Interaction-Continuous Predictors

Paul E. Johnson\textsuperscript{1}  \textsuperscript{2}

\textsuperscript{1}Department of Political Science
\textsuperscript{2}Center for Research Methods and Data Analysis, University of Kansas

2014
Note: This demonstrates some features in rockchalk
Outline

1. Introduction

2. Why $x_{1i} \cdot x_{2i}$?

3. Simple Slopes

4. Weird t / p Value problem and the Mirage of “Centering”
Outline

1. Introduction
2. Why $x_{1i} \cdot x_{2i}$?
3. Simple Slopes
4. Weird t / p Value problem and the Mirage of “Centering”
Definition: Interaction

- **Linear Model:**
  \[ y_i = b_0 + b_1 x_1 i + b_2 x_2 i + e_i \]

- **Social/Behavioral researchers often assert an additional “interaction effect”**
  \[ y_i = b_0 + b_1 x_1 i + b_2 x_2 i + b_3 x_1 i \cdot x_2 i + e_i \]
Outline

1 Introduction

2 Why $x_{1i} \cdot x_{2i}$?

3 Simple Slopes

4 Weird t / p Value problem and the Mirage of “Centering”
Justification 1. The “Moderated Slope” Model

- My best explanation in English:
  
  *The effect of one variable depends on another.*

- Moderator: name for a variable that “moderates” (changes) the effect of a variable
Examples Of Interaction in Literature


Visualize the Slope’s Dependence: \( x_1; \) influence depends
Suppose: $x_2$ moderates $x_1$.

Re-group the terms so that this:

$$y_i = b_0 + b_1x_1 + b_2x_2 + b_3x_2 \cdot x_1 + e_i$$  \hspace{1cm} (1)

becomes this:

$$y_i = b_0 + (b_1 + b_3x_2)x_1 + b_2x_2 + e_i$$  \hspace{1cm} (2)

- $b_3$ is an “interaction effect”
- $b_1$, $b_2$ often called the “main effects” of $x_1$ and $x_2$
Concentrate: Interpret Coefficients!

\[ y_i = b_0 + (b_1 + b_3 x_2) x_1 + b_2 x_2 + e_i \]

Note substantive importance of 
\[ b_1 \mathrel{\overset{\Leftrightarrow}{\gtrless}} - b_3 x_2 \].

- If \( > \), the marginal effect of \( x_1 \) is positive
- if \( < \), the marginal effect of \( x_1 \) is negative
- If \( = \), then \( x_1 \) has no marginal effect

I’ll explain how to make this later:
Warning: Always include $x_1$ and $x_2$ if you fit $x_1 \cdot x_2$ as well.

- If a model includes an interaction $x_1 \cdot x_2$, it should always include $x_1$ and $x_2$ (even if they appear to be “not statistically significant”).
- $x_1$ and $x_2$ are said to be “marginal” to $x_1 \cdot x_2$
- Wm Venables, “Exegesis on Linear Models”
  
  http://www.stats.ox.ac.uk/pub/MASS3/Exegeses.pdf
Reason 2: Approximation of a Function

- Taylor’s Theorem says that any function $f$ at a point $x$, $f(x_1, x_2)$, can be approximated by a clever choice of coefficients.

\[
y = f(x_1, x_2) + \beta_1(x_1 - x_1_0) + \beta_2(x_2 - x_2_0) + \beta_3(x_1 - x_1_0)(x_2 - x_2_0) + \frac{1}{2}\beta_4(x_1 - x_1_0)^2 + ... \tag{3}
\]

\[
+ \beta_3(x_1 - x_1_0)(x_2 - x_2_0) + \frac{1}{2}\beta_4(x_1 - x_1_0)^2 + ... \tag{4}
\]

- $(x_1, x_2)$ is value where we “approximate from”

- If curvature of $f$ is mild, then the Taylor approximation will stay close to true values.

- We throw away the higher order terms, asserting/hoping they are small.
An Identification Problem

- Identification: Ability to estimate parameters with data at hand.
- The theoretical model boils down to this:

\[ y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{1i} \cdot x_{2i} + e_i \]  \hspace{1cm} (5)

- Expression (5) is equivalent to both of these interpretations:
  - \( x_{1i} \)'s slope depends on \( x_{2i} \)
    \[ y_i = b_0 + (b_1 + b_3 x_{2i}) \cdot x_{1i} + b_2 x_{2i} + e_i \]  \hspace{1cm} (6)
  - \( x_{2i} \)'s slope depends on \( x_{1i} \).
    \[ y_i = b_0 + b_1 x_{1i} + (b_2 + b_3 x_{1i}) \cdot x_{2i} + e_i \]  \hspace{1cm} (7)

- Data cannot differentiate those 2 models, hence we say there is an “identification problem”.
Outline

1. Introduction

2. Why $x_{1i} \cdot x_{2i}$?

3. Simple Slopes

4. Weird t / p Value problem and the Mirage of “Centering”
Same Advice, many disciplines

Recent Chorus:
Must compare predicted values from various predictor combinations to understand their effects.
Each predictors “example values” must be set and understood while focusing on some particular values.

- Psychology

- Political Science:
Same Advice, many disciplines ... 


Economics:

Choose Some “For Instance” Values of One Variable, Plot the other

- We want to make a 2D plot of $y_i$ on $x_{1i}$
- User must supply some “substantively interesting values” of $x_{2i}$, say 1, 2, 3, so that this

$$y_i = b_0 + (b_1 + b_3 x_{2i}) \cdot x_{1i} + b_2 x_{2i} + e_i \quad (8)$$

- Generates a family of lines,

$$x_{2i} = 1 \quad : \quad y_i = (b_0 + b_2 \cdot (1)) + (b_1 + b_3 (1)) \cdot x_{1i} + e_i \quad (9)$$
$$x_{2i} = 2 \quad : \quad y_i = (b_0 + b_2 \cdot (2)) + (b_1 + b_3 (2)) \cdot x_{1i} + e_i \quad (10)$$
$$x_{2i} = 3 \quad : \quad y_i = (b_0 + b_2 \cdot (3)) + (b_1 + b_3 (3)) \cdot x_{1i} + e_i \quad (11)$$
The Marginal Effect of $x_1$ is $(b_1 + b_3 \times \text{moderator}_i)$

- The full model might be
  \[
  y_i = b_0 + b_1 x_1 i + b_2 x_2 i + b_3 x_1 i \cdot x_2 i + e_i 
  \]  
  (12)

- You choose $x_2 i = \text{whatever value is interesting for the moderator}$:
  \[
  y_i = (b_0 + b_2 \cdot (\text{whatever})) + (b_1 + b_3 \cdot (\text{whatever})) \cdot x_1 i + e_i 
  \]  
  (13)

- After you set a particular $x_2 i = \text{whatever}$, then the line for the “simple slope” has
  - The “New Intercept:” $(b_0 + b_2 \text{whatever})$: AKA “shifted intercept”
  - The “New Slope:” $(b_1 + b_3 \text{whatever})$: AKA the “marginal effect” of $x_1 i$. 
Suppose $y_i = 2 + 0.5 \cdot x_1 + 0.2 \cdot x_2 + 0.03 \cdot x_1 \cdot x_2$
One Line Per Value of $x_2$; while plotting Simple Slope
Translate Between 3-D and the Simple Slopes in 2-D

These four lines are highlighted in yellow in the graph on the right.
Ways To Choose "whatever" Values

- Problem specific interesting cases that suit your project!
  - Fahrenheit temperatures? Pick \{32, 100, 212\}
  - Salary (dollars)? Pick \{20,000, 100,000, 500,000, 1,000,000\}
- The rockchalk package has routines to choose, based either on
  - quantiles (break a range into values that correspond with, for example, the lowest 25%, the median (50%), and the top 75%).
    - I originally developed the plotSlopes function with this in mind
  - standard deviation-based ranges.
    - psychologists suggest it is easier for them to conceptualize special values like the mean, the mean - 1 standard deviation, mean + 1 standard deviation, and so forth.
A Special Hypothesis Test

- The simple line $y_i = (2 + 0.2 \times x_{2i}) + (0.5 + 0.03 \times x_{2i}) \times x_{1i}$
- Concentrate on the slope, the “marginal effect”: For a given value of $x_{2i}$, of course, that is just a sum like

$$\text{for } x_{2i} : \text{whatever, the slope is } 0.5 + 0.03 \times \text{whatever} \quad (14)$$

- Some people ask, “is that particular slope statistically significantly different from 0?”
And a Fancy T-Test Pops Out (Not Entirely Unexpected)

- Suppose whatever $= 10$. They are asking “is the estimate $(0.5 + 0.03 \times 10)$ statistically significantly different from 0?”
- And if you put in estimates from a regression, that’s a fancy t-test

$$H_0 : b_1 + b_3x_2 = 0$$  \hspace{1cm} (15)

$$\hat{t} = \frac{\hat{b}_1 + \hat{b}_3x_2}{\sqrt{\text{Var}(\hat{b}_1 + \hat{b}_3x_2)}}$$, $x_2$ is selected value  \hspace{1cm} (16)

$$\hat{t} = \frac{\hat{b}_1 + \hat{b}_3x_2}{\sqrt{\text{Var}(\hat{b}_1) + x_2^2 \text{Var}(\hat{b}_3) + 2x_2 \text{Cov}(\hat{b}_1, \hat{b}_3)}}$$  \hspace{1cm} (17)
J-N Interval: The *whatever* over and over problem

- Imagine letting your research director says, over and over again,
  - What if $x_{20} = 10$, is it significant then?
  - What if $x_{20} = 13$, is it significant then?
- That drives you crazy! Over and over, you calculate
  \[
  \hat{t} = \frac{\hat{b}_1 + \hat{b}_3x^2}{\sqrt{Var(\hat{b}_1) + x_{20}^2 Var(\hat{b}_3) + 2x_{20} Cov(\hat{b}_1, \hat{b}_3)}}
  \]  
  (18)
- Wish you could find a formula to say “$\hat{b}_1 + \hat{b}_3x^2$ is statistically significant if $x^2$ is in ‘this range’?”
- It is necessary to solve for $|\hat{t}| > 1.98$, to get the values of $x_{20}$ that cause $\hat{t}$ to be statistically different from zero.
- That interval is known as the Johnson-Neyman interval.
rockchalk plotting approaches for both of these

- Described in vignette (run vignette(“rockchalk”) with rockchalk version 1.5.4 or later).
- Step 1: use regression to fit a model with multiplicative terms
- Plot Type 1: plotSlopes() will to draw the 2 dimensional plot with several lines, one for each value of a moderator
- Plot Type 2: The “J-N interval” plot.
  - The testSlopes() function finds an interval on which the marginal effect is not 0.
  - A plot method for testSlopes objects creates a “marginal effect” plot. I find these confusing, but some people love them!
Basic Idea Behind: plotSlopes, plotCurves, plotPlane

- Fit any regression with interactions and as many other variables you like
  
  ```
  m1 <- lm(y ~ x1 * x2 + x3 + x4, data=dat)
  ```

- You want to focus on the predictive effect of \(x_1\) and \(x_2\).

- draw a plot with \(x_2\) on the horizontal axis and a line for some focal values of modx.

  ```
  m1ps <- plotSlopes(m1, plotx = "x2", modx = "x1")
  ```

- That creates a plot, but also an output object for further analysis.

- Give the output to testSlopes(), like so:

  ```
  m1psts <- testSlopes(m1ps)
  ```

- There is a plot method for that type of object

  ```
  plot(m1psts)
  ```
Basic Idea Behind: plotSlopes, plotCurves, plotPlane ...

- Make a reasonable 3D plot of the pair of variables:
  - the defaults are mostly good
    ```r
    m1pp <- plotPlane(m1, plotx1="x1", plotx2="x2")
    ```
  - there are many options that can customize
    ```r
    m1pp <- plotPlane(m1, plotx1="x1", plotx2="x2", plotPoints=F, drawArrows=F, ticktype="detailed", theta=-20, npp=7)
    ```

- plotSlopes has to generate a particular kind of output object for its eventual input into testSlopes(). If we ignore that problem, then we can have a more flexible line plotter. That is called plotCurves().

- Unlike plotSlopes(), the plotCurves() and plotPlane() functions can handle many types of nonlinear functions. They do work (or should work) on formula like
  ```r
  alpha <- 2
  m2 <- lm(y ~ x1*log(x2 + alpha) + sin(x3) + x4, data=dat)
  ```
In rockchalk 1.6.3, a new function called addLines() was introduced. It can take the lines from a “plotSlopes()” or “plotCurves()” and superimpose them on the 3d output from plotPlane. That *should* make a plot of the sort displayed above “Translate Between 3-D and ...”
Default plotSlopes output

```r
m1ps <- plotSlopes(m1, plotx="x1", modx="x2")
```
plotSlopes variations

```r
m1ps <- plotSlopes(m1, plotx="x1",modx="x2",modxVals = "std.dev", n = 2, interval = "conf")
```
plotSlopes or plotCurves equivalents are in the literature

**Figure 2** The Impact of Viability and Intraparty Competition on the Likelihood of Cheating

*Notes:* The figure reports simulations of the impact of the interaction of viability and intraparty competition on probability of cheating based on Table 2 Model 3, with other variables held at their median values. The lower line with a light gray 95% confidence interval represents the probability of cheating as viability varies with no intraparty competition. The upper line with dark gray confidence intervals represents the probability of cheating with high levels of intraparty competition (six other candidates in the same camp).

Admittedly, that’s a nonlinear model, but it shows the same basic thing. predicted values for 2 moderator values.
Ask `plotSlopes` for Confidence Intervals if you want

```r
m1ps2 <- plotSlopes(m1, plotx="x1", modx="x2", interval = "conf", modxVals = round(range(dat$x2, na.rm=TRUE), 2))
```
J-N confidence interval: testSlopes

\[
m1psts <- \text{testSlopes}(m1ps)
\]

Values of \(x_2\) INSIDE this interval:

<table>
<thead>
<tr>
<th>lo</th>
<th>hi</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.30582</td>
<td>122.54847</td>
</tr>
</tbody>
</table>

cause the slope of \((b_1 + b_2 \times x_2) \times x_1\) to be statistically significant
Rather than calculating that interval, some would plot it.

Figure 3. Marginal effect of Parliamentarism on Political_Deaths across levels of ethnic fractionalization.
plot.testSlopes tries to make this type of plot more clear

```
plot(m1psts)
```

- Marginal Effect of $x_1$: $(\hat{b}_{x_1} + \hat{b}_{x_2} : x_1 \cdot x_2)_i$
  - 20.31
- 95% Conf. Int.
- Shaded Region: Null Hypothesis $b_{x_1} + b_{x_2} : x_1 \cdot x_2_i = 0$ rejected
To me, this is a more understandable representation
I like the 3d representation for a simple problem
Can Superimpose one on the other
The Alternative J-N Plot is Preferred by Many

\[
\text{Marginal Effect of } x_1 : (\hat{b}_{x_1} + \hat{b}_{x_1x_2} x_2)
\]

\[
\begin{align*}
-3.19 & \quad 24.52
\end{align*}
\]

95% Conf. Int.

Shaded Region: Null Hypothesis

\[
b_{x_1} + b_{x_1x_2} x_2 = 0 \text{ rejected}
\]
testSlopes can have various patterns of significant regions

The Moderator:  $x_2$

Marginal Effect of $x_1$: $(\hat{b}_{x_1} + \hat{b}_{x_2:x_1}x_2)_i$

55.16

Marginal Effect

95% Conf. Int.

Shaded Region: Null Hypothesis

$b_{x_1} + b_{x_2:x_1}x_2_i = 0$ rejected

The Moderator:  $x_2$
Outline

1. Introduction

2. Why $x_{1i} \cdot x_{2i}$?

3. Simple Slopes

4. Weird t / p Value problem and the Mirage of “Centering”
Centering: My Theme

- rockchalk vignette has long-ish explanation of mean-centering
- Recall quadratic regression lecture, “centering does not really help” with multicollinearity between $x$ and $x^2$.
- If you “really understand” regression, you should see that centering doesn’t help here either
- It does not “fix” a multicollinearity problem, it was a mistake for its proponents to think so
- Centering does facilitate superficial interpretation in one situation:
  - centering of all $x$’s has effect of making the intercept the predicted value for the “mean case”.
  - Intercept is same as using non-centered model to calculate predicted value: $\hat{y}_i = \hat{b}_0 + \hat{b}_1 x + \hat{b}_2 x^2 + \hat{b}_3 x^3 + ...$
3 Variables I Found Lying About

```r
dat <- genCorrelatedData(N=400, rho=0.1, stde=250, beta=c(2,0.1,-0.1,0.5))
m1 <- lm(y ~ x1 + x2, data=dat)
fit1 <- plotPlane(m1, plotx1="x1", plotx2="x2", ticktype="detailed")
```

That manufactures data with the true coefficients

\[
y = 2 + 0.1x_1 - 0.1x_2 + 0.3x_1 \cdot x_2 + e_i, \; e_i \sim N(0, 300^2) \quad (19)
\]
The 3D Plot for that
## The Fitted Linear Model

```r
Call:
lm(formula = y ~ x1 + x2, data = dat)

Residuals:
   Min     1Q  Median     3Q    Max
-646.23 -179.72  -22.99  152.93  726.21

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)          -1464.760   92.051  -15.91  <2e-16 ***
x1                    26.539     1.388   19.12  <2e-16 ***
x2                    27.837     1.339   20.79  <2e-16 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 259.4 on 397 degrees of freedom
Multiple R^2:  0.6972   Adjusted R^2:  0.6957
F-statistic:  457 on 2 and 397 DF,  p-value: < 2.2e-16
```
That model seems persuasive!

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate (S.E.)</td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-1464.760***</td>
</tr>
<tr>
<td></td>
<td>(92.051)</td>
</tr>
<tr>
<td>x1</td>
<td>26.539***</td>
</tr>
<tr>
<td></td>
<td>(1.388)</td>
</tr>
<tr>
<td>x2</td>
<td>27.837***</td>
</tr>
<tr>
<td></td>
<td>(1.339)</td>
</tr>
<tr>
<td>N</td>
<td>400</td>
</tr>
<tr>
<td>RMSE</td>
<td>259.449</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.697</td>
</tr>
<tr>
<td>adj $R^2$</td>
<td>0.696</td>
</tr>
</tbody>
</table>

* $p \leq 0.05$  ** $p \leq 0.01$  *** $p \leq 0.001$

- Plenty of “stars” indicating statistical significance!
- Easy to interpret parameter estimates
"Throw In" an Interaction to "See" if it is Needed

- Researcher wonders, “should I add $x_1 \times x_2$ as a predictor?”
- Code change
  - Change the model from
    ```r
    m1 <- lm(y ~ x1 + x2, data=dat)
    ```
  - To:
    ```r
    m2 <- lm(y ~ x1*x2, data=dat)
    ```
- R will automatically return equivalent of
  ```r
  m2 <- lm(y ~ x1 + x2 + x1:x2, data=dat)
  ```
Descriptive

Weird t / p Value problem and the Mirage of "Centering"

Oh My God! Your p’s Exploded

m2 <- lm(y ~ x1*x2, data=dat)
summary(m2)

Call:
  lm(formula = y ~ x1 * x2, data = dat)

Residuals:
   Min     1Q   Median     3Q    Max
-634.81 -173.75 -17.76  160.99  715.48

Coefficients:
                  Estimate Std. Error   t value  Pr(>|t|)
(Intercept)   -424.2604   379.0010   -1.119   0.26364
x1             6.5226     7.2091    0.905    0.36613
x2             7.3553     7.3616    0.999    0.31833
x1:x2         -0.3922     0.1387    2.829    0.00491 **

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 257.2 on 396 degrees of freedom
Multiple R^2:  0.7032 ,  Adjusted R^2:  0.7009
F-statistic: 312.7 on 3 and 396 DF,  p-value: < 2.2e-16
Problem: “Nothing is significant anymore!”

<table>
<thead>
<tr>
<th>M1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
</tr>
<tr>
<td>(S.E.)</td>
</tr>
<tr>
<td>(Intercept)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>x1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>x2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>x1:x2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>RMSE</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
<tr>
<td>$adj\ R^2$</td>
</tr>
</tbody>
</table>

$p \leq 0.05\ast\ast p \leq 0.01\ast\ast\ast p \leq 0.001$
Aiken & West (later, Cohen, Cohen, West and Aiken) claim the constructed variable \( x_1_i \cdot x_2_i \) is multi-collinear with \( x_1_i \) and \( x_2_i \), thus causing the results to become “poor.”

As we recall, multicollinearity causes standard errors to inflate, and \( \hat{t} \)’s shrink.

They recommend “mean-centering” as a way to ameliorate the “nonessential collinearity”.
Weird t / p Value problem and the Mirage of “Centering”

CCWA advice

- p. 266
  “We recommend that continuous predictors be centered before being entered into regression analyses containing interactions. ... Doing so yields two straightforward, meaningful interpretations of each first-order regression coefficient of predictors entered into the regression equation: (1) effects of the individual predictors at the mean of the sample, and (2) average effects of each individual predictors across the range of the other variables. Doing so also eliminates nonessential multicollinearity between first-order predictors and predictors that carry their interaction with other predictors.”

- This advice has been followed VERY widely.

- My counter-argument will be that
  - benefits 1 and 2 are not wrong, but not beneficial either, and
  - the “nonessential multicollinearity” argument is just completely wrong.
Run Again, But Center the Data First

- We “manually” center predictors (either use R’s scale() function or the more literal):

  ```r
  dat$x1c <- dat$x1 - mean(dat$x1, na.rm = TRUE)
dat$x2c <- dat$x2 - mean(dat$x2, na.rm = TRUE)
m3 <- lm(y ~ x1c * x2c, data=dat)
  ```

- meanCenter() in rockchalk will do this for us:

  ```r
  m2centered <- meanCenter(m2)
snmary(m2centered)
fit1 <- plotPlane(m2centered, plotx1="x1c", plotx2="x2c",
ticktype="detailed")
  ```
meanCenter Output

These variables were mean-centered before any transformations were made on the design matrix.

[1] "x1c" "x2c"

The centers and scale factors were

\[
\begin{align*}
\text{mean} & : & 50.48552 & & 50.69844 \\
\text{scale} & : & 1.00000 & & 1.00000 
\end{align*}
\]

The summary statistics of the variables in the design matrix (after centering).

\[
\begin{array}{ccc}
\text{mean} & \text{std. dev.} \\
\hline
y & 1286.3620 & 470.3068 \\
x1c & 0.0000 & 9.4335 \\
x2c & 0.0000 & 9.7806 \\
x1c:x2c & 11.7519 & 94.5357 \\
\end{array}
\]

The following results were produced from:

\[
\text{meanCenter.default(model = m2)}
\]

Call:

\[
\text{lm(formula = y \sim x1c \ast x2c, data = stddat)}
\]

Residuals:

\[
\begin{array}{cccccc}
\text{Min} & \text{1Q} & \text{Median} & \text{3Q} & \text{Max} \\
-634.81 & -173.75 & -17.76 & 160.99 & 715.48 \\
\end{array}
\]
Weird t / p Value problem and the Mirage of “Centering”

meanCenter Output ...

| Coefficients        | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------------|----------|------------|---------|----------|
| (Intercept)         | 1281.7531| 12.9624    | 98.883  | < 2e-16  *** |
| x1c                 | 26.4058  | 1.3770     | 19.177  | < 2e-16  *** |
| x2c                 | 27.1550  | 1.3490     | 20.129  | < 2e-16  *** |
| x1c:x2c             | 0.3922   | 0.1387     | 2.829   | 0.00491  ** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 257.2 on 396 degrees of freedom

Multiple R^2: 0.7032, Adjusted R^2: 0.7009

F-statistic: 312.7 on 3 and 396 DF, p-value: < 2.2e-16
Plots Same, but moving Y-axis makes fitted models appear different!

The NON CENTERED FIT

The CENTERED FIT

- Recall beginning of course: rescaling by subtraction shifts intercept, does not change slope
- Find the “y axis” in each plot. Understand why centering seems to matter now?
Look at your p’s. The Centered Fit Is Super Awesome!

```
outreg (m2centered)

<table>
<thead>
<tr>
<th></th>
<th>Estimate (S.E.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1281.753***</td>
</tr>
<tr>
<td></td>
<td>(12.962)</td>
</tr>
<tr>
<td>x1c</td>
<td>26.406***</td>
</tr>
<tr>
<td></td>
<td>(1.377)</td>
</tr>
<tr>
<td>x2c</td>
<td>27.155***</td>
</tr>
<tr>
<td></td>
<td>(1.349)</td>
</tr>
<tr>
<td>x1c:x2c</td>
<td>0.392**</td>
</tr>
<tr>
<td></td>
<td>(0.139)</td>
</tr>
</tbody>
</table>

N 400
RMSE 257.191
$R^2$ 0.703
adj $R^2$ 0.701

*p ≤ 0.05 **p ≤ 0.01 ***p ≤ 0.001
Don’t Get Carried Away: It’s the Same Model!

<table>
<thead>
<tr>
<th></th>
<th>Not Centered Estimate (S.E.)</th>
<th>Centered Estimate (S.E.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-424.260 (379.001)</td>
<td>1281.753*** (12.962)</td>
</tr>
<tr>
<td>x1</td>
<td>6.523 (7.209)</td>
<td>.</td>
</tr>
<tr>
<td>x2</td>
<td>7.355 (7.362)</td>
<td>.</td>
</tr>
<tr>
<td>x1:x2</td>
<td>0.392** (0.139)</td>
<td>.</td>
</tr>
<tr>
<td>x1c</td>
<td>.</td>
<td>26.406*** (1.377)</td>
</tr>
<tr>
<td>x2c</td>
<td>.</td>
<td>27.155*** (1.349)</td>
</tr>
<tr>
<td>x1c:x2c</td>
<td>.</td>
<td>0.392** (0.139)</td>
</tr>
<tr>
<td>N</td>
<td>400</td>
<td>400</td>
</tr>
<tr>
<td>RMSE</td>
<td>257.191</td>
<td>257.191</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.703</td>
<td>0.703</td>
</tr>
<tr>
<td>adj $R^2$</td>
<td>0.701</td>
<td>0.701</td>
</tr>
</tbody>
</table>

* $p \leq 0.05$ ** $p \leq 0.01$ *** $p \leq 0.001$

As CCWA observe, the estimates for the “highest order” slope coefficients are same in both models.
Don’t Get Carried Away: Its The SAME MODEL! ...

So is $R^2$, $RMSE$, etc.
Predicted Values of The Two Models
See why they are the Same Model?

- Same predicted values from same values of input X values
- Same estimates of “slopes” at any given combination of X values
- Same uncertainty (variance) of slope estimates at any given combination of X values
- Why the interaction coefficient the same in the 2 models?
  - Answer: it is the only parameter that is a “constant” in the nonlinear model (cross partial derivative same at all points)
Can Reproduce "Mean Centered" Parameter Estimates From Uncentered Model Estimates

The uncentered model fit is

\[ \hat{y}_i = \hat{b}_0 + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} + \hat{b}_3 x_{1i} \cdot x_{2i} \]

The predicted value AT THE MEANS \( \bar{x_1} \) and \( \bar{x_2} \) is

\[ \hat{y}_{mean} = \hat{b}_0 + \hat{b}_1 \bar{x_1} + \hat{b}_2 \bar{x_2} + \hat{b}_3 \bar{x_1} \cdot \bar{x_2} \]

```r
bs <- coef(m2)

(Intercept)    1281.753
```

Which is the estimated intercept of the “centered regression”.

Weird t/p Value problem and the Mirage of “Centering”
Can Reproduce "Mean Centered" Parameter Estimates From UnCentered Model Estimates

Uncentered: \( \hat{y}_i = \hat{b}_0 + \hat{b}_1x_{1i} + \hat{b}_2x_{2i} + \hat{b}_3x_{1i} \times x_{2i} \)

The partial slope—the effect of a change in either IV—can be evaluated AT THE MEANS \( \bar{x}_1 \) and \( \bar{x}_2 \).

The effect of \( x_{1i} \) (for example) is:

\[
\frac{\partial y_i}{\partial x_i} = \hat{b}_1 + \hat{b}_3\bar{x}_2
\]

(20)

\[
\text{partial}_x1 \leftarrow \text{bs}[2] + \text{bs}[4] \times \text{mean(dat}$x2$)
\]

\[
\begin{array}{c}
\text{partial}_x1 \\
26.40583
\end{array}
\]

Which is the estimated slope of \( x_1 \) in the “centered regression”.
Can Reproduce "Mean Centered" Parameter Estimates From UnCentered Model Estimates

Previous showed the partial slope at the mean of $x_1, x_2$ is:

$$\frac{\partial y_i}{\partial x_i} = \hat{b}_1 + \hat{b}_3 x_2$$

Calculate the Variance of that estimated value:

$$Var[\hat{b}_1 + \hat{b}_3 x_2] = Var[\hat{b}_1] + x_2^2 Var[\hat{b}_3] + 2x_2 Cov(\hat{b}_1, \hat{b}_3) \quad (21)$$

```r
V <- vcov(m2)
varsum

[1] 1.895999

sqrt(varsum)

[1] 1.376953
```
Can Reproduce "Mean Centered" Parameter Estimates From UnCentered Model Estimates ...

Notice that the square root of the estimated $\text{Var}[\hat{b}_1 + \hat{b}_2 \bar{x}_2]$ is EXACTLY the same standard error that is reported in the Centered Regression for the coefficient $x_{1c}$. 
How To Explain “Centered Mirage”? 

Two components cause the illusion that the Centered Regression Line is somehow better

- Recall the uncertainty around a regression line is hour shaped. If we place the y axis into the center of the data, we are going to the smallest part of the hourglass, so the standard errors are at their smallest possible values.

- Centering (accidentally, really) may move from a “flat spot” on the curving surface to a place that has steeper slope. This will make the estimated coefficients “bigger” because we are at a steeper spot
  - intercept is $y$ at $x_1 = x_2 = 0$
  - slope coefficients $\hat{b}_2$ and $\hat{b}_3$ are linear effects of $x_1$ and $x_2$ at $x_1 = x_2 = 0$. 

Look Again

<table>
<thead>
<tr>
<th>Not Centered Data</th>
<th>Centered Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Not Centered Data" /></td>
<td><img src="image2" alt="Centered Data" /></td>
</tr>
</tbody>
</table>

"Many empirical marketing researchers commonly mean-center their moderated regression data hoping that this will improve the precision of estimates from ill conditioned, collinear data, but unfortunately, this hope is futile. Therefore, researchers using moderated regression models should not mean-center in a specious attempt to mitigate collinearity between the linear and the interaction terms. Of course, researchers may wish to mean-center for interpretive purposes and other reasons.”
"Specifically, we demonstrate that (1) in contrast to Aiken and West’s (1991) suggestion, mean centering does not improve the accuracy of numerical computation of statistical parameters, (2) it does not change the sampling accuracy of main effects, simple effects, and/or interaction effects (point estimates and standard errors are identical with or without mean centering), and (3) it does not change overall measures of fit such as R2 and adjusted-R2. It does not hurt, but it does not help, not one iota.”

See Also: