct effects $\hat{\beta}_1$ and $\hat{\beta}_2$ have natural interpretations as the conditional differences in the means of Y.
- Similarly, the test of whether $\hat{\beta}_3 = 0$ amounts to a direct test of the restriction that leads to the model in (11); in this model only, then, is direct interpretation of the "statistical significance" of $\hat{\beta}_3$ a good idea.

The larger point here is that, with two interacted dummy variables, the model is a very simple one, and straightforward to interpret.

Figure 2: Boxplots of Y for Various Combinations of Values of D_1 and D_2



One Dichotomous and One Continuous Covariate

A more common situation is where we have an interaction of a dichotomous covariate with a continuous one:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i D_i + u_i \tag{12}$$

In this model, we can think of the expected value of Y as:

$$\begin{aligned} \mathrm{E}(Y|X, D = 0) &= \beta_0 + \beta_1 X \\ \mathrm{E}(Y|X, D = 1) &= (\beta_0 + \beta_2) + (\beta_1 + \beta_3)X \end{aligned}$$

This suggests four possibilities:

- Vis-a-vis Y, X has both the same slope and the same intercept for $D = 0$ and $D = 1$; that is, $\beta_2 = \beta_3 = 0$.
- Vis-a-vis Y, X has the same slope but different intercepts for $D = 0$ and $D = 1$; that is, $\beta_2 \neq 0$ and $\beta_3 = 0$.
- Vis-a-vis Y, X has the same intercept, but different slopes for $D = 0$ and $D = 1$; that is, $\beta_2 = 0$ and $\beta_3 \neq 0$.

- Finally, X's effect on Y can have both different slopes and different intercepts for $D = 0$ and $D = 1$; so, $\beta_2 \neq 0$ and $\beta_3 \neq 0$.

Graphically, we can represent these four possibilities as:

Figure 3: Scatterplot and Regression Lines of Y on X for $D = 0$ and $D = 1$: No Slope or Intercept Differences ($\beta_2 = \beta_3 = 0$)

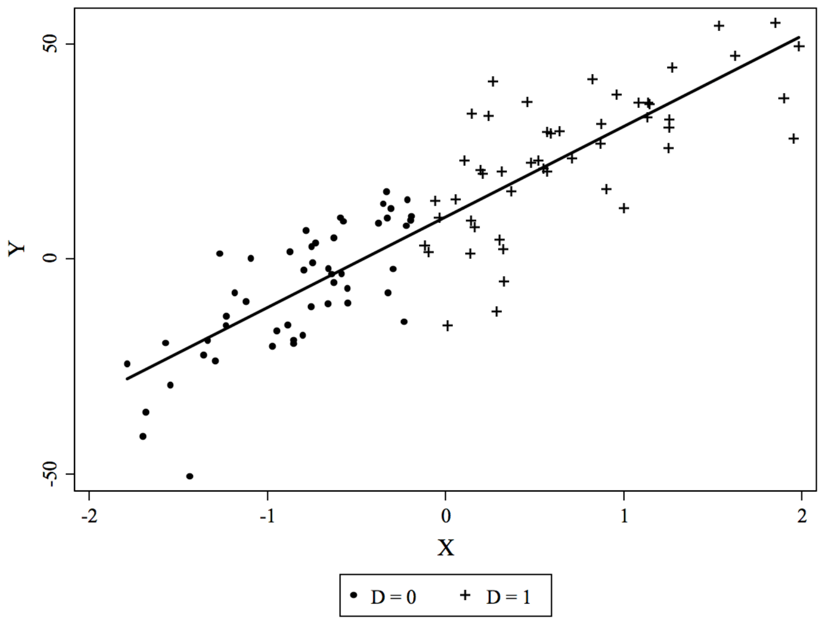


Figure 4: Scatterplot and Regression Lines of Y on X for $D = 0$ and $D = 1$: Intercept Shift ($\beta_2 \neq 0$, $\beta_3 = 0$)

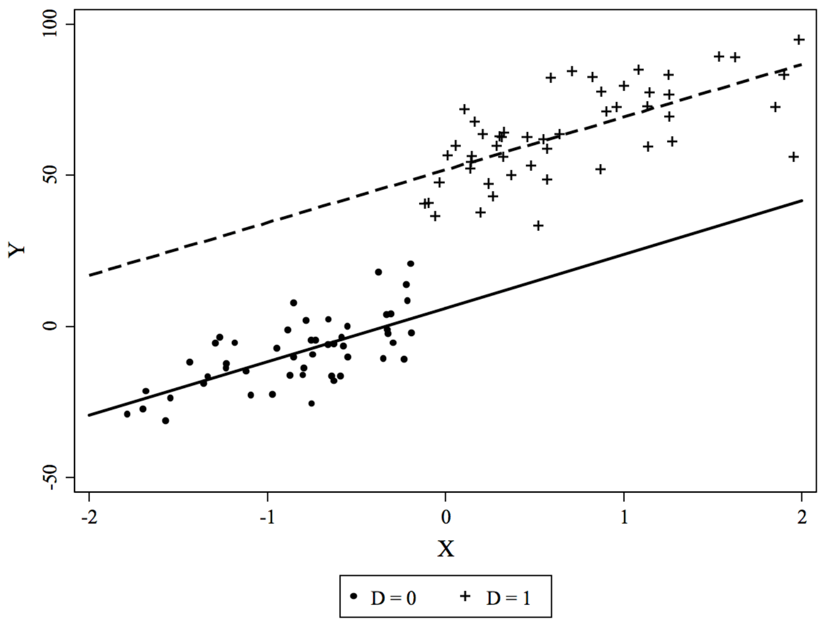


Figure 5: Scatterplot and Regression Lines of Y on X for $D = 0$ and $D = 1$: Slope Change ($\beta_2 = 0$, $\beta_3 \neq 0$)

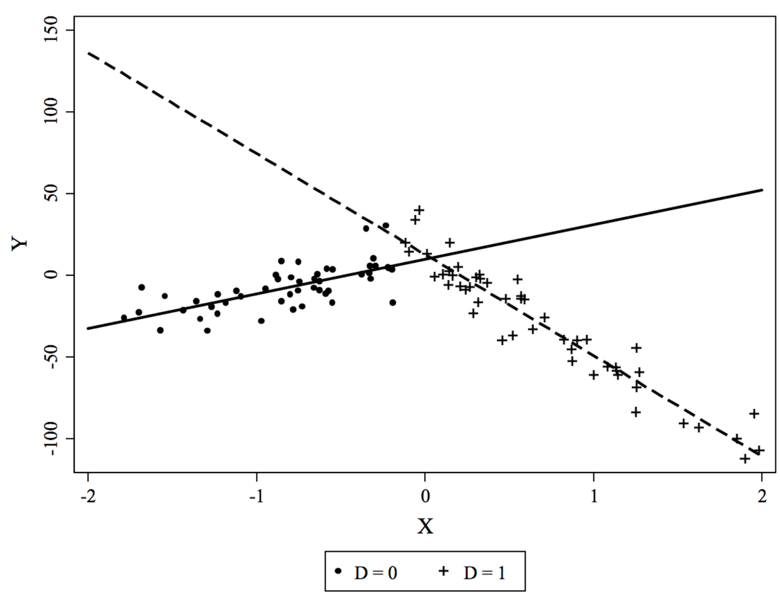
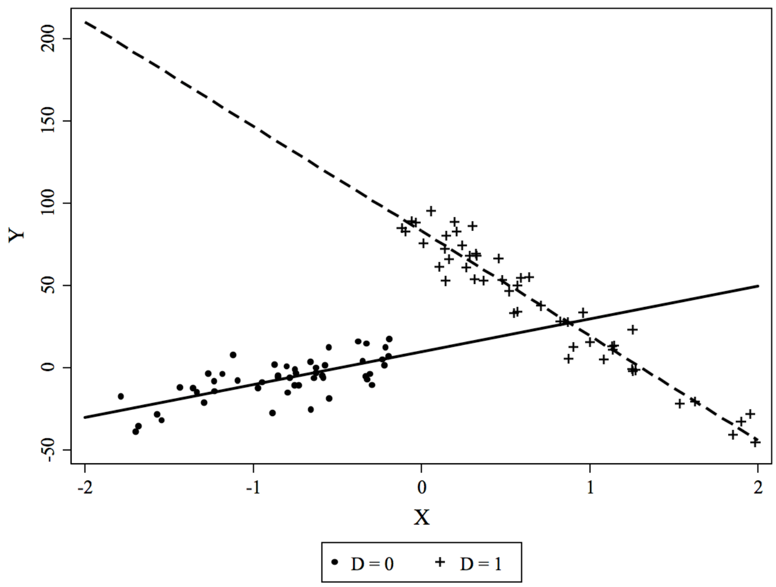


Figure 6: Scatterplot and Regression Lines of Y on X for $D = 0$ and $D = 1$: Slope and Intercept Change ($\beta_2 \neq 0$, $\beta_3 \neq 0$)



In the context of a model with one dichotomous and one continuous interacted covariate, we can test for the sorts of differences illustrated in Figures 3-6 by examining and testing for the sizes and statistical significance of the coefficients in (12). Note also that there is another approach to estimating such models. Rather than the model in (12), we can instead estimate:

$$Y_i = \gamma_0 + \gamma_1 D_i + \gamma_2 X_i D_i + \gamma_3 X_i (1 - D_i) + u_i \tag{13}$$

In other words, we can create two "new" variables, one ($X_i D_i$ that equals X_i only when $D_i = 1$ (and zero otherwise), and the other that equals X_i only when $D_i = 0$ (and zero otherwise). This is both mathematically and substantively identical to the model in (12), and is the approach suggested by Wright (1976).

Two Continuous Covariates

It is also common to have a model in which both interacted covariates are continuous:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + u_i. \tag{14}$$

The interpretation here is that each covariate's effect on Y changes *smoothly* and *monotonically* as a function of the other covariate. Note first that, for such models,

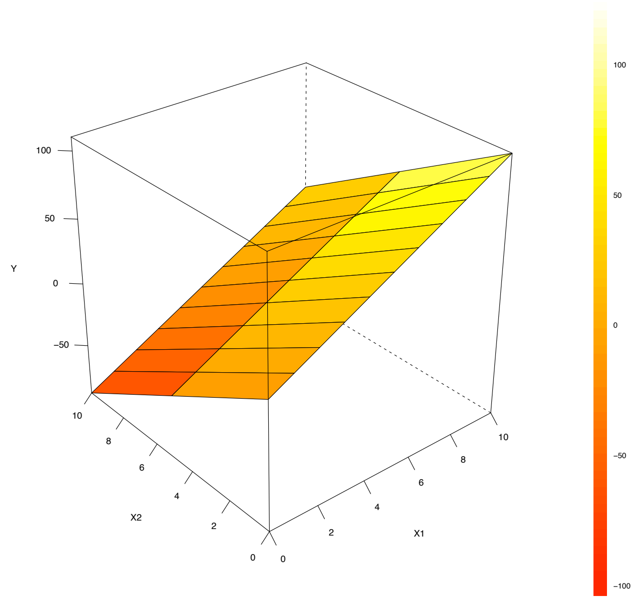
- Absent any interaction effects (that is, if $\beta_3 = 0$), the impact of X_1 (X_2) on Y remains constant across all levels of X_2 (X_1).

- On the other hand, if $\beta_3 \neq 0$, then the "slope" of X_1 (X_2) on Y is different at different values of X_2 (X_1).

This is akin to the unions/government ideology example I gave at the beginning of class. Such relationships are is best represented graphically using 3-D figures of one sort or another; while I generally recommend contour plots, in honor of my friend Luke Keele I'll use a wireframe[3] here:

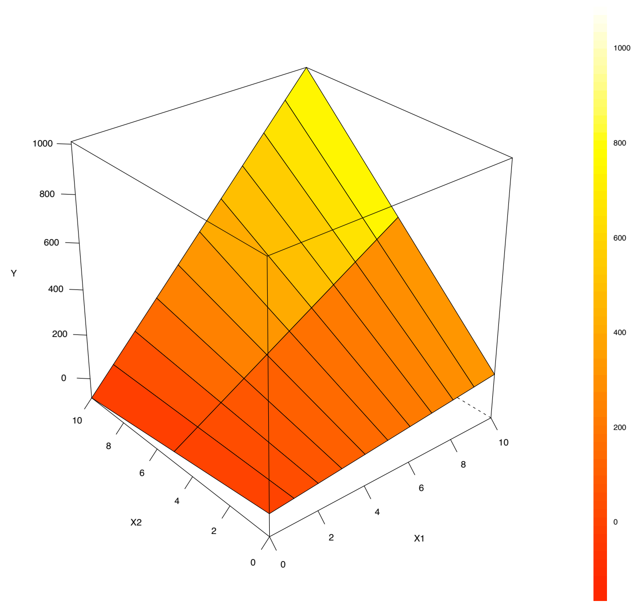
[3] Stata won't do this sort of graphs; I am pretty sure I used the `scatterplot3d` package in R to create these.

Figure 7: Two-Variable Model: No Interactive Effects



Note: Plane denotes $Y_i = 10 + 10X_{1i} - 10X_{2i}$.

Figure 8: Two-Variable Model: Interactive Effects



Note: Plane denotes $Y_i = 10 + 10X_{1i} - 10X_{2i} + 10X_{1i}X_{2i}$.

Note that in Figure 7, the marginal influence of a change in each X on Y is not dependent on the value of the other covariate. By contrast, Figure 8 shows that when values of (e.g.) X_1 are small, the effect of X_2 on Y is both negative and relatively small; however, when X_1 is large, the effect of X_2 on Y is both large and positive.

Note as well that while such models are generally quite flexible (and so are widely used), they also have a few restrictions that should be kept in mind:

- The model assumes that the effects of each of the covariates changes smoothly with the values of the other. Thus, if there is reason to believe that this is not the case (perhaps there are "threshold" effects), then an alternative coding for the variable in question is in order.

- The model also has the effect of each covariate changing monotonically across values of the other. Again, think about whether or not this makes sense; if, for example, your theory suggests that (say) political extremists behave one way, and moderates another, then a simple interaction of the sort described here will not accurately represent what is going on. (For more on that, see the discussion of polynomial models, below).

Quadratic, Cubic, and Other Polynomial Effects

With a continuous covariate, we sometimes want to consider the possibility of a "curvilinear" effect. For example, the effect of age on the probability of a traffic accident is curvilinear: beginning at age 16, it declines steadily until about age 40, then begins to increase again around age 60 and after. A commonly-used way of estimating curvilinear effects is a polynomial regression model, of the form:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + ... + \beta_j X_i^j + u_i. \tag{15}$$

We refer to this general model as a jth-order polynomial regression. Specific cases include the quadratic model ($j = 2$):

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i. \tag{16}$$

and the cubic model ($j = 3$):

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i. \tag{17}$$

Models of this sort are often used to represent curvilinear relationships. If we consider the marginal effect of X on Y in (say) the quadratic model, we find that it is:

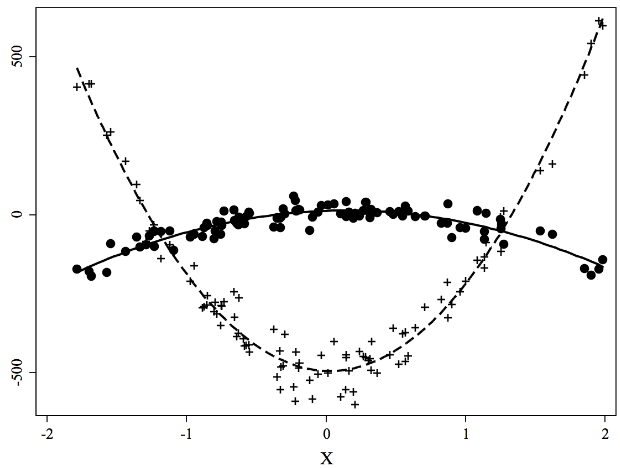
$$\frac{\partial \mathrm{E}(Y)}{\partial X} = \beta_1 + 2\beta_2 X$$

This tells us that the marginal effect of X on Y depends linearly on the value of X itself. More generally, the marginal effect of X in the more general polynomial model (15) is:

$$\frac{\partial \mathrm{E}(Y)}{\partial X} = \beta_1 + 2\beta_2 X + 3\beta_3 X^2 + ... + j\beta_j X^{j-1}$$

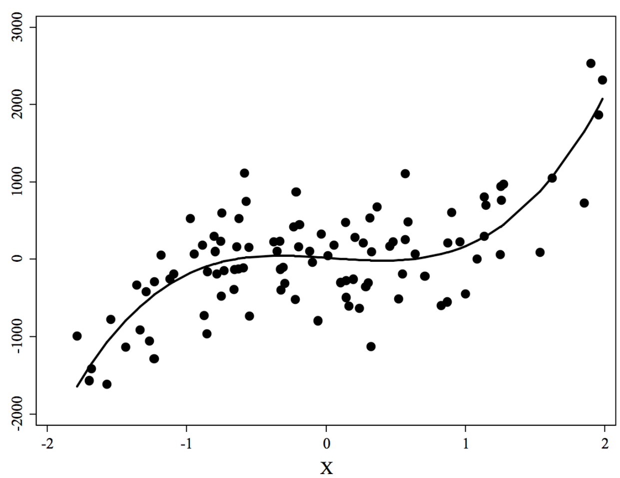
In other words, polynomial models capture the possibility that the effect of X depends on higher-order X terms. We illustrate these sorts of models in Figures 9 and 10:

Figure 9: Examples of Two Quadratic Relationships



Note: Solid line is $Y_i = 10 + 10X_i - 50X_i^2 + u_i$; dashed line is $Y_i = -500 - 20X_i + 300X_i^2 + u_i$.

Figure 10: Example of a Cubic Relationship



Note: Solid line is $Y_i = 10 + 10X_i - 50X_i^2 + 300X_i^3 + u_i$.

Intuitively, there are a few things to remember:

- The first key point to remember is that one needs a polynomial model of one order higher than you have "bends" in the relationship. So, if theory tells you that the effect of some X on Y first increases, then decreases, then increases again (two "bends,") then a cubic model is called for.
- Second, interpretation:
    - In the quadratic model, the sign of β_1 indicates the direction of the slope of Y on X at lower levels of X, while the sign of β_2 tells the slope of X on Y at higher values of X. So,
    - $\beta_1 < 0$ and $\beta_2 > 0$ yields a "U-shaped" function, while
    - $\beta_1 > 0$ and $\beta_2 < 0$ yields an "inverted-U" shaped function.
- Finally, as with any other function, we can find the "inflection points" of a polynomial model by setting the first derivatives of the statistical estimates equal to zero and solving for X; second derivatives will tells us which are maxima and which are minima (of course, you could also just graph the predicted values...).

Higher-Order Interactive Models

Once we've considered the possibility that one variable conditions the impact of another, why not expand that thought? It is perfectly possible – and sometimes useful – to adopt higher-order interaction models, of the form:

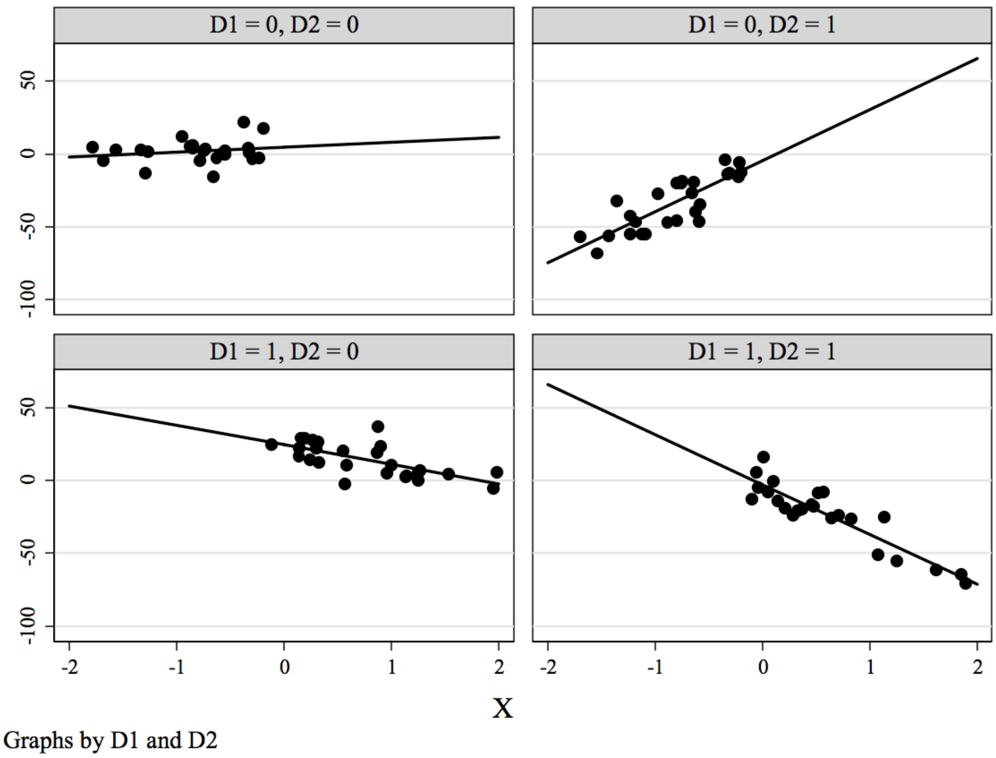
$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{1i} X_{2i} + \beta_4 X_{1i} X_{3i} u_i + \beta_5 X_{2i} X_{3i} + \beta_6 X_{1i} X_{2i} X_{3i} + u_i \quad (18)$$

The various Xs may be continuous or dichotomous, with corresponding interpretations. In this model, each of the three Xs conditions the influence of the other. This is, obviously, a complex animal; it's easiest to grasp if you consider two dummy variables and one continuous X:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_{1i} + \beta_3 D_{2i} + \beta_4 X_i D_{1i} + \beta_4 X_i D_{2i} u_i + \beta_5 D_{1i} D_{2i} + \beta_6 X_i D_{1i} D_{2i} + u_i \quad (19)$$

Think of this as a model in which the "slope" of Y on X is allowed to be different for each of the four groups defined by the values of D_1 and D_2; we can illustrate this model graphically as:

Figure 11: Three-Way Interaction: Two Dummy and One Continuous Covariate



Note: Relationship is
$Y_i = 10 + 10X_i + 10D_{1i} - 20D_{2i} - 20X_iD_{1i} + 20X_iD_{2i} - 50X_iD_{1i}D_{2i} + u_i$.

Here, the slope of Y on X is positive when $D_1 = 0$, and negative when $D_1 = 1$. In addition, D_2 has an exacerbating effect: its presence makes the effect of X on Y stronger, irrespective of whether that effect is positive or negative. Of course, this is just one example; there are lots of other possibilities for such models.

Summary Thoughts

As you can see, there are lots of potential complications one can address with interaction terms. We haven't even talked about combinations of the above (e.g., interactions of covariates with quadratic effects, etc.). Here are a few more general tips:

- **It is almost never a good idea to omit the "direct effects" components from an interactive regression model.** The short answer to the question "Why not?" is that it will lead to bias; the longer answer is spelled out and discussed at length in Brambor et al. (2006).

- Because the direct effects are the impact of one variable on the dependent when the other is zero, *it is important that zero be a meaningful value on the variable in question.*
  - For example, often one might interact a variable with time, e.g. in studies of institutionalization, the influence of heterogeneity might vary with year.
  - The direct effect on the heterogeneity term will reflect the `year` variable being zero; for most studies, 0 A.D. is pretty out-of-sample.
- Relatedly, one can almost always rescale the variables in an interactive model so that the "direct" effects are statistically differentiable from zero; that doesn't mean that the variables' effects are meaningful. (Brambor et al. (2006) have a nice discussion of this as well).
- Quadratics, cubics, etc. are far from the only way to model "curvilinear" effects; and, in fact, are some of the most restrictive functional forms for such relationships. We'll discuss some other, flexible alternatives for such relationships a bit later in the course.
- In general, interaction terms higher than second-order are very, very difficult to understand substantively. Often times, it is a better idea to split one's sample by a dummy variable, since that way it is easier for most people to grasp what is going on.

An Example

Data Preliminaries

We'll consider data from the 1996 American National election study. The main variable of interest (response) is the *Clinton Feeling Thermometer*, which ranges from 0 to 100. We'll consider three other covariates:

- `RConserv` – the respondent's self-placement on a liberal/conservative scale, ranging in value from 1 (extremely liberal) to 7 (extremely conservative),
- `ClintonConserv` – the respondent's placement of President Clinton on the same 1 (extremely liberal) to 7 (extremely conservative) scale, and
- `GOP` – a dummy variable coded 1 if the respondent self-identified as a Republican, and 0 otherwise.

The data (with cases containing missing data omitted) look like this:[4]

[4]Note: At this writing, the code throughout these notes works, and gives the same results, but doesn't exactly match the code on github or the slides.

```
> ClintonTherm<-read.dta("ClintonTherm.dta",convert.factors=FALSE)
> attach(ClintonTherm)
> summary(ClintonTherm)
     caseid       ClintonTherm     RConserv      ClintonConserv
 Min.   :1001   Min.   :  0   Min.   :1.000   Min.   :1.000
 1st Qu.:1440   1st Qu.: 30   1st Qu.:3.000   1st Qu.:2.000
 Median :1854   Median : 60   Median :4.000   Median :3.000
 Mean   :2001   Mean   : 57   Mean   :4.323   Mean   :2.985
 3rd Qu.:2262   3rd Qu.: 85   3rd Qu.:5.000   3rd Qu.:4.000
 Max.   :3403   Max.   :100   Max.   :7.000   Max.   :7.000
      PID             GOP
 Min.   :1.000   Min.   :0.0000
 1st Qu.:1.000   1st Qu.:0.0000
 Median :2.000   Median :0.0000
 Mean   :2.059   Mean   :0.3161
 3rd Qu.:3.000   3rd Qu.:1.0000
 Max.   :5.000   Max.   :1.0000
```

A basic linear model of the Clinton feeling thermometer might look like this:

```
> summary(lm(ClintonTherm~RConserv+GOP))

Call:
lm(formula = ClintonTherm ~ RConserv + GOP)

Residuals:
    Min      1Q  Median      3Q     Max
-86.989 -12.886   2.471  17.471  78.600

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  93.4756     2.2278   41.96   <2e-16 ***
RConserv     -6.4866     0.5373  -12.07   <2e-16 ***
GOP         -26.6699     1.6056  -16.61   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 23.65 on 1294 degrees of freedom
Multiple R-squared: 0.3795,Adjusted R-squared: 0.3786
F-statistic: 395.7 on 2 and 1294 DF,  p-value: < 2.2e-16
```

This suggests that (surprise!) conservatives and Republicans like Clinton less than Democrats and/or Liberals...

Estimating Models with Multiplicative Interactions

Of course, the model above is a very oversimplified one. We'll actually consider two different types of interactive models today. In the first, consider the fact that, while Clinton was a polarizing figure in partisan terms, he was actually something of a moderate as late-20th century American Democrats go. In fact, Clinton was simultaneously able to attract moderate Republicans (especially those old "Reagan Democrats") and to infuriate the more conservative wing of that party.

All this might suggest that, while most people will like or hate Clinton on the basis of their own ideology, Republicans will be especially responsive in this regard: even more so than for others (especially Democrats), we might expect conservative Republicans to *really* dislike Clinton, and moderate or liberal Republicans to support him. This suggests a mode of the form:

$$\begin{aligned}\text{Clinton Thermometer}_i &= \beta_0 + \beta_1(\text{Respondent Conservatism}_i) + \beta_2(\text{GOP Respondent}_i) + \\ &= \beta_3(\text{Respondent Conservatism} \times \text{GOP Respondent}) + u_i \end{aligned} \tag{20}$$

Here, the expectation is that $\beta_1 < 0$ (conservatives will be more negative towards Clinton), $\beta_2 < 0$ (Republicans will be more negative towards Clinton), and $\beta_3 < 0$ (the responsiveness of Clinton's thermometer score to the respondent's ideology will be even greater – that is, more negative – among Republicans).

The model looks like this:

```
> fit1<-lm(ClintonTherm~RConserv+GOP+RConserv*GOP)
> summary(fit1)

Call:
lm(formula = ClintonTherm ~ RConserv + GOP + RConserv * GOP)

Residuals:
    Min      1Q  Median      3Q     Max
-84.357 -13.216   2.355  15.071  83.957

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   89.9271     2.4866  36.165  < 2e-16 ***
RConserv      -5.5705     0.6085  -9.154  < 2e-16 ***
GOP           -6.4840     6.5690  -0.987  0.32379
RConserv:GOP  -4.0581     1.2808  -3.168  0.00157 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
Residual standard error: 23.57 on 1293 degrees of freedom
Multiple R-squared: 0.3843,Adjusted R-squared: 0.3829
F-statistic:   269 on 3 and 1293 DF,  p-value: < 2.2e-16
```

Note a few things about these estimates:

- The results are as expected, though the estimate for the "direct effect" of `GOP` does not appear to be "statistically significant" (but see below).
- More important, these results suggest that, for non-Republicans (that is, when `GOP`=0), we have:

$$\begin{aligned}\text{E(Clinton Thermometer)}_i &= 89.9 - 6.5(0) - 5.6(\text{Respondent Conservatism}_i) - \\ &\quad 4.0(0 \times \text{Respondent Conservatism}_i) \\ &= 89.9 - 5.6(\text{Respondent Conservatism}_i)\end{aligned}$$

while for Republicans (that is, when `GOP`=1), the equation is:

$$\begin{aligned}\text{E(Clinton Thermometer)}_i &= (89.9 - 6.5(1)) + (-5.6 - 4.0(1 \times \text{Respondent Conservatism}_i)) \\ &= 83.4 - 9.6(\text{Respondent Conservatism}_i)\end{aligned}$$

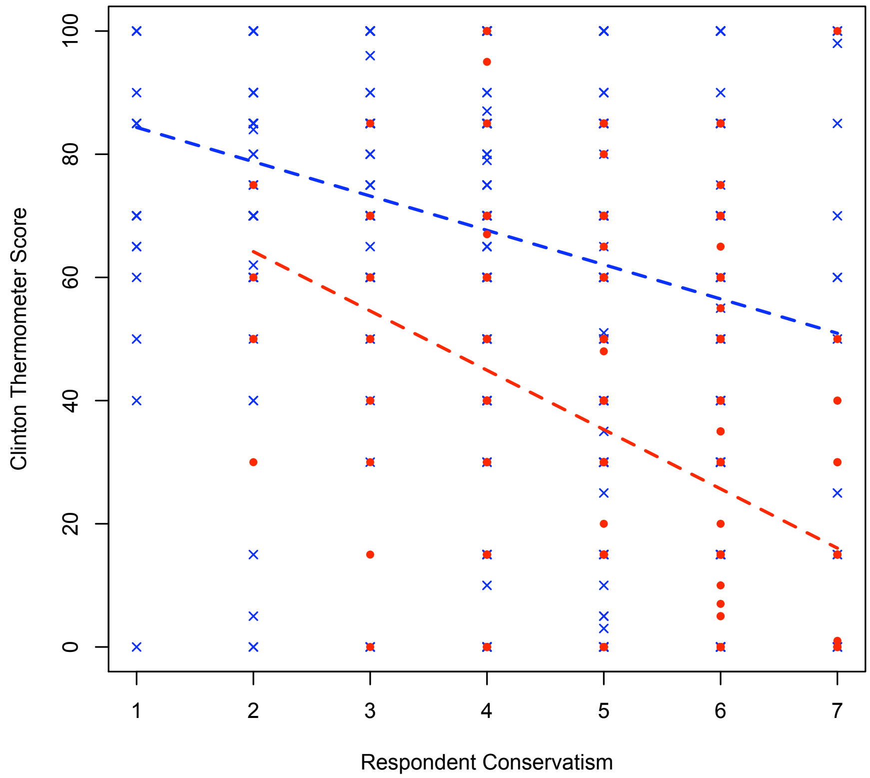
In the language that we discussed last time, we can think of the value -9.6 as $\hat{\psi}_1$ – that is, the "quasi-coefficient" estimate for the effect of `RConserv` on `ClintonTherm`. In other words, the intercept for Republicans is somewhat lower than for non-Republicans, but the (negative) slope on conservatism for Republicans is almost twice as large as for non-Republicans. This can be seen in the plot in Figure 12, below. The red symbols and line are for respondents identifying with the GOP; the blue symbols and line are for non-Republicans.

Note two additional things about these results:

1) They are (almost) identical to the results we get if we estimate separate regression models for Republicans and non-Republicans. This is because the model with the interaction terms allows both different slopes (on `RConserv`) and intercepts for Republicans and non-Republicans:

```
> NonReps<-subset(ClintonTherm,GOP==0)
> summary(lm(ClintonTherm~RConserv,data=NonReps))
```

Figure 12: Scatterplot of Clinton Thermometer Scores by Respondent Conservatism, GOP and Non-GOP, with OLS Fits




```
Call:
lm(formula = ClintonTherm ~ RConserv, data = NonReps)

Residuals:
    Min      1Q  Median      3Q     Max
-84.357  -8.786   2.355  17.355  49.067

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  89.9271     2.4695  36.416   <2e-16 ***
RConserv     -5.5705     0.6043  -9.217   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 23.41 on 885 degrees of freedom
Multiple R-squared: 0.08759,Adjusted R-squared: 0.08656
```

```
F-statistic: 84.96 on 1 and 885 DF,  p-value: < 2.2e-16

> Reps<-subset(ClintonTherm,GOP==1)
> summary(lm(ClintonTherm~RConserv,data=Reps))

Call:
lm(formula = ClintonTherm ~ RConserv, data = Reps)

Residuals:
    Min      1Q  Median      3Q     Max 
-54.557 -18.014  -4.372  15.071  83.957 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   83.443      6.170  13.524  < 2e-16 ***
RConserv      -9.629      1.144  -8.419 6.52e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 23.92 on 408 degrees of freedom
Multiple R-squared: 0.148,Adjusted R-squared: 0.1459 
F-statistic: 70.88 on 1 and 408 DF,  p-value: 6.518e-16
```

The "almost" qualifier is necessary because while the point estimates are identical, the standard errors are not precisely so. This is because the interactive model estimates the two implicit models "together," and therefore is (very slightly) more efficient than the separate models. In large data sets, however, the difference is tiny.

2) It is also the case that, if GOP membership moderates the association between conservatism and Clinton's thermometer score, then it also muse be true that conservatism moderates the association between those scores and party membership. That is:

1. We can think of the estimate $\hat{\beta}_2 = -6.48$ as the expected difference in Clinton's thermometer score between Republicans and non-Republicans *when the respondent's conservatism equals zero*. Here, that is not such an interesting thing, since there are no values in the data where `RConserv` $= 0$ (that would correspond to someone who was even more liberal than "extremely liberal").

2. To get a quantity that is of more interest, we have to pick values for `RConserv` that are in the data. So, for example, for an "extremely liberal" survey respondent (with `RConserv` $= 1$), the difference in the expected Clinton thermometer score between a Republican and a non-Republican is:

$$
\begin{aligned}
\hat{\psi}_2 &= \hat{\beta}_2 + \hat{\beta}_3(1) \\
&= -6.48 + (-4.06)(1) \\
&= -10.54
\end{aligned}
$$

3. Similarly, the expected interparty difference for an "extremely conservative" respondent (`RConserv` $= 7$) is:

$$
\begin{aligned}
\hat{\psi}_2 &= \hat{\beta}_2 + \hat{\beta}_3(7) \\
&= -6.48 + (-4.06)(7) \\
&= -34.9
\end{aligned}
$$

   Consistent with our theory, the differences (here, between Republicans and non-Republicans) in opinions about Clinton are largest among those who are very conservative, and smallest among those who are most liberal.

Inference, Testing, and Presentation of Results

Thus far, we've focused only on point estimates of the βs (and the ψs). If we rewrite (20) in terms of the marginal impact of the `RConserv` variable, for example, we get

$$
\begin{aligned}
\text{Clinton Thermometer}_i &= \beta_0 + (\beta_1 + \beta_3 \text{GOP}_i)\text{Respondent Conservatism}_i + \\
&\quad \beta_2 \text{GOP}_i + u_i \\
&= \beta_0 + \psi_1 \text{Respondent Conservatism}_i + \beta_2 \text{GOP}_i + u_i \qquad (21)
\end{aligned}
$$

where, as you recall from before, ψ_1 is a simplified way of writing (in this case) $\frac{\partial \text{E}(Y)}{\partial \texttt{RConserv}}$.

Thankfully, it's straightforward to calculate both the point estimates $\hat{\psi}$ as well as its implicit standard error – which, you will recall, is equal to

$$
\hat{\sigma}_{\psi_1} = \sqrt{\widehat{\text{Var}(\hat{\beta}_1)} + (\text{GOP})^2 \widehat{\text{Var}(\hat{\beta}_3)} + 2(\text{GOP})\widehat{\text{Cov}(\hat{\beta}_1, \hat{\beta}_3)}}.
$$

We can recover the point estimate and implicit standard error of $\hat{\psi}_1$ in a straightforward way:

```
> fit1$coeff[2]+fit1$coeff[4]
 RConserv
-9.628577

> sqrt(vcov(fit1)[2,2] + (1)^2*vcov(fit1)[4,4] + 2*1*vcov(fit1)[2,4])
[1] 1.127016
```

Here, `vcov(fit1)[A,B]` denotes the element in the Ath row and Bth column of the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ in the model called `fit1`. The ratio of these two numbers yields a t-score of -8.54, which indicates that the marginal association between the thermometer score for Clinton and respondent conservatism is very precisely estimated ($p < .001$).

We can do the same thing if we're interested in conducting inference on $\hat{\psi}_2$ (the marginal effect of `GOP` on Y), which, for any given value of *Respondent Conservatism* (say, k) is:

$$\hat{\psi}_2|(\text{Respondent Conservatism} = k) = \hat{\beta}_2 + k\,\hat{\beta}_3.$$

We also know that, as was the case above,

$$\hat{\sigma}_{\psi_2} = \sqrt{\widehat{\text{Var}(\hat{\beta}_2)} + k^2\widehat{\text{Var}(\hat{\beta}_3)} + 2k\widehat{\text{Cov}(\hat{\beta}_2, \hat{\beta}_3)}}.$$

This means that for (say) extreme liberals (those with `GOP` = 1), we have a point estimate equal to:

```
> fit1$coeff[3]+(1 * fit1$coeff[4])
      GOP
-10.54208
```

and an associated standard error of:

```
> sqrt(vcov(fit1)[3,3] + (1)^2*vcov(fit1)[4,4] + 2*1*vcov(fit1)[3,4])
[1] 5.335847
```

yielding a t-score of around -1.98. If we do the same thing for extreme conservatives (those with `RConserv` = 7), we get:

```
> fit1$coeff[3]+(7 * fit1$coeff[4])
      GOP
-34.89045

> sqrt(vcov(fit1)[3,3] + (7)^2*vcov(fit1)[4,4] + 2*7*vcov(fit1)[3,4])
[1] 3.048302
```

yielding a t-score of about -11.5.

But all of this seems remarkably clunky. And, in fact, it is; we can actually get the same hypothesis tests on our estimates of the ψs using a straightforward `R` command: `linear.hypothesis()` in the `car` package. This operates similarly to the `lincom` command in `Stata` in that it does F-tests (or, alternatively and mostly equivalently, Wald tests) on linear combinations of parameters in our model. So, to test the hypothesis that $\hat{\psi}_1 = 0$ (that is, that $\hat{\beta}_1 + (1)\hat{\beta}_3 = 0$) in the model above, we use:

```
> library(car)
> linear.hypothesis(fit1,"RConserv+RConserv:GOP")
Linear hypothesis test

Hypothesis:
RConserv + RConserv:GOP = 0

Model 1: ClintonTherm ~ RConserv + GOP + RConserv * GOP
Model 2: restricted model

  Res.Df    RSS Df Sum of Sq     F    Pr(>F)
1   1293 718173
2   1294 758714 -1    -40541 72.99 < 2.2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Note that this yields a p-value for the hypothesis test that is exactly equal to that for the t-test given above; this is because (as we discussed previously in our class on F-tests) an F-test for a single marginal effect is nothing more than the square of the associated t-statistic:

```
> sqrt(72.99)
[1] 8.543419
```

We can do the same thing for $\hat{\psi}_2$, say, if we want to test the hypothesis that the effect of GOP on Clinton's thermometer rating for extreme conservatives (`RConserv` $= 7$) is zero:

```
> linear.hypothesis(fit1,"GOP+7*RConserv:GOP")
Linear hypothesis test

Hypothesis:
GOP + 7 RConserv:GOP = 0

Model 1: ClintonTherm ~ RConserv + GOP + RConserv * GOP
Model 2: restricted model

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1   1293 718173
2   1294 790938 -1    -72766 131.01 < 2.2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Once again, this is exactly equal to $(-11.5)^2$, and so yields the same inferences.

Plots (3-D and Otherwise) for Multiplicative Interactions

Marginal Effects Plots...

(See the slides.)

Plots of Predicted Values...

(See the slides.)

Wrap-Up

Finally, a couple key points to remember about presenting results in models with multiplicative interaction terms:

- Whenever you are discussing these models, it is imperative to keep in mind what the actual values of the different covariates in your data are. Beyond the fact that our inferences get shakier outside of the range of the data (a phenomenon reflected in the correspondingly broader confidence intervals around such predictions), it's important to keep in mind that our counterfactuals shouldn't be *too* counterfactual – that is, so counterfactual as to strain the limits of the theoretical model that underlies them in the first place. This is, of course, always good advice, but can be particularly important when multiplicative interactions mean that variables' effects on Y are no longer orthogonal to one another. Gary King and Langche Zeng have a nice paper on this in the Spring 2006 issue of *Political Analysis*.

- As we've said from the very beginning, whenever possible one should include measures of uncertainty as well as point predictions, irrespective of whether the quantity we're talking about is $\hat{\psi}$ or $\hat{Y}$. In the case of line/scatterplots, that's relatively easy to do; for contour and other three-dimensional ones, in contrast, it is almost impossible. The latter have "confidence surfaces" which are guaranteed to uglify just about any otherwise nice-looking plot into which they are introduced. That in turn means that, if there is something particularly important about the confidence intervals (e.g., if they change very significantly over different values of one of the Xs), a better approach is to skip the contour plots and go with something more like Figure **??**. Alternatively, one can choose 2-3 values of the interacted/mediating variable, and then plot $\hat{Y}$ vs. X for those different values, along with the confidence intervals for each.