PLSC 503 – Spring 2025 Variable Selection, Specification Bias, and Multiplicative Interactions

March 3, 2025

Among our OLS / linear regression assumptions:

- $\text{Cov}(\mathbf{X}, \mathbf{u}) = 0$, and
- the distribution of $\mathbf{X}$ does not depend on either $\boldsymbol{\beta}$ or σ^2.

Model Specification

"Truth":

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

Fitted model:

$$Y_i = \gamma_0 + \gamma_1 X_{1i} + e_i$$

Then:

$$e_i = \beta_2 X_{2i} + u_i$$

Omitted Variable Bias

Note:

$$
\begin{aligned}
\mathsf{E}(e) &= \mathsf{E}(\beta_2 X_2 + u) \\
&= X_2 \mathsf{E}(\beta_2) + \mathsf{E}(u) \\
&\neq 0
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{E}(\gamma_1) &= \beta_1 + \frac{\sum_{i=1}^{N}(X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)}{\sum_{i=1}^{N}(X_{1i} - \bar{X}_1)^2}\beta_2 \\
&= \beta_1 + b_{X_2 X_1}\beta_2
\end{aligned}
$$

where $b_{X_2X_1}$ is the "slope" coefficient one obtains from regressing X_2 on X_1.

Omitted Variable Bias, continued

If X_2 is omitted and $\text{Cov}(X_1, X_2) = 0$, then

- $E(\hat{\gamma}_1) = \beta_1$, but
- $E(\hat{\gamma}_0) \neq \beta_0$.

If $\text{Cov}(X_1, X_2) \neq 0$ then

- $E(\hat{\gamma}_1) \neq \beta_1$ and $E(\hat{\gamma}_0) \neq \beta_0$
- In the simple bivariate case,
  - if $Cov(X_1, X_2) > 0$ then $E(|\hat{\gamma}_1|) > |\beta_1|$,
  - if $Cov(X_1, X_2) < 0$ then $E(|\hat{\gamma}_1|) < |\beta_1|$.

Recall that for one X:

$$\widehat{\text{Var}(\hat{\beta}_1)} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}.$$

and for two Xs:

$$\widehat{\text{Var}(\hat{\beta}_1)} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2(1 - R^2_{X_1 X_2})}$$

Also, because $\hat{e}_i \neq \hat{u}_i$,

$$\text{E}(\sigma_e^2) = \sigma_u^2 + f\left(\beta_2, X_1\right) \;\leftarrow\; \text{Bias}$$

For the "true" DGP

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

and fitted model

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\Gamma} + \mathbf{e}$$

where $\mathbf{Z} \subset \mathbf{X}$, we have

$$\begin{aligned}\boldsymbol{\Gamma} &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\ &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{u}\end{aligned}$$

and so

$$\begin{aligned}\mathsf{E}(\boldsymbol{\Gamma}) &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{P}\boldsymbol{\beta}.\end{aligned}$$

Now assume a "true" model:

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

and fitted model:

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\Gamma} + \mathbf{e}$$

where $\mathbf{X} \subset \mathbf{Z}$. This means:

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

where $\mathbf{Z} = \mathbf{X} \cup \mathbf{W}$ and $\boldsymbol{\Theta} = 0$.

Results:

- $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ and $E(\hat{\sigma}^2) = \sigma^2$, but
- $\widehat{\text{Var}(\boldsymbol{\beta})} > \text{Var}(\boldsymbol{\beta}) \leftarrow$ Inefficiency

Implication: *Pre-Test Bias*

Omitted Variable Bias: Simulated Example

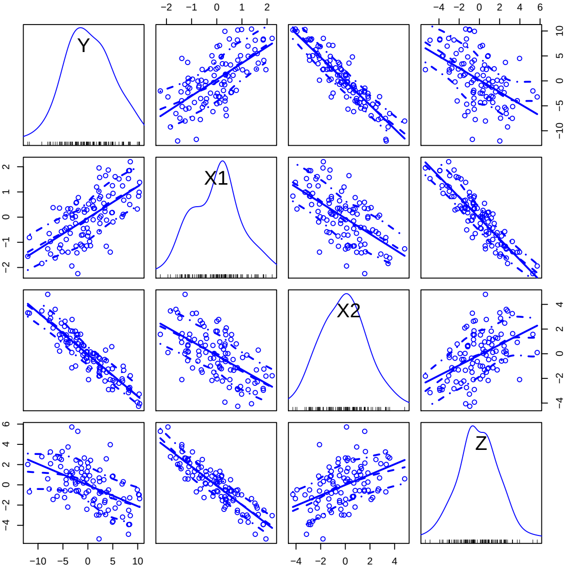
"True" model:

$$Y_i = 0 + 1.0X_{1i} - 2.0X_{2i} + u_i$$

Simulation:

```
> N <- 100
> X1<-rnorm(N)                   # <- X1
> X2<-(-X1)+1.5*(rnorm(N))       # <- correlated w/X1
> Y<-X1-(2*X2)+(2*(rnorm(N))) # <- Y
> Z <- (-2*X1) + rnorm(N)        # <- correlated w/X1 but irrelevant
> data <- data.frame(Y=Y,X1=X1,X2=X2,Z=Z)
```

Scatterplot Matrix



Correctly Specified Model

```
> correct<-lm(Y~X1+X2)
> summary(correct)

Residuals:
   Min     1Q Median     3Q    Max
-5.721 -1.209  0.093  1.198  5.915

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.03311    0.21249  -0.156  0.87651
X1           0.81690    0.26718   3.057  0.00288 **
X2          -2.13652    0.13844 -15.433  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.116 on 97 degrees of freedom
Multiple R-squared:  0.8295, Adjusted R-squared:  0.826
F-statistic:   236 on 2 and 97 DF,  p-value: < 2.2e-16
```

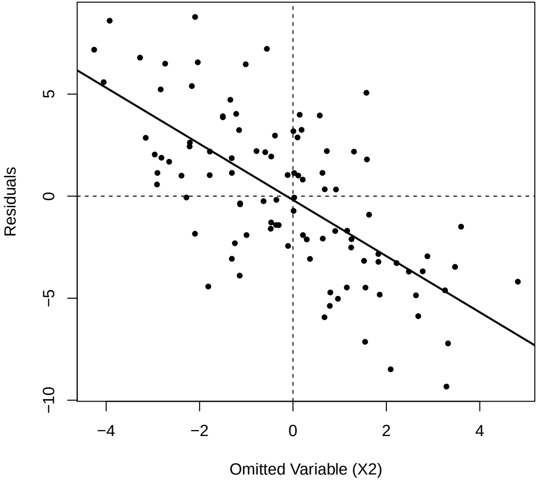
Overspecified Model

```
> overspec<-lm(Y~X1+X2+Z)
> summary(overspec)

Residuals:
    Min      1Q  Median      3Q     Max
-5.9809 -1.0442 -0.0265  1.2609  6.0201

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.01570    0.21420  -0.073  0.94173
X1           0.82148    0.26785   3.067  0.00281 **
X2          -2.11735    0.14105 -15.011  < 2e-16 ***
Z            0.01662    0.02202   0.755  0.45220
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.12 on 96 degrees of freedom
Multiple R-squared:  0.8306, Adjusted R-squared:  0.8253
F-statistic: 156.8 on 3 and 96 DF,  p-value: < 2.2e-16
```

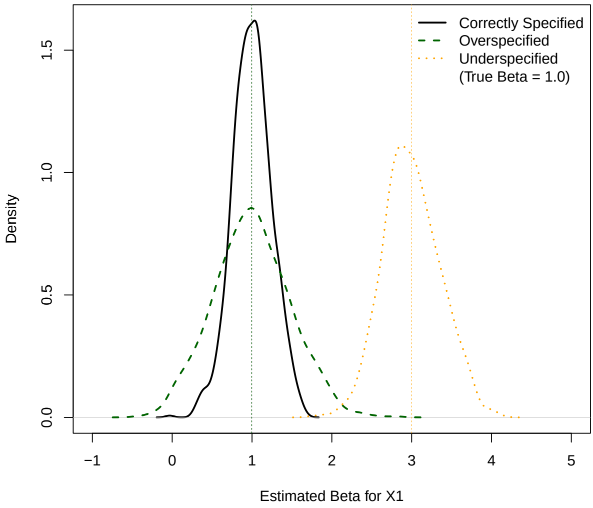
Underspecified Model

```
> incorrect<-lm(Y~X1)
> summary(incorrect)

Residuals:
    Min      1Q  Median      3Q     Max
-9.3297 -2.9762 -0.0672  2.4828  8.7787

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.2704     0.3913   0.691    0.491
X1            3.2783     0.3964   8.270 6.71e-13 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 3.913 on 98 degrees of freedom
Multiple R-squared:  0.411, Adjusted R-squared:  0.405
F-statistic: 68.39 on 1 and 98 DF,  p-value: 6.714e-13
```

Omitted Variable Plot



Repeated 1000 Times...



Nothing Beats a Good Theory. Period.

Also:

- "Model specification tests" $\leftarrow$ meh
- Examine residuals
- Proxy variables...
- *Resist the urge to overspecify!*

Multiplicative Interactions

Our old friend:

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

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

means:

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

and

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

Consider:

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

$$\begin{aligned} \mathrm{E}(Y_i) &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} \\ &= \beta_0 + \beta_2 X_{2i} + (\beta_1 + \beta_3 X_{2i}) X_{1i} \\ &= \beta_0 + \beta_2 X_{2i} + \psi_1 X_{1i} \end{aligned}$$

where $\psi_1 = \beta_1 + \beta_3 X_{2i}$. This means:

$$\frac{\partial \mathrm{E}(Y_i)}{\partial X_1} = \beta_1 + \beta_3 X_{2i}.$$

Similarly:

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

which implies:

$$
\frac{\partial \mathrm{E}(Y_i)}{\partial X_2} = \beta_2 + \beta_3 X_{1i}.
$$

"Direct Effects"

If $X_2 = 0$, then:

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

Similarly, for $X_1 = 0$:

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

Key Point

In most instances, the quantities we care about are not β_1 and β_2, but rather ψ_1 and ψ_2.

Inference

Point estimates:

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

and

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

For variance, recall that:

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

Means that:

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

and

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

Types of Interactions: Dichotomous Xs

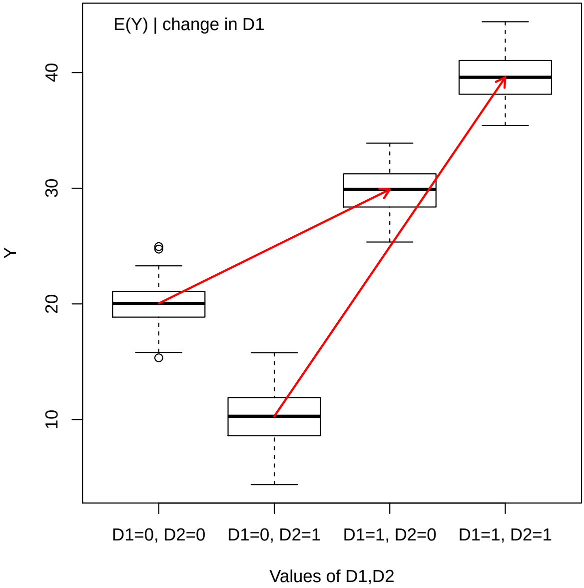
For

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

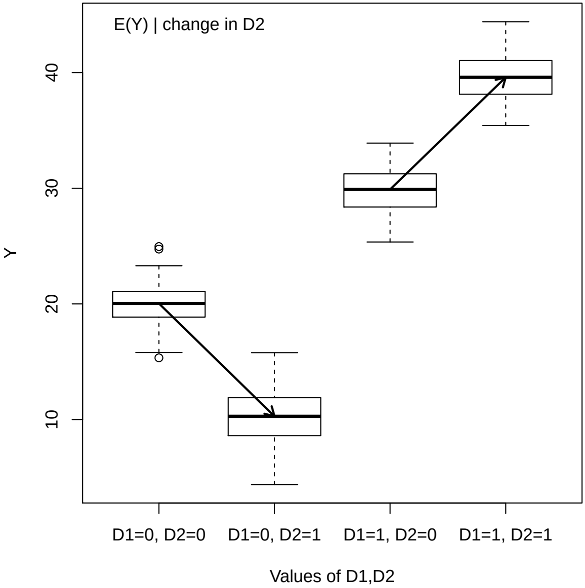
we have:

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

Values of $E(Y)$ for Changes in D_1



Values of $E(Y)$ for Changes in D_2



Writing:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i D_i + u_i$$

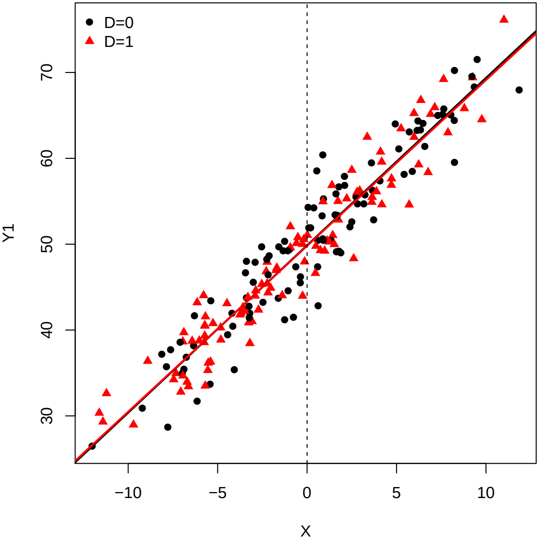
gives:

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

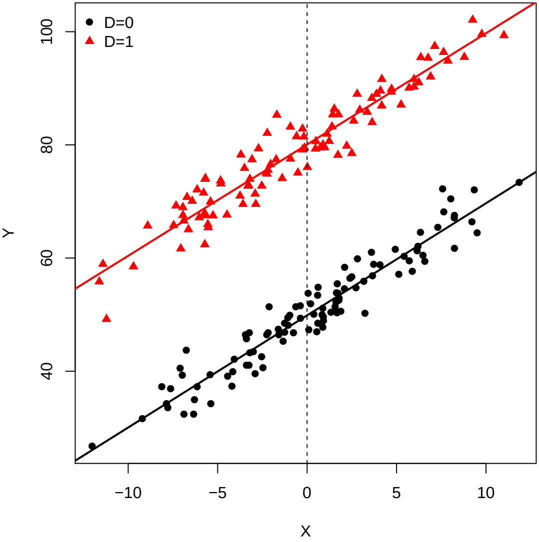
Four possibilities:

- $\beta_2 = \beta_3 = 0$
- $\beta_2 \neq 0$ and $\beta_3 = 0$
- $\beta_2 = 0$ and $\beta_3 \neq 0$
- $\beta_2 \neq 0$ and $\beta_3 \neq 0$

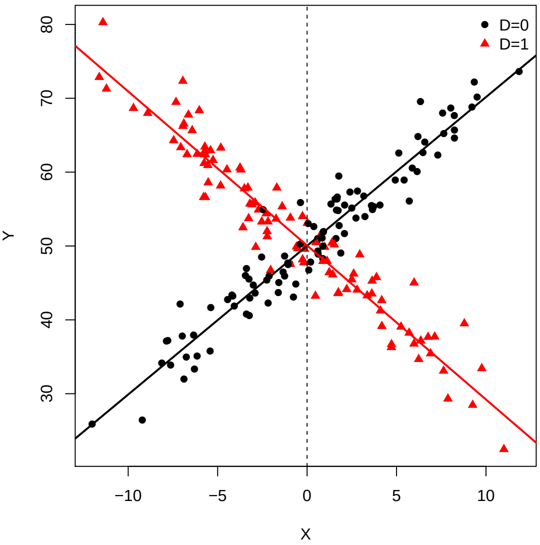
Scatterplot and Regression Lines of Y on X for $D = 0$ and $D = 1$: No Slope or Intercept Differences ($\beta_2 = \beta_3 = 0$)



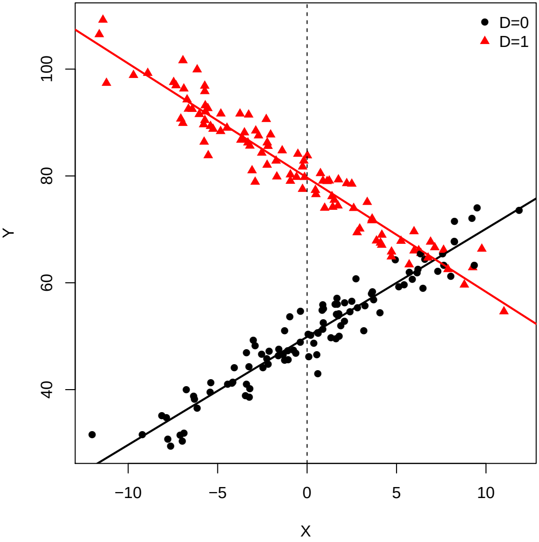
Scatterplot and Regression Lines of Y on X for $D = 0$ and $D = 1$: Intercept Shift ($\beta_2 \neq 0$, $\beta_3 = 0$)



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



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



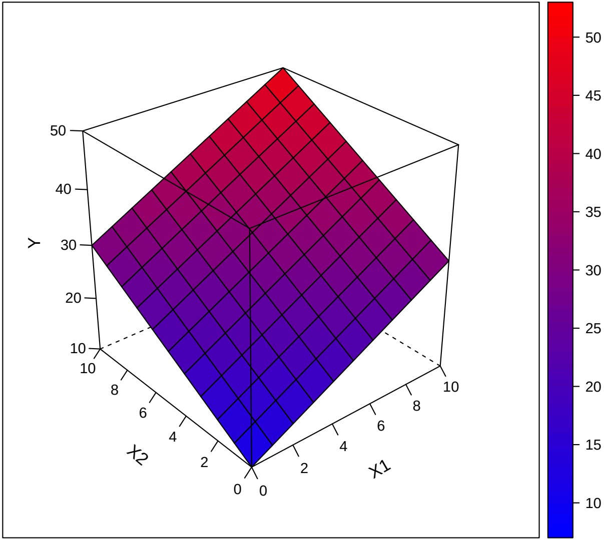
Two Continuous Xs

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

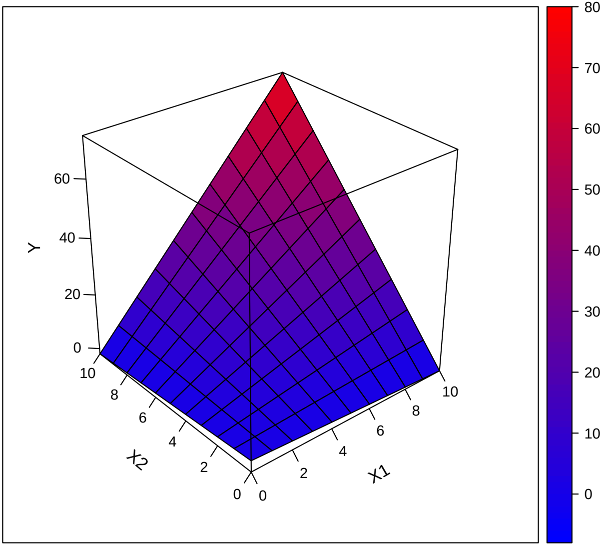
Implies

$$\beta_3 = 0 \; \rightarrow \; \frac{\partial E(Y)}{\partial X_1} = \beta_1 \, \forall \, X_2 \text{ and } \frac{\partial E(Y)}{\partial X_2} = \beta_2 \, \forall \, X_1$$

Two Continuous Variables: No Interactive Effects



Two Continuous Variables: Interaction Present



Quadratic, Cubic, and Other Polynomial Effects

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

In general:

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

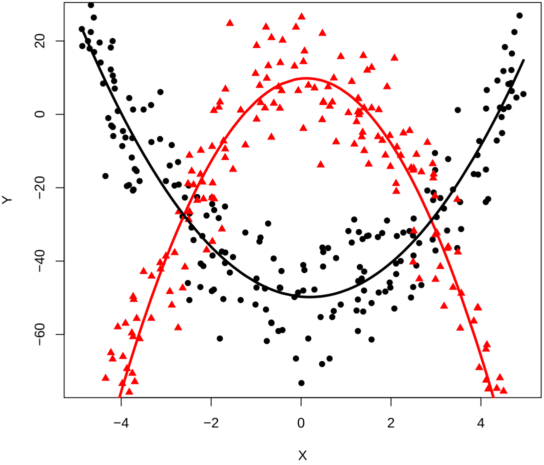
Quadratic case ($j = 2$):

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i.$$

implies

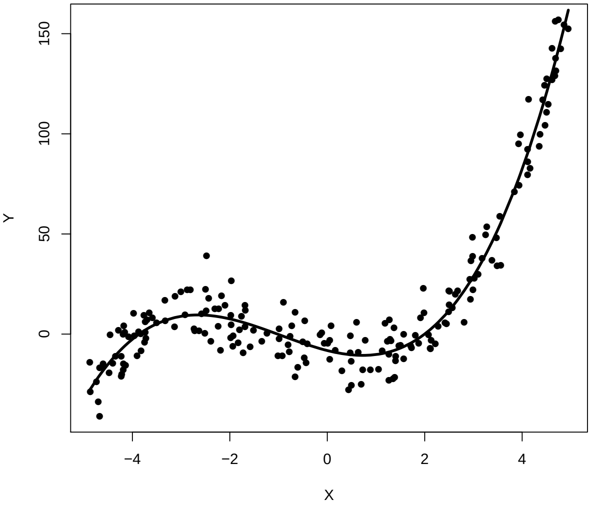
$$\frac{\partial \mathrm{E}(Y)}{\partial X} = \beta_1 + 2\beta_2 X$$

Two Quadratic Relationships



Note: Red line is $Y_i = 10 + 1X_i - 5X_i^2 + u_i$; black line is $Y_i = -50 - 1X_i + 3X_i^2 + u_i$.

Example of a Cubic Relationship



Note: Solid line is $Y_i = -1 + 1X_i - 8X_i^2 + 5X_i^3 + u_i$.


INTERACTIONS
INTERACTIONS EVERYWHERE

Higher-Order Interactive Models

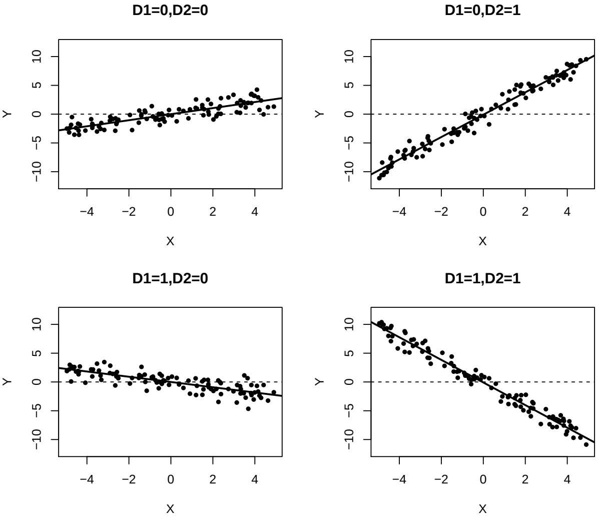
Three-way interaction:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \\ &\quad \beta_4 X_{1i} X_{2i} + \beta_5 X_{1i} X_{3i} + \beta_6 X_{2i} X_{3i} + \beta_7 X_{1i} X_{2i} X_{3i} + u_i \end{aligned}$$

Special case with dichotomous X_1, X_2:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \beta_2 D_{1i} + \beta_3 D_{2i} + \\ &\quad \beta_4 X_i D_{1i} + \beta_5 X_i D_{2i} + \beta_6 D_{1i} D_{2i} + \beta_7 X_i D_{1i} D_{2i} + u_i \end{aligned}$$

Three-Way Interaction: Two Dummy and One Continuous Covariates



Example: President Clinton's 1996 "Thermometer Score"

Details:

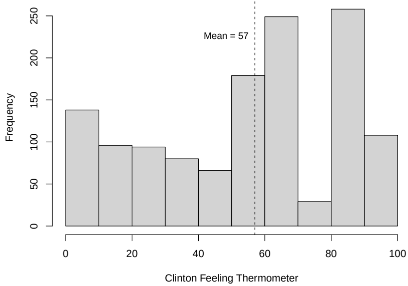
- Source: 1996 American National Election Study
- $N \approx 1300$
- "Feeling Thermometer":

  "In the list that follows, rate that person/party using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person/party. Ratings between 0 and 50 degrees mean that you don't feel favorable toward the person/party and that you don't care too much for that person. You would rate the person at the 50-degree mark if you don't feel particularly warm or cold toward the person."

- Ideology measures:
  - Seven-point scale (1 = "very left / liberal," 7 = "very right / conservative")
  - Each respondent places themself + President Clinton
- Party ID: Republican (GOP=1) vs. Democrat/Independent (GOP=0)

Clinton "Feeling Thermometer" Data

Distribution:



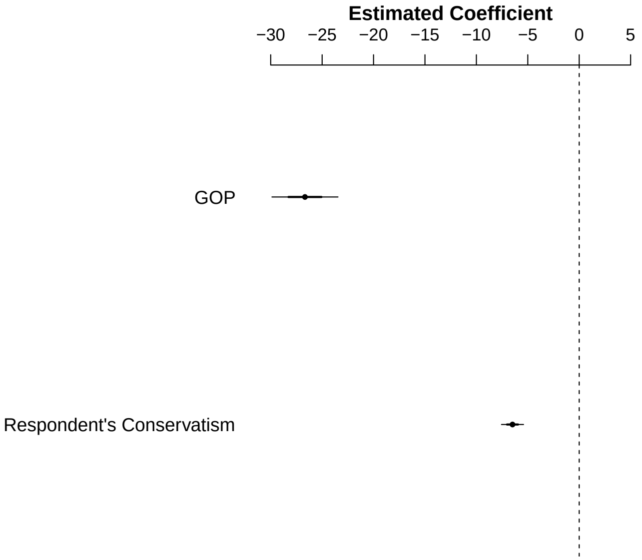
Summary statistics:

```
> describe(ClintonTherm,skew=FALSE)
              vars    n    mean     sd median  min  max range    se
caseid           1 1297 2000.70 718.42   1854 1001 3403  2402 19.95
ClintonTherm     2 1297   57.00  30.00     60    0  100   100  0.83
RConserv         3 1297    4.32   1.39      4    1    7     6  0.04
ClintonConserv   4 1297    2.98   1.37      3    1    7     6  0.04
PID              5 1297    2.06   1.05      2    1    5     4  0.03
GOP              6 1297    0.32   0.47      0    0    1     1  0.01
```

A Basic Regression

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

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

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
```

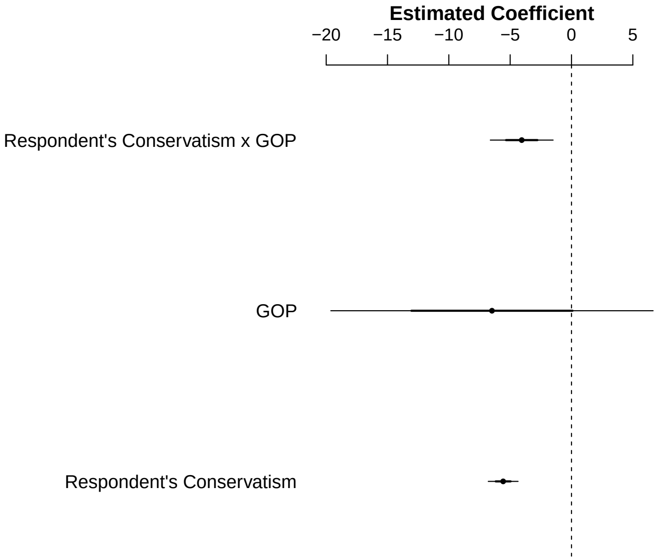
Coefficient Plot: Non-Interactive Model



An Interactive Model

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

> summary(fit1)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    89.927      2.487   36.16   <2e-16 ***
RConserv       -5.571      0.609   -9.15   <2e-16 ***
GOP            -6.484      6.569   -0.99   0.3238
RConserv:GOP   -4.058      1.281   -3.17   0.0016 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23.6 on 1293 degrees of freedom
Multiple R-squared:  0.384, Adjusted R-squared:  0.383
F-statistic:  269 on 3 and 1293 DF,  p-value: <2e-16
```

Coefficient Plot: Interactive Model



Syntax Matters

Different ways to specify the (mathematically identical, but sometimes computationally different) interactive model in R :

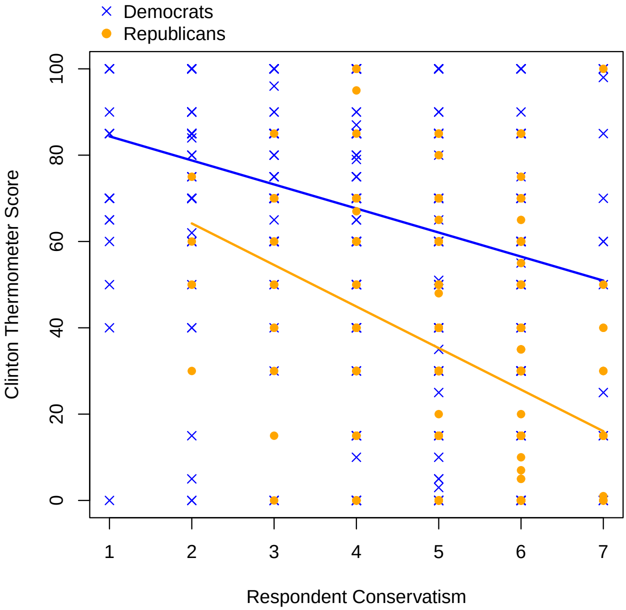
```
> summary(lm(ClintonTherm~RConserv+GOP+RConserv*GOP,data=CT))
> summary(lm(ClintonTherm~RConserv+GOP+RConserv:GOP,data=CT))
> summary(lm(ClintonTherm~RConserv*GOP,data=CT))
> summary(lm(ClintonTherm~(RConserv+GOP)^2,data=CT))
>
> CT$GOPxRC<-CT$GOP * CT$RConserv
> summary(lm(ClintonTherm~RConserv+GOP+GOPxRC,data=CT))
```

Two Regressions, Sort Of

$$\begin{aligned}
\text{E(Thermometer} \mid \text{Non-GOP)}_i &= 89.9 - 6.5(0) - 5.6(\text{R's Conservatism}_i) \\
&\quad -4.0(0 \times \text{R's Conservatism}_i) \\
&= \mathbf{89.9 - 5.6}(\text{R's Conservatism}_\mathbf{i})
\end{aligned}$$

$$\begin{aligned}
\text{E(Thermometer} \mid \text{GOP)}_i &= [89.9 - 6.5(1)] + [-5.6 - 4.0(1 \times \text{R's Conservatism}_i)] \\
&= \mathbf{83.4 - 9.6}(\text{R's Conservatism}_\mathbf{i})
\end{aligned}$$

Thermometer Scores by Conservatism, GOP and Non-GOP



Interactive Results $\approx$ Separate Regressions

```
> NonReps<-subset(CT,GOP==0)
> Reps<-subset(CT,GOP==1)
> split<-list("Interactive" = lm(ClintonTherm~RConserv*GOP,data=CT),
+             "Non-GOP" = lm(ClintonTherm~RConserv,data=NonReps),
+             "GOP" = lm(ClintonTherm~RConserv,data=Reps))
> modelsummary(split,output="SplitModels-25.tex",
+              gof_omit="DF|Deviance|LogLik|AIC|BIC")
```

Feeling Thermometer Models

	Interactive	Non-GOP	GOP
(Intercept)	89.927	89.927	83.443
	(2.487)	(2.469)	(6.170)
Respondent Conservatism	−5.571	−5.571	−9.629
	(0.609)	(0.604)	(1.144)
GOP Respondent	−6.484		
	(6.569)		
Respondent Conservatism × GOP Respondent	−4.058		
	(1.281)		
Num.Obs.	1297	887	410
R2	0.384	0.088	0.148
R2 Adj.	0.383	0.087	0.146
F	269.011	84.961	70.878
RMSE	23.53	23.38	23.86

Discovering $\hat{\psi}_1$ and $\hat{\psi}_2$

For `RConserv`:

$$\begin{aligned} \text{Clinton Thermometer}_i &= \beta_0 + (\beta_1 + \beta_3 \text{GOP}_i)\text{R's Conservatism}_i + \\ &\quad \beta_2 \text{GOP}_i + u_i \\ &= \beta_0 + \psi_{1i}\text{R's Conservatism}_i + \beta_2 \text{GOP}_i + u_i. \end{aligned}$$

So:

$$\hat{\psi}_{1i} = \hat{\beta}_1 + \hat{\beta}_3 \times \text{GOP}_i$$

and

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

Discovering $\hat{\psi}_1$ and $\hat{\psi}_2$

For GOP:

$$\begin{aligned}\text{Clinton Thermometer}_i &= \beta_0 + (\beta_2 + \beta_3 \times \text{R's Conservatism}_i)\text{GOP}_i + \\ &\quad \beta_1(\text{R's Conservatism}_i) + u_i \\ &= \beta_0 + \psi_{2i}\text{GOP}_i + \beta_1(\text{R's Conservatism}_i) + u_i.\end{aligned}$$

So:

$$\hat{\psi}_{2i} = \hat{\beta}_2 + \hat{\beta}_3 \times (\text{R's Conservatism}_i).$$

and

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

Discovering $\hat{\psi}_1$ and $\hat{\psi}_2$

```
> Psi1<-fit1$coeff[2]+fit1$coeff[4]

> Psi1
 RConserv
-9.628577

> SPsi1<-sqrt(vcov(fit1)[2,2] + (1)^2*vcov(fit1)[4,4] + 2*1*vcov(fit1)[2,4])
> SPsi1
[1] 1.127016

> Psi1 / SPsi1 # <-- t-statistic
 RConserv
-8.543422
```

Discovering $\hat{\psi}_1$ and $\hat{\psi}_2$

```
> # psi_2 | RConserv = 1
> fit1$coeff[3]+(1 * fit1$coeff[4])
      GOP
-10.54208

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

# Implies t is approximately 2



> # psi_2 | RConserv = 7
> 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

# t is approximately 11
```

An Easier Way: `linearHypothesis()`

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

Hypothesis:
RConserv  + RConserv:GOP = 0

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

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

> # Note: Same as t-test:
> sqrt(72.99)
[1] 8.543419
```

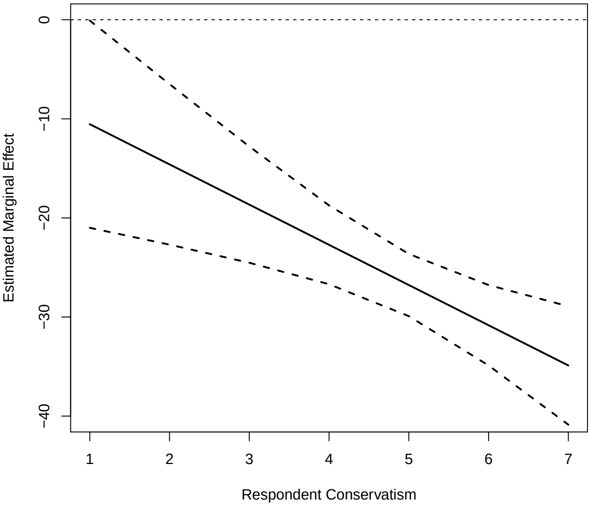
An Easier Way: `linearHypothesis()`

```
> # psi_2 | RConserv = 7:
> linearHypothesis(fit1,"GOP+7*RConserv:GOP")
Linear hypothesis test

Hypothesis:
GOP  + 7 RConserv:GOP = 0

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

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

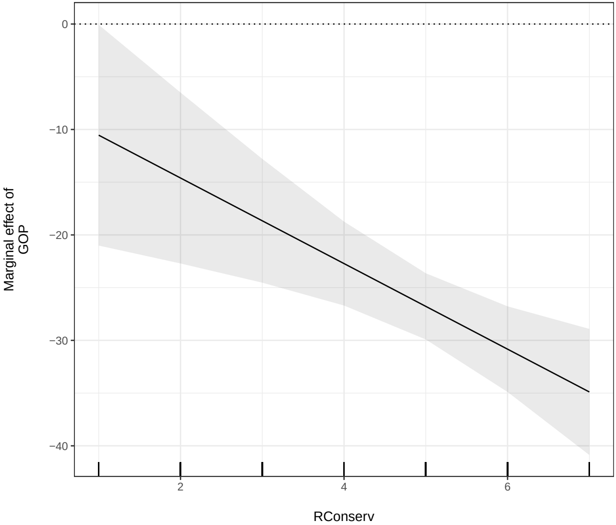
Marginal Effects Plots, I

```
> ConsSim<-seq(1,7,1)
> psis<-fit1$coeff[3]+(ConsSim * fit1$coeff[4])
> psis.ses<-sqrt(vcov(fit1)[3,3] +
  (ConsSim)^2*vcov(fit1)[4,4] + 2*ConsSim*vcov(fit1)[3,4])

> plot(ConsSim,psis,t="l",lwd=2,xlab="Respondent Conservatism",
  ylab="Estimated Marginal Effect",ylim=c(-40,0))
> lines(ConsSim,psis+(1.96*psis.ses),lty=2,lwd=2)
> lines(ConsSim,psis-(1.96*psis.ses),lty=2,lwd=2)
> abline(h=0,lwd=1,lty=2)
```

The Plot...


0
−10
−20
−30
−40
Estimated Marginal Effect
1
2
3
4
5
6
7
Respondent Conservatism

Same, Using `plot_me` (`plotMElm` package)



A Simple Spatial Model

Suppose that for respondent i:

$$\begin{aligned}\texttt{Clinton Thermometer}_i &= \alpha + \delta(\texttt{Respondent's Ideology}_i - \\ &\quad \texttt{Clinton's Ideology}_i)^2 + \epsilon_i\end{aligned}$$

with $\delta < 0$. Note that:

$$\begin{aligned}\texttt{Clinton Thermometer}_i &= \alpha + \delta[\texttt{R's Ideology}_i^2 + \\ &\quad \texttt{Clinton's Ideology}_i^2 - \\ &\quad 2(\texttt{R's Ideology}_i \times \texttt{C's Ideology}_i)] + \epsilon_i \\ &= \alpha + \delta_1(\texttt{R's Ideology}_i^2) + \\ &\quad \delta_2(\texttt{Clinton's Ideology}_i^2) + \\ &\quad \delta_3(\texttt{R's Ideology}_i \times \texttt{C's Ideology}_i)] + \epsilon_i\end{aligned}$$

Interacting Two Continuous Covariates

```
> fit2<-with(ClintonTherm,
+       lm(ClintonTherm~RConserv+ClintonConserv+RConserv*ClintonConserv))
> summary(fit2)

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

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)             119.3515     5.1634  23.115  < 2e-16 ***
RConserv                -19.5673     1.0362 -18.884  < 2e-16 ***
ClintonConserv           -7.9311     1.6477  -4.813 1.66e-06 ***
RConserv:ClintonConserv   3.6293     0.3394  10.695  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 22.03 on 1293 degrees of freedom
Multiple R-squared: 0.4619, Adjusted R-squared: 0.4606
F-statistic:   370 on 3 and 1293 DF,  p-value: < 2.2e-16
```

```
> fit2$coef[2]+(1*fit2$coef[4])
 RConserv
-15.93803

> sqrt(vcov(fit2)[2,2] + (1)^2*vcov(fit2)[4,4] + 2*1*vcov(fit2)[2,4])
[1] 0.7439696
```

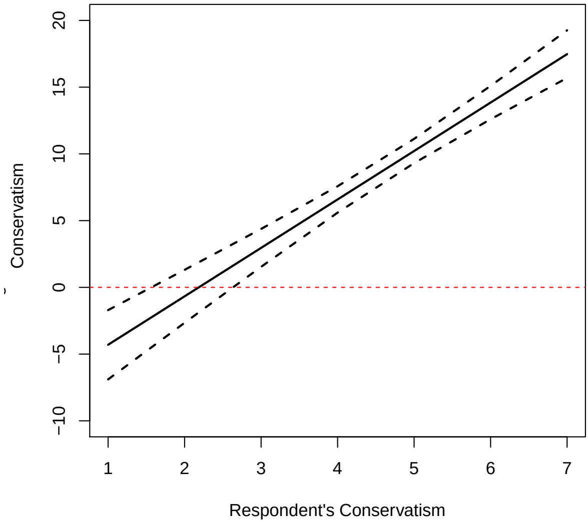
```
> linearHypothesis(fit2,"RConserv+1*RConserv:ClintonConserv")
Linear hypothesis test

Hypothesis:
RConserv  + RConserv:ClintonConserv = 0

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

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1   1294 850442
2   1293 627658  1    222784 458.94 < 2.2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
> # psi_1 | ClintonConserv = mean
> fit2$coef[2]+((mean(ClintonTherm$ClintonConserv))*fit2$coef[4])

 RConserv
-8.735424

> sqrt(vcov(fit2)[2,2] + (mean(ClintonTherm$ClintonConserv)^2*vcov(fit2)[4,4] +
+                            2*(mean(ClintonTherm$ClintonConserv))*vcov(fit2)[2,4]))

[1] 0.4507971

> pt(((fit2$coef[2]+(2.985*fit2$coef[4])) / sqrt(vcov(fit2)[2,2] +
+     (2.985)^2*vcov(fit2)[4,4] + 2*2.985*vcov(fit2)[2,4])),df=1293)

   RConserv
6.483788e-74

> # psi_2 | RConserv = 1
> fit2$coef[3]+(1*fit2$coef[4])

ClintonConserv
     -4.301803

> # psi_2 | RConserv = 6
> fit2$coef[3]+(6*fit2$coef[4])

ClintonConserv
      13.84463
```

Marginal Effect Plot, II

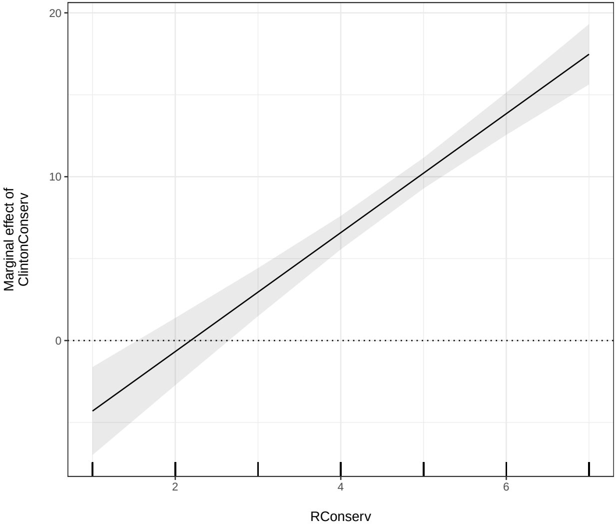
```
> psis2<-fit2$coef[3]+(ConsSim*fit2$coef[4])
> psis2.ses<-sqrt(vcov(fit2)[3,3] + (ConsSim)^2*vcov(fit2)[4,4]
  + 2*ConsSim*vcov(fit2)[3,4])

> plot(ConsSim,psis2,t="l",lwd=2,xlab="Respondent's
  Conservatism",ylab="Marginal Effect of Clinton's
  Conservatism",ylim=c(-10,20))
> lines(ConsSim,psis2+(1.96*psis2.ses),lty=2,lwd=2)
> lines(ConsSim,psis2-(1.96*psis2.ses),lty=2,lwd=2)
> abline(h=0,lty=2,lwd=1,col="red")
```

The Plot...



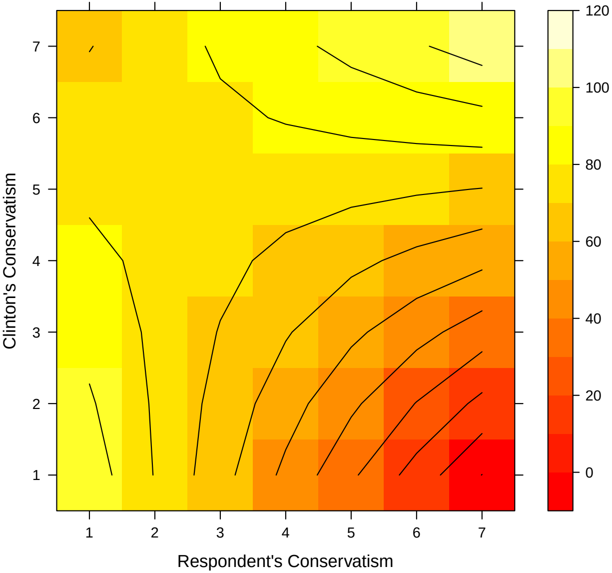
20
15
10
5
0
−5
−10
1
2
3
4
5
6
7
Conservatism
Respondent's Conservatism

Same, Using `plot_me`



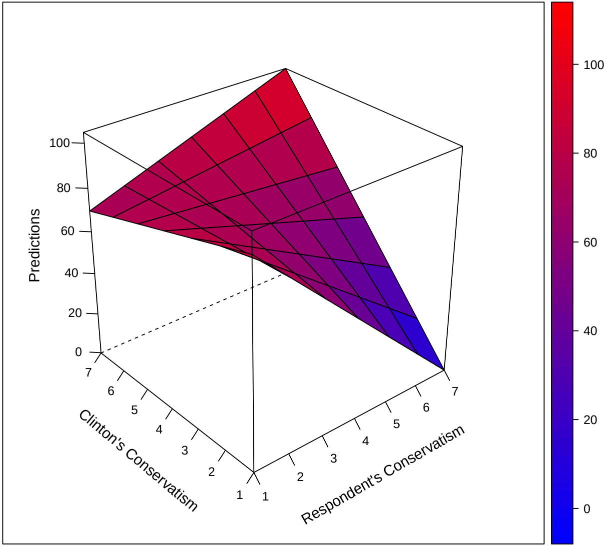
Predicted Values: A Contour Plot

```
> library(lattice)
> grid<-expand.grid(RConserv=seq(1,7,1),
  ClintonConserv=seq(1,7,1))
> hats<-predict(fit2,newdata=grid)

> levelplot(hats~grid$RConserv*grid$ClintonConserv,
  contour=TRUE,
  cuts=12,pretty=TRUE,xlab="Respondent's Conservatism",
  ylab="Clinton's Conservatism",
  col.regions=heat.colors)
```


120
100
80
60
40
20
0
7
6
5
4
3
2
1
Clinton's Conservatism
1
2
3
4
5
6
7
Respondent's Conservatism

Predicted Values: A Wireframe Plot

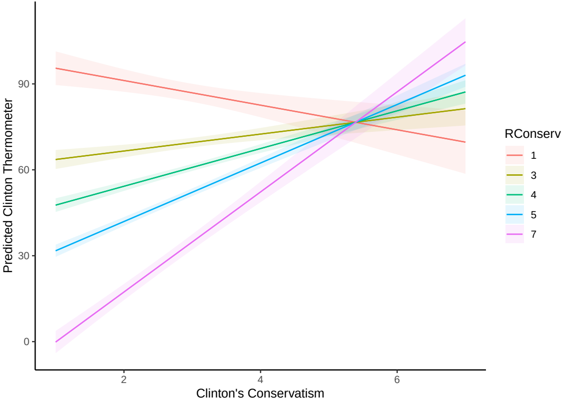
```
> trellis.par.set("axis.line",list(col="transparent"))

> wireframe(hats~grid$RConserv*grid$ClintonConserv,
  drape=TRUE,
  xlab=list("Respondent's Conservatism",rot=30),
  ylab=list("Clinton's Conservatism",
  rot=-40),zlab=list("Predictions",rot=90),
  scales=list(arrows=FALSE,col="black"),
  zoom=0.85,pretty=TRUE),
  col.regions=colorRampPalette(c("blue","red"))(100))
```



Predictions
100
80
60
40
20
0
Clinton's Conservatism
1
2
3
4
5
6
7
Respondent's Conservatism
1
2
3
4
5
6
7
100
80
60
40
20
0

Predictions Using marginaleffects

```
> library(marginaleffects)
> p <- plot_predictions(fit2,condition=c("ClintonConserv","RConserv")) +
       theme_classic() +
       labs(x = "Clinton's Conservatism",
            y = "Predicted Clinton Thermometer")
> p
```



Interpreting Interactions: Other Tools

A partial list of packages and things:

- All the various model summary and plotting tools in Vincent Arel-Bundock's `modelsummary` and `marginaleffects` packages
- `interactions` (vignette)
- `emmeans` (see, e.g., this vignette)
- `plot_model` in `sjPlot` (tutorial)
- `interactionR`
- `interplot`
- `InteractionPoweR` (power analyses for interactive models)