Such a nuisance (parameter):
Interactions & cross-model comparisons in binary response models

Shawna N. Smith, PhD
University of Michigan

18 June 2020
Université de Montréal
Research Chair in Electoral Studies

****With extra-super significant acknowledgment to J. Scott Long (Indiana University) & Trenton Mize (Purdue) + countless ICPSRCDA students & TAs (2009-present)*

The main takeaway:
Beta coefficients in binary response models (BRMs) function as little more than *nuisance parameters*

Nuisance parameter: *Any parameter which is not of immediate interest, but which must be accounted for in the analysis of parameters that are of interest.*

The main takeaway:
Beta coefficients in binary response models (BRMs) function as little more than ***nuisance parameters***

Nuisance parameter: *Any parameter which is not of immediate interest, but which must be accounted for in the analysis of parameters that are of interest.*

This is particularly true when trying to understand *how or whether variable effects change or differ*.

Live and Let Die (1973)

This is particularly true when trying to understand ***how or whether variable effects change or differ***.

- **Interactions** *(Is the effect of x_1 contingent on the value of x_2?)*
- **Mediation** *(Does the effect of x_1 disappear when I account for x_2?)*
- **Cross-model comparisons** *(Does the effect of x_1 differ for $sample_1$ compared to $sample_2$?)*

This is particularly true when trying to understand ***<u>how or whether variable effects change or differ</u>***.

- **Interactions** *(Is the effect of x_1 contingent on the value of x_2?)*
- **Mediation** *(Does the effect of x_1 disappear when I account for x_2?)*
- **Cross-model comparisons** *(Does the effect of x_1 differ for $sample_1$ compared to $sample_2$?)*

In all of these situations, using information about **magnitude, direction, or significance** of an effect **directly from the coefficient** may lead you to ***<u>draw an erroneous conclusion</u>***.

Nuisance: *That which causes offence,* ***annoyance, inconvenience****, or injury.*

self-discovery of a gadfly

Lecture Overview

Part 1: Review the BRM (logit/probit)

- Identification assumptions
- Functional form

Part 2: Interactions in BRMs

- Why coefficients are pointless
- What we should do instead

Part 3: Cross-model comparisons in BRM

- Why coefficients are pointless (redux)
- What we should do instead (redux)

Part 1: BRM Refresher

Part 1: BRM Refresher

Two major points:

- Identification assumptions mean the $\beta's$ are not identified (& thus of limited utility)
- Functional form means that these models are inherently interactive…
  - But extent of that interactivity depends on where your data are.

Latent variable derivation

There is an ***underlying propensity*** that generates the observed state.

- We can't directly observe $y*$ but at some point, a change in $y*$ results in a change in what we observe (e.g., 0 or 1)

Structural model

$$y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i$$

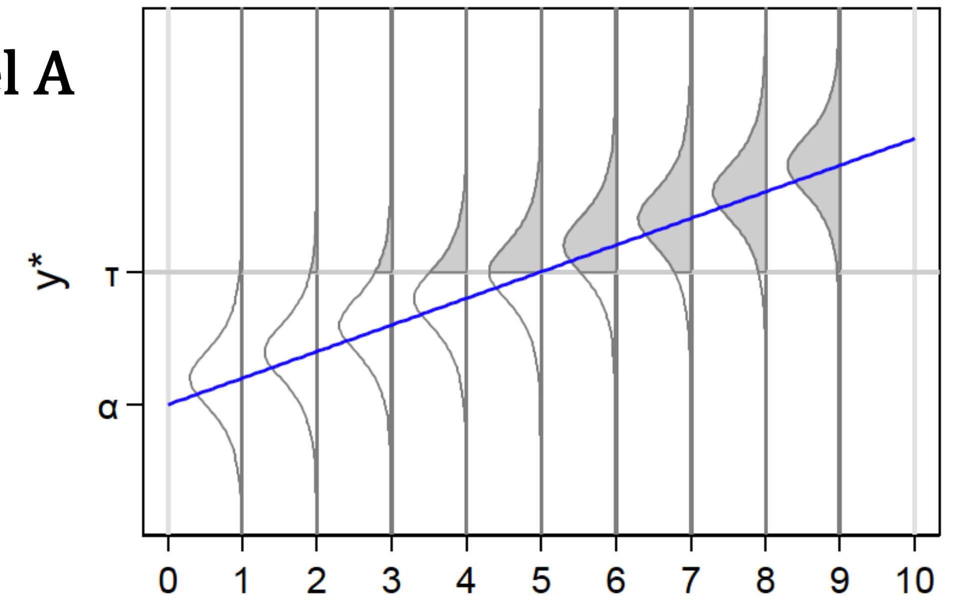
Measurement model

y^* is linked to the observed y by the measurement equation:

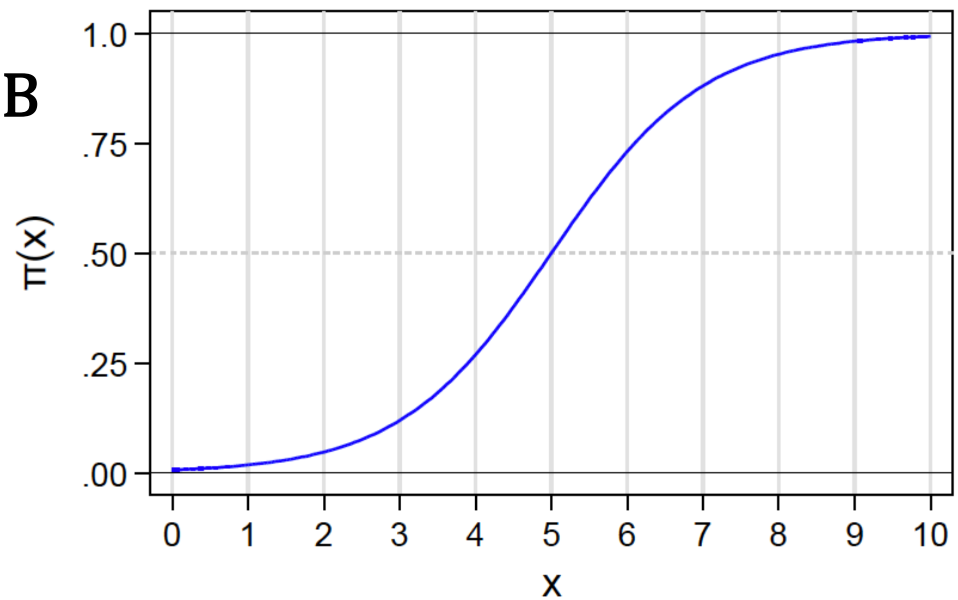
$$y = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases}$$

Graphically

Panel A



Panel B



Identification Assumption

Because we don't know the true error variance for y*, it must be assumed for the model to be identified

Logit: $Var(\varepsilon) = \frac{\pi^2}{3}$

Probit: $Var(\varepsilon) = 1$

Probit:

$$\Pr\left(y=1 \mid \mathbf{x}\right)=\int_{-\infty}^{\mathbf{x}\boldsymbol{\beta}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt = \Phi\left(\mathbf{x}\boldsymbol{\beta}\right)$$

Logit:

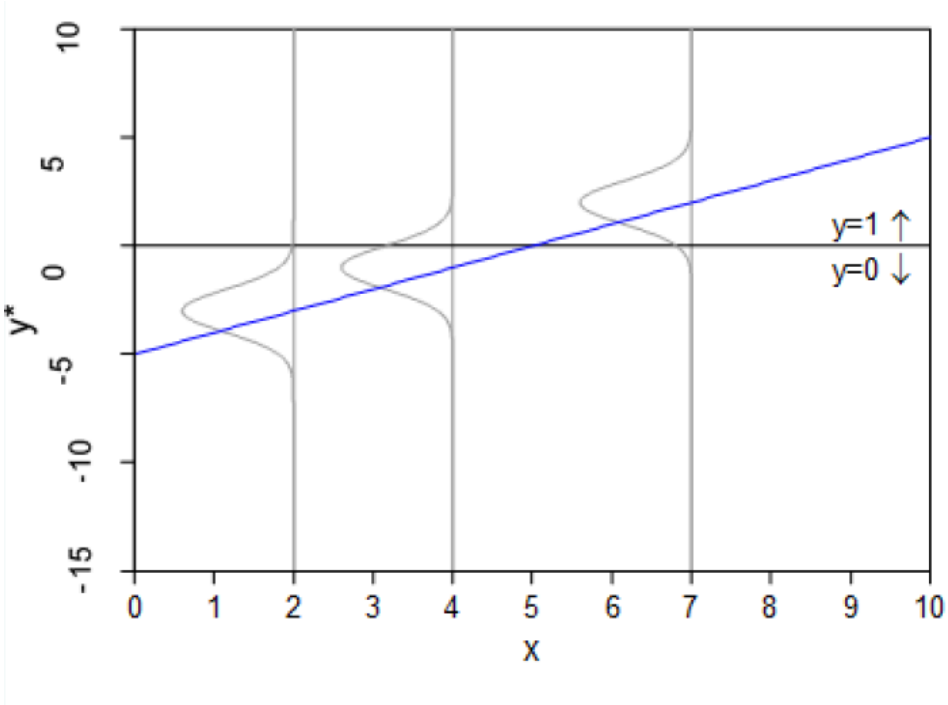
$$\Pr\left(y=1 \mid \mathbf{x}\right)=\frac{\exp\left(\mathbf{x}\boldsymbol{\beta}\right)}{1+\exp\left(\mathbf{x}\boldsymbol{\beta}\right)}=\Lambda\left(\mathbf{x}\boldsymbol{\beta}\right)$$

Consequence

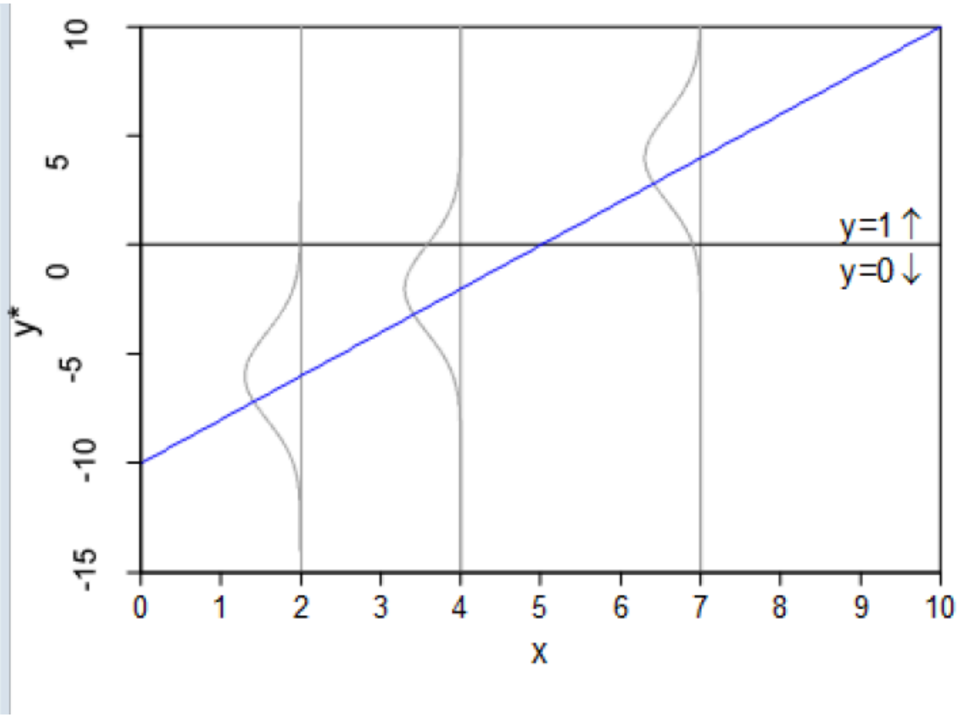
The identification assumption we make about the variance of the errors also sets our metric for our coefficients.

As such, βs reflect the magnitude of the relationship between our y* and our x's *and* the metric/scale of our underlying y*.

i.e., the βs are not identified individually.

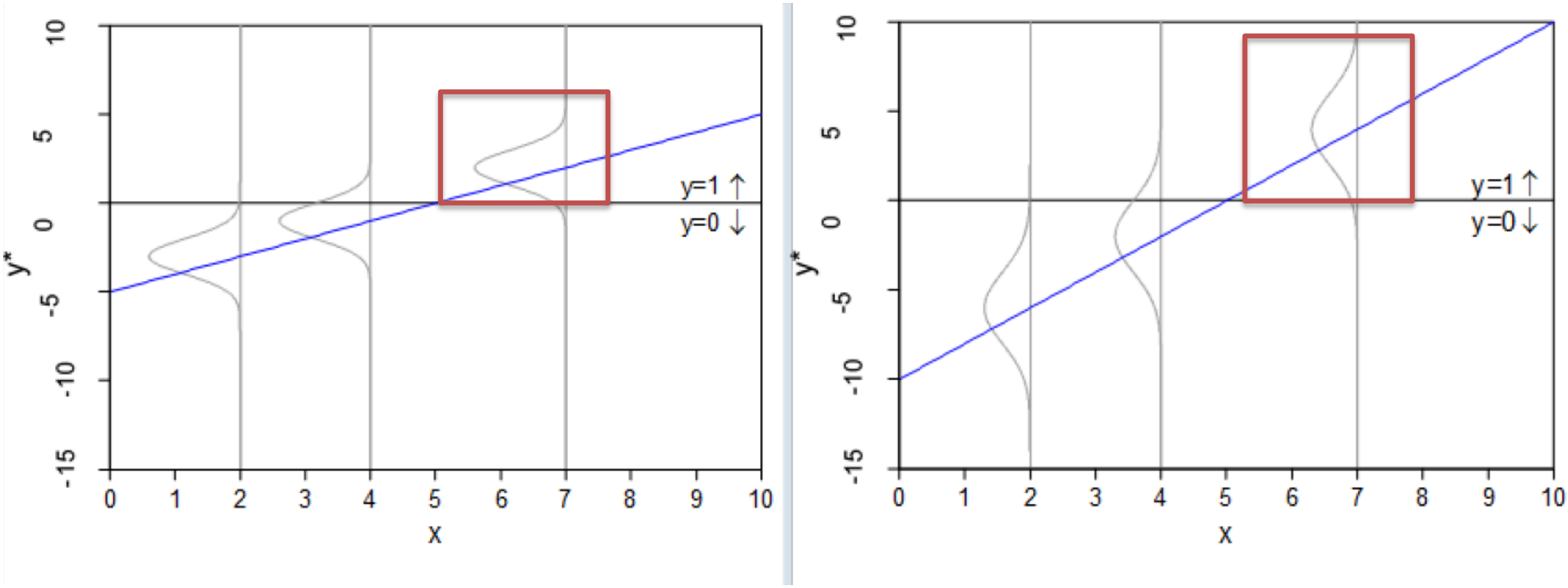


Panel A



Panel B

$$\boldsymbol{\beta}_L \approx \sqrt{Var(\varepsilon_L \mid \mathbf{x})}\boldsymbol{\beta}_P \approx \sqrt{\pi^2/3}\boldsymbol{\beta}_P \approx 1.81\boldsymbol{\beta}_P$$

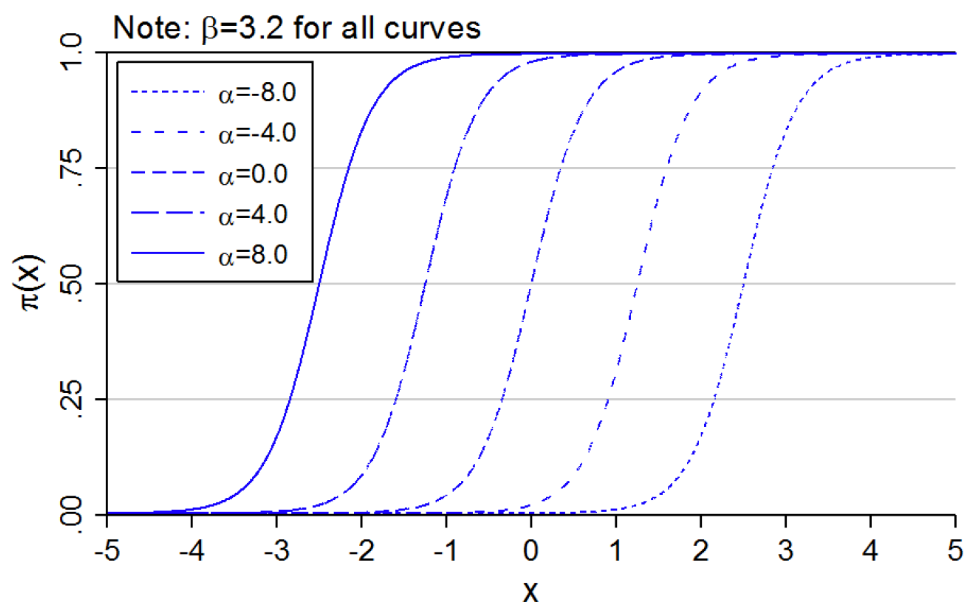


Panel A **Panel B**

$$\boldsymbol{\beta}_L \approx \sqrt{Var(\varepsilon_L \,|\, \mathbf{x})}\boldsymbol{\beta}_P \approx \sqrt{\pi^2/3}\boldsymbol{\beta}_P \approx 1.81\boldsymbol{\beta}_P$$

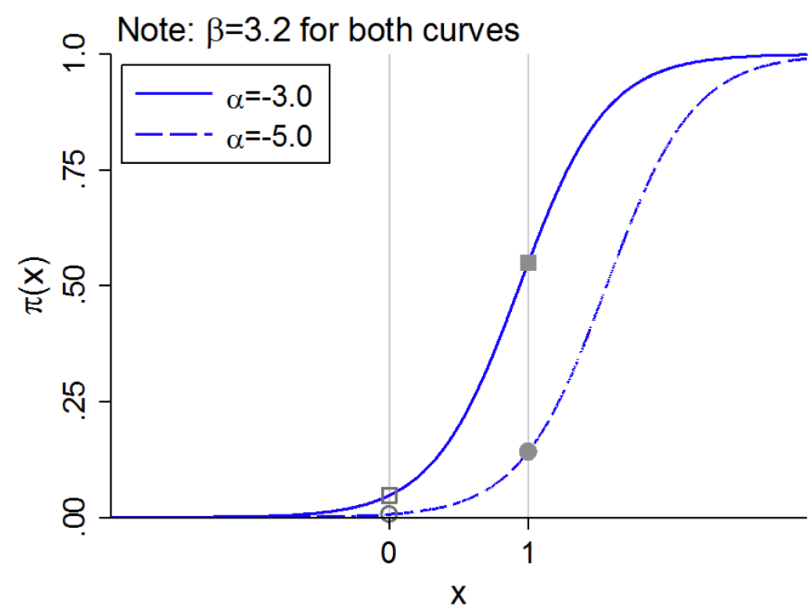
BUT! The identification assumption does not affect $\Pr\left(y=1 \mid \mathbf{x}\right)$ (or $\Delta \Pr\left(y=1 \mid \mathbf{x}\right) \mid \Delta\mathbf{x}$)

Functional Form: BRMs as "Inherently Interactive"



Intercept changes result in a series of 'marching curves'

These different curves result in different changes in Pr(Y=1) [marginal change], even when the β is the same.



Every Combination of X = New Probability Curve!

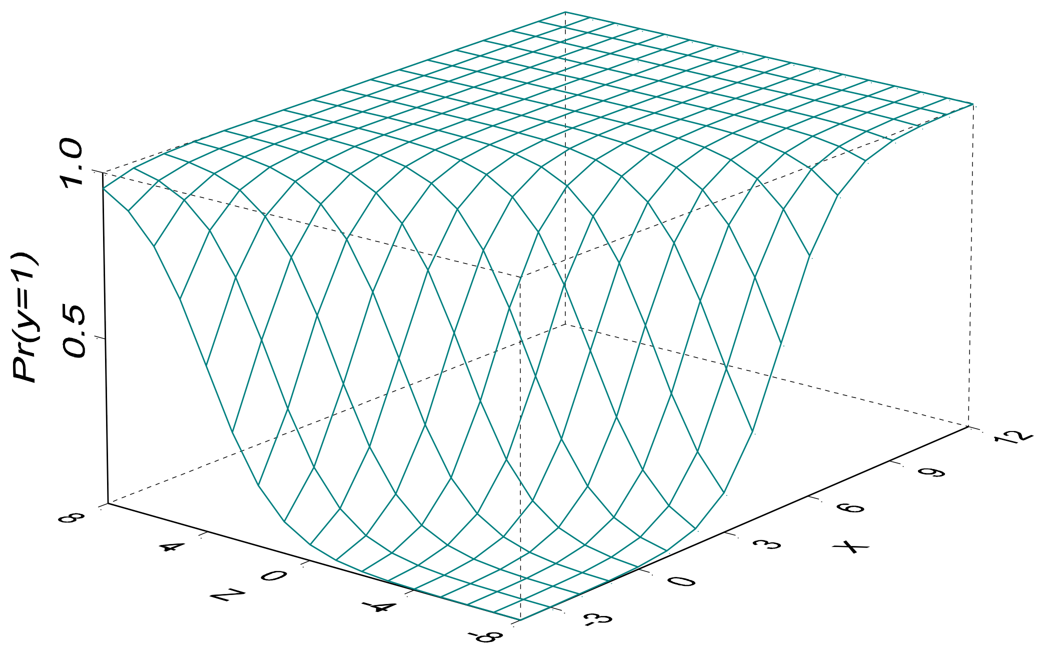
Consider the model:

$$\Pr(y=1 \mid x,z) = \Phi(1+1x+.75z)$$

Fix $z=-8$:

$$\Pr(y=1 \mid x,z=-8) = \Phi(1+1x+[.75*-8]) = \Phi(-5+1x)$$

Increase z to -4: $\Pr(y=1 \mid x,z=-4) = \Phi(1+1x+[.75*-4]) = \Phi(-2+1x)$



But extent of interactivity differs by location

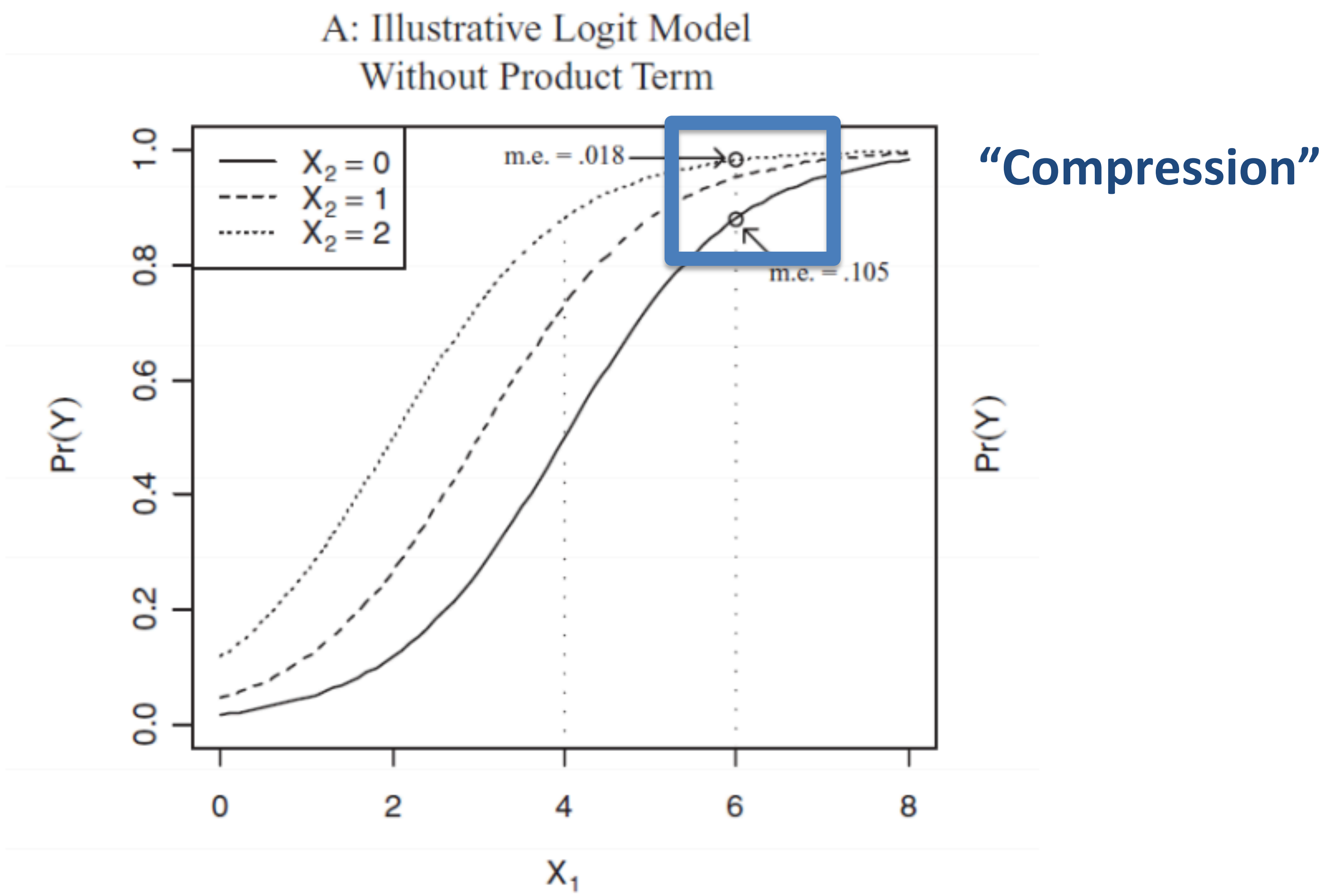


Figure from Berry, DeMeritt, & Esarey, AJPS 2010

But extent of interactivity differs by location

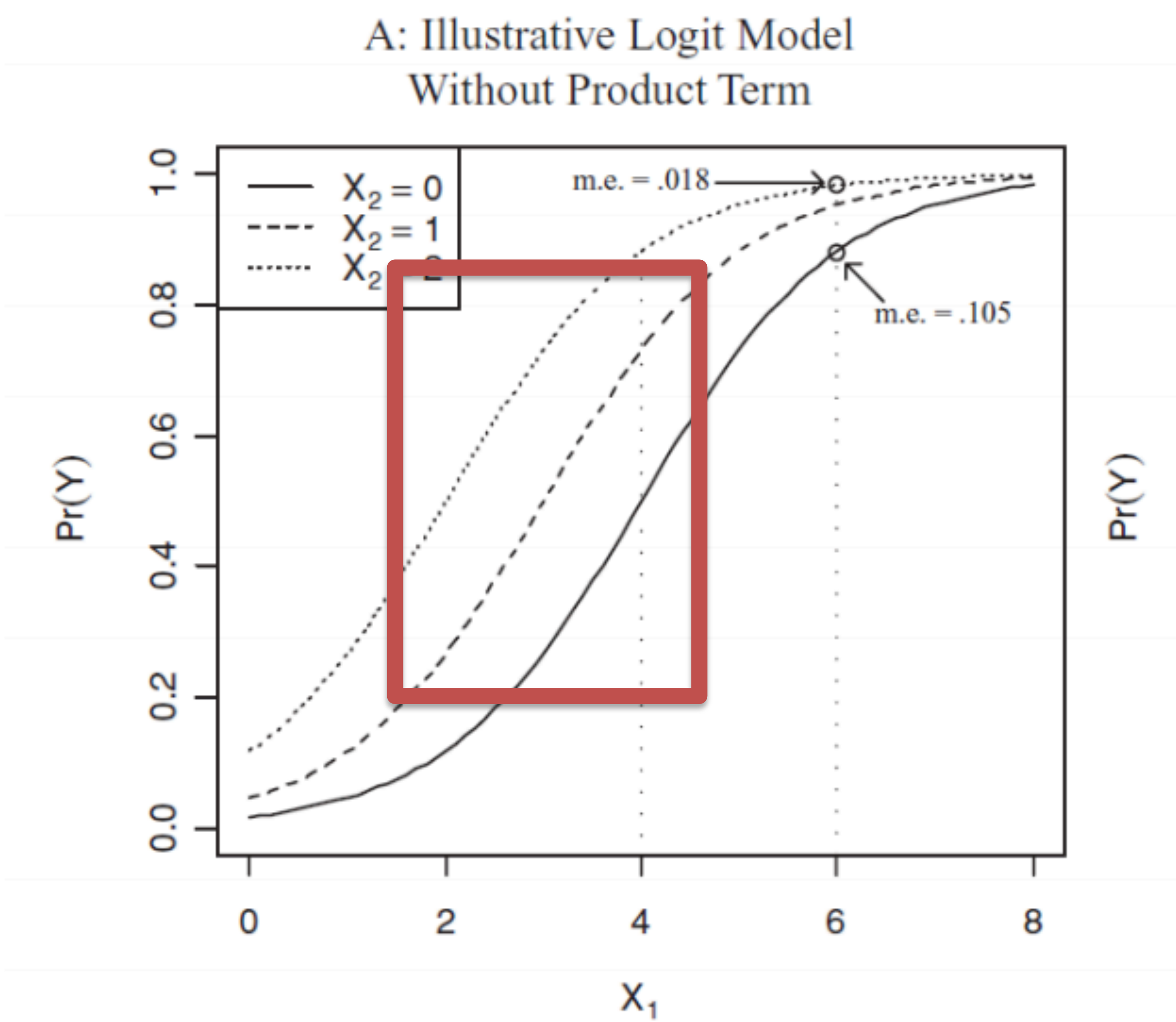


Figure from Berry, DeMeritt, & Esarey, AJPS 2010

Part 2: Interactions in BRMs

Part 2: Interactions in BRMs

Main points:

- Unlike in LRMs, **coefficients for product terms** in BRMs tell us <u>nothing</u> about the direction, magnitude, or significance of ΔPr(Y=1)
- Adding product terms in BRMs may make our model ***less interactive***.
  - *<u>And we may want that</u>*!

- Interaction terms must be ***understood*** via ***visualization.***
- Interaction effects must be ***tested*** via **test(s) of second differences** in the effect on Pr(Y=1).

Review: Product Terms in LRMs

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 * x_2)$$

Recent LRM Example

- Binary treatment indicator
- Binary moderator
- Outcome = Δ in # of patients receiving therapy/mo

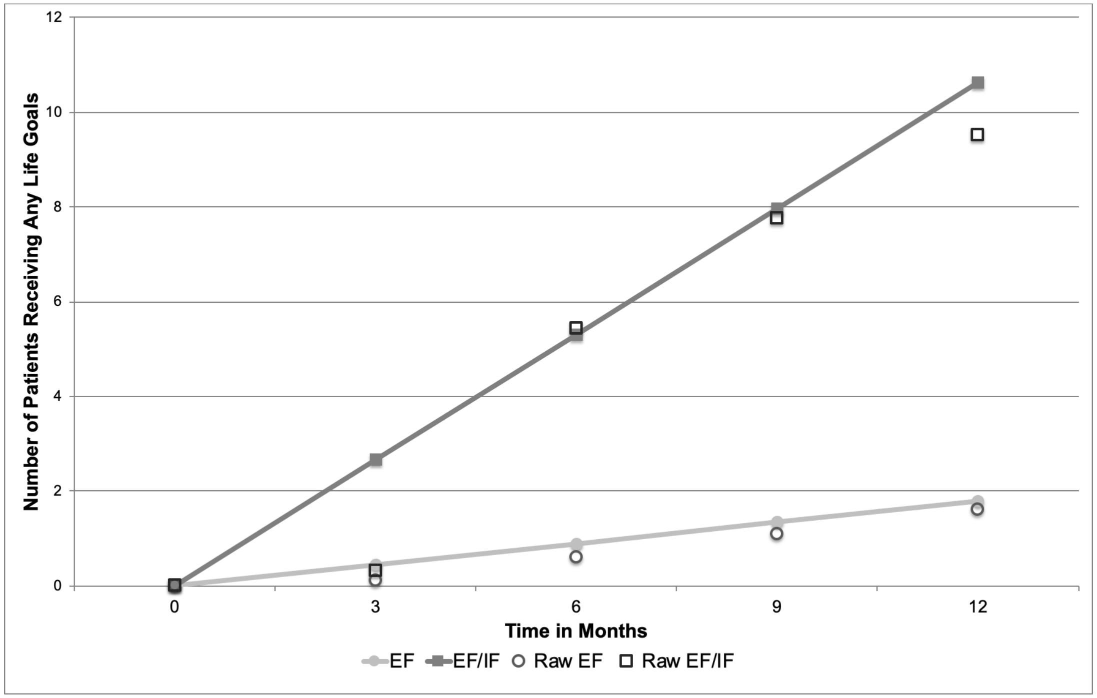
Coefficients of interest:

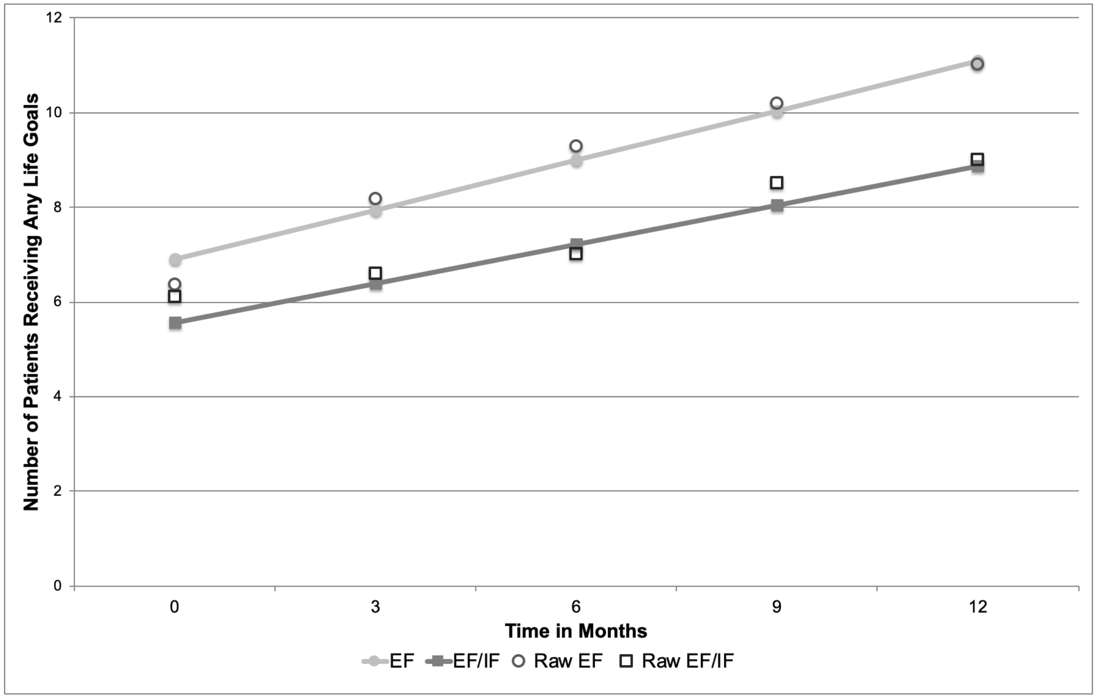
$\beta_{tx1} = 0.74$

$\beta_{(tx1 * group1)} = -0.81$

Recent LRM Example

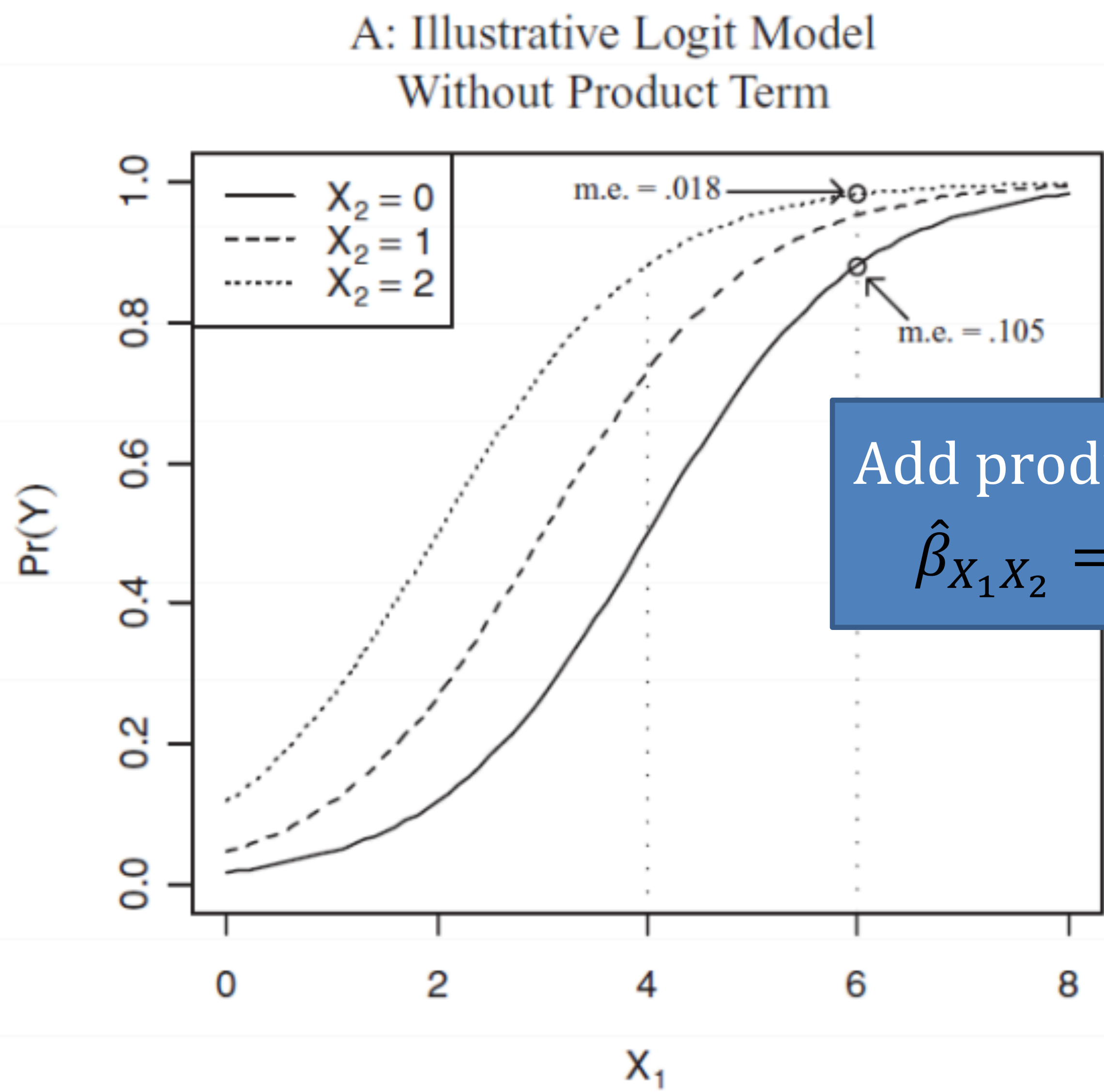
Group 1





Group 2

Can we use β to visualize an interaction in a BRM?

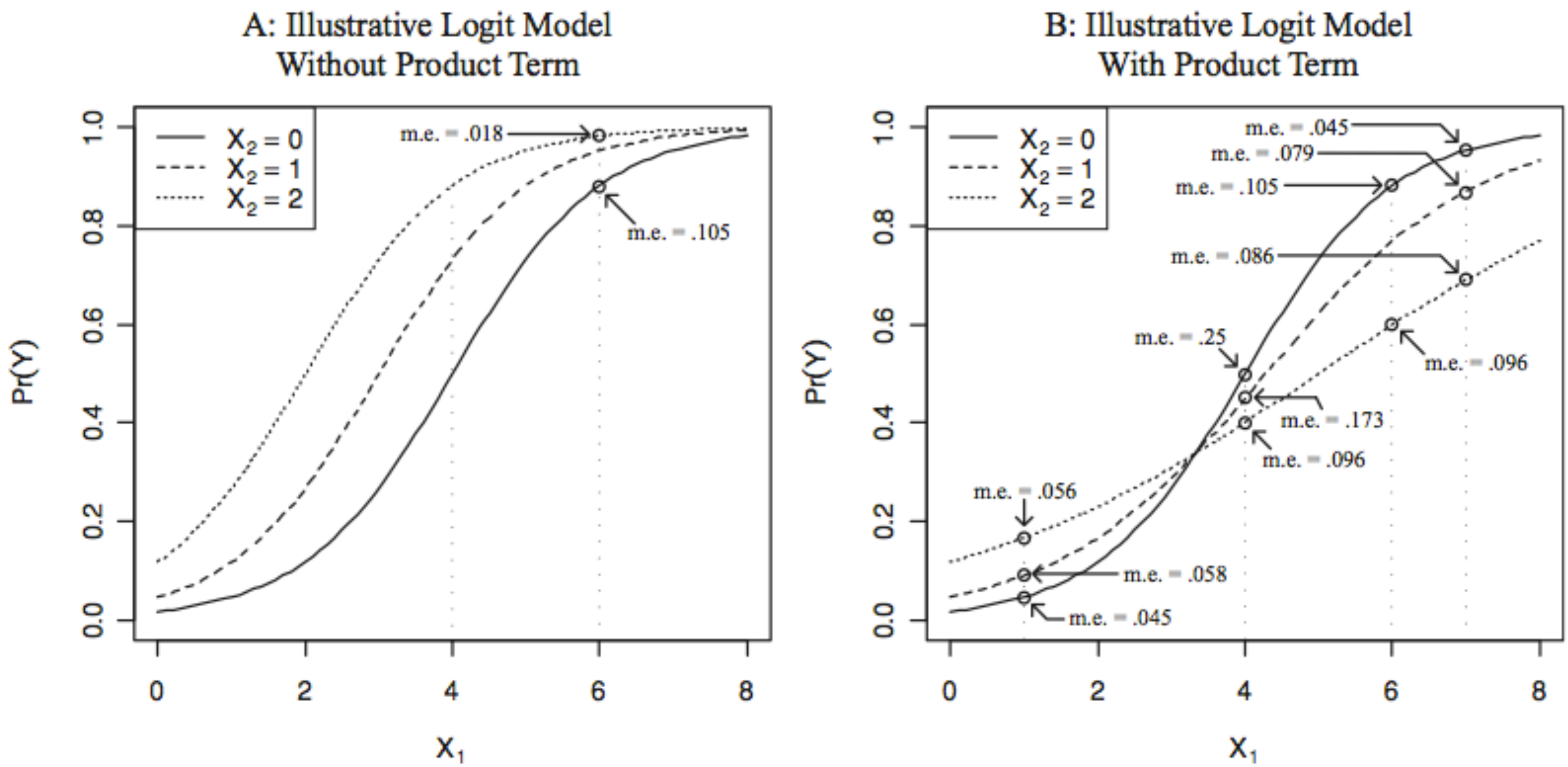


Add product term:

$$\hat{\beta}_{X_1 X_2} = -0.30$$

How'd you do??

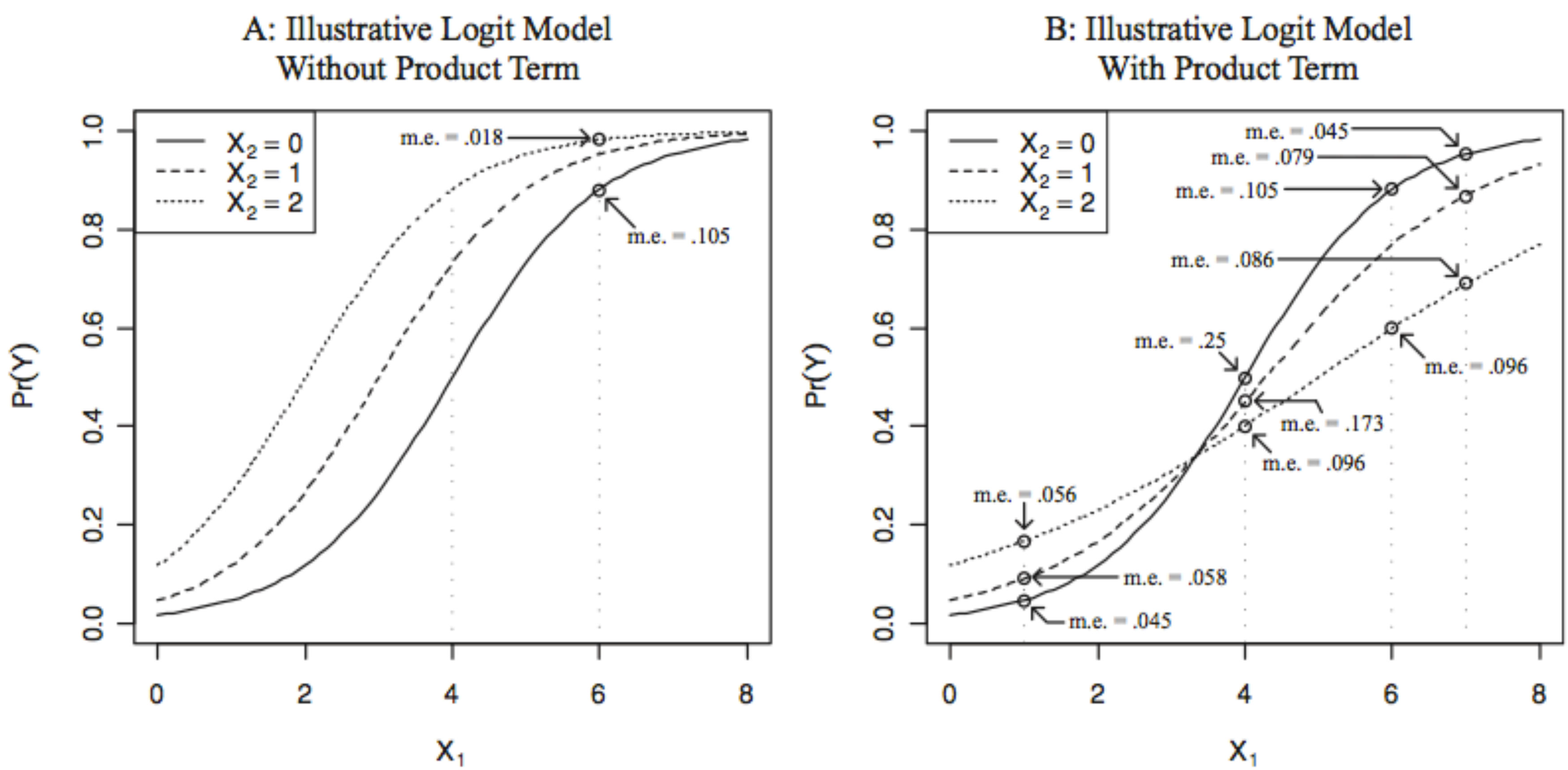
FIGURE 2 Logit Models Illustrating How Marginal Effects on Pr(Y) Vary with the Values of Independent Variables



Panel A depicts the model $\Pr(Y) = G(-4 + X_1 + X_2)$. Panel B depicts the model $\Pr(Y) = G(-4 + X_1 + X_2 - 0.30X_1X_2)$. In both cases, $G()$ is the logit link function. Arrows indicate the marginal effect (m.e.) of the curve, i.e., $\partial \Pr(Y)/\partial X_1$, at the indicated point.

Problem #1: Interactions are hard to visualize

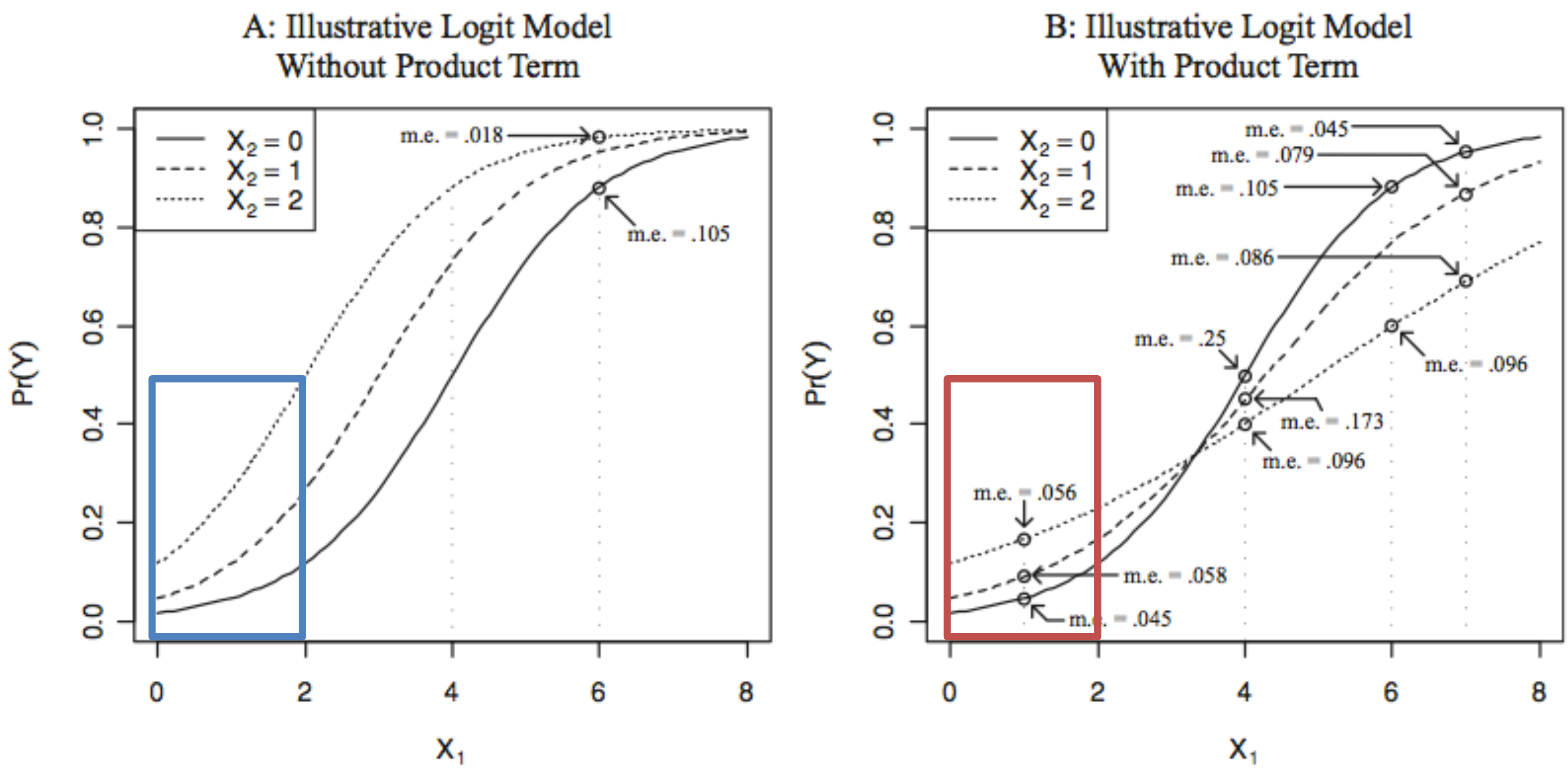
Figure 2 Logit Models Illustrating How Marginal Effects on Pr(*Y*) Vary with the Values of Independent Variables



Panel A depicts the model $\Pr(Y) = G(-4 + X_1 + X_2)$. Panel B depicts the model $\Pr(Y) = G(-4 + X_1 + X_2 - 0.30X_1X_2)$. In both cases, $G()$ is the logit link function. Arrows indicate the marginal effect (m.e.) of the curve, i.e., $\partial \Pr(Y)/\partial X_1$, at the indicated point.

Problem #2: Product term might make model more linear

FIGURE 2 Logit Models Illustrating How Marginal Effects on Pr(Y) Vary with the Values of Independent Variables



Panel A depicts the model $\Pr(Y) = G(-4 + X_1 + X_2)$. Panel B depicts the model $\Pr(Y) = G(-4 + X_1 + X_2 - 0.30X_1X_2)$. In both cases, $G()$ is the logit link function. Arrows indicate the marginal effect (m.e.) of the curve, i.e., $\partial \Pr(Y)/\partial X_1$, at the indicated point.

{Aside: Sometimes that's a good thing!}

When compression effects **don't exist** or are overestimated by the functional form of the BRM, adding product terms can help to diminish those effects--i.e., **the product term is necessary for the effect not to be interactive.**

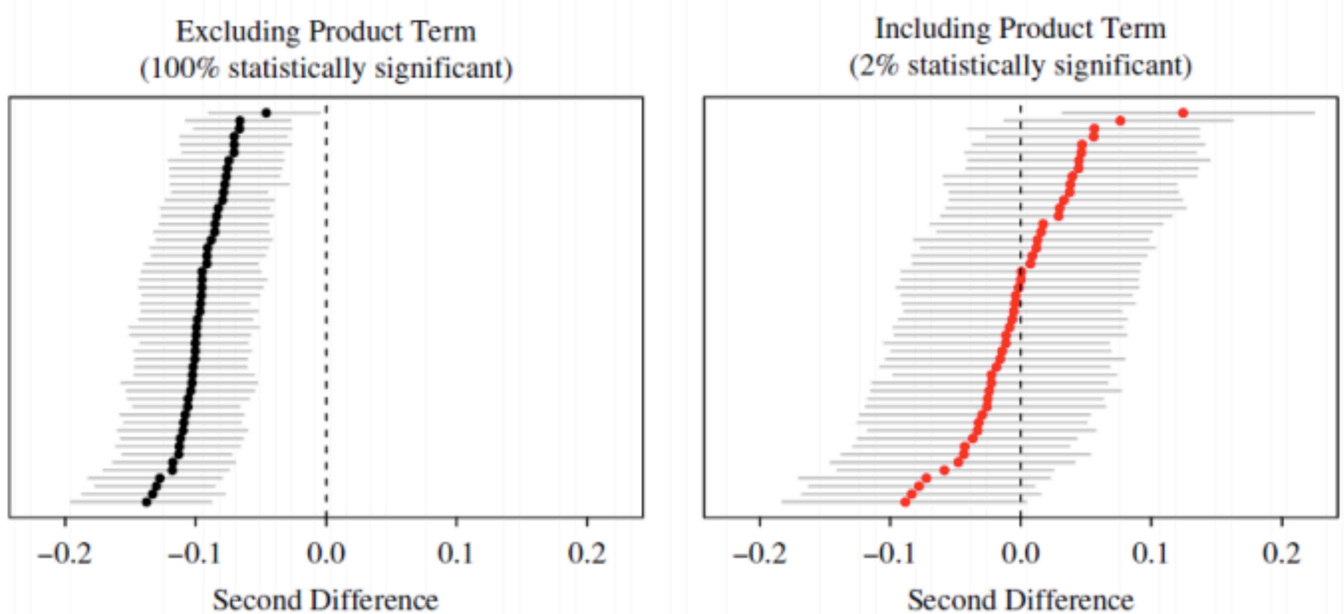
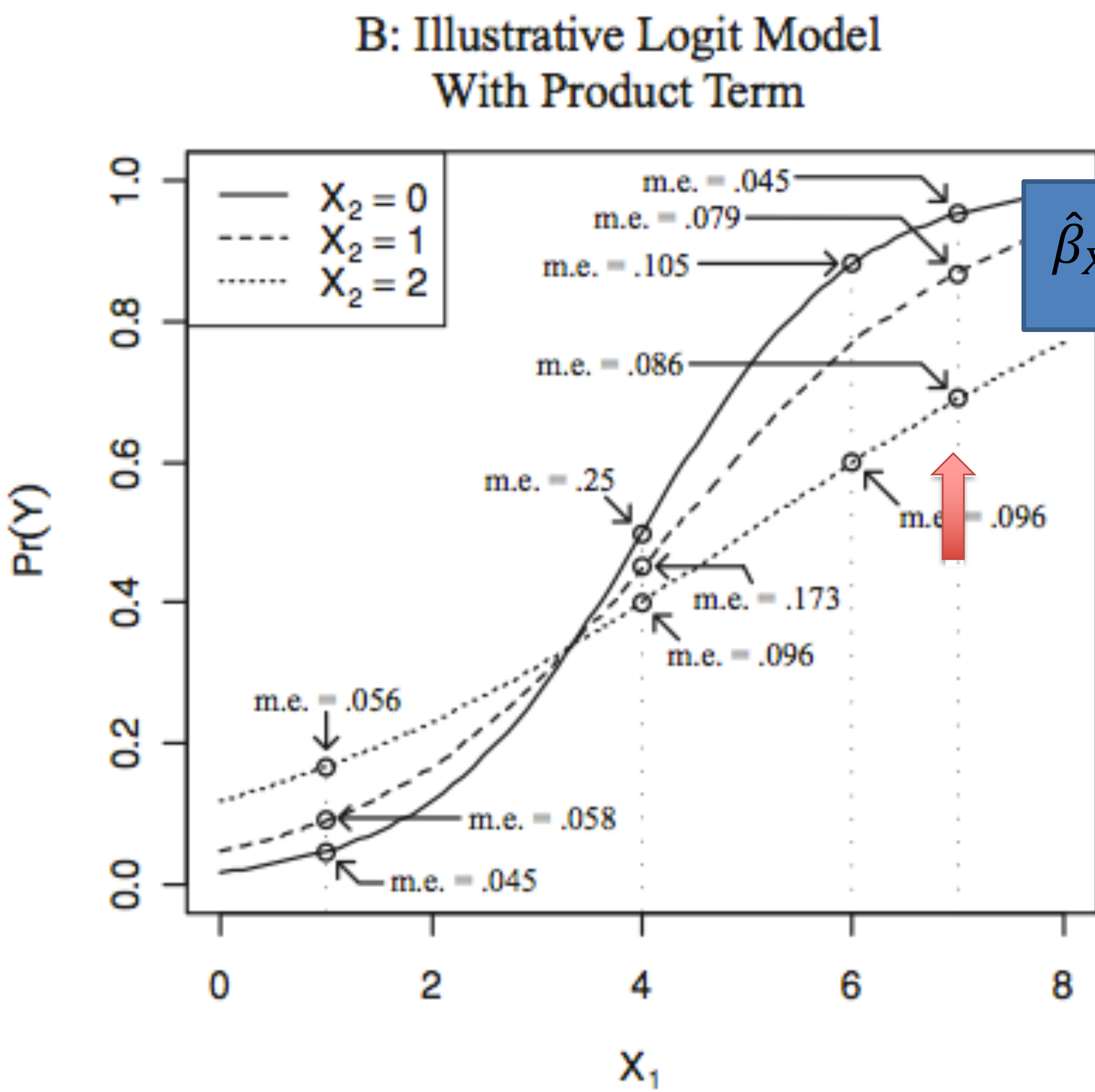


Fig. 2. Fifty estimated second differences and confidence intervals
Note: this figure shows 50 simulations of estimated second-differences and confidence intervals using models with and without a product term. Although second-difference is actually 0 (a non-interactive relationship), the model with no product term consistently finds interaction. Including a product term removes almost all of this bias.

Figure from Rainey, PSRM 2016

Problem #3: The sign on the coefficient may also be misleading



$\hat{\beta}_{X_1 X_2} = -0.30$

Problem #4a: A significant product term does not necessarily mean that there are significant second differences at any point in your data space.

Problem #4a: A significant product term does not necessarily mean that there are significant second differences at any point in your data space.

Problem #4b: A non-significant product term does not necessarily mean that there are *not* significant second differences at any point in your data space.

Problem #4a: A significant product term does not necessarily mean that there are significant second differences at any point in your data space.

Problem #4b: A non-significant product term does not necessarily mean that there are *not* significant second differences at any point in your data space.

You must test for the interaction effect in terms of the second difference of ΔPr(Y=1)

i.e., are the marginal effects on (ΔPr(Y=1)) significantly different?

Most People Get This Wrong

"A review of 13 economics journals listed on JSTOR found 72 articles published between 1980 and 1999 that used interaction terms in nonlinear models. **None of the studies interpreted the coefficient on the interaction term correctly**." (Ai and Norton 2003 p.123).

"… I examined the 53 articles published in the *American Sociological Review* between 2004 and 2016 that examined interaction effects in nonlinear models. **50 of the 53 referred only to the coefficient on the product term to determine the significance of the interaction effect—an improper test of interaction in terms of the predicted probabilities.**" (Mize 2019 p. 82)

A Simple Example

```
. usecda cda_tenure01
. codebook tenure female articles prestige , compact

Variable     Obs Unique      Mean   Min   Max  Label
-------------------------------------------------------------------------------
tenure      2797      2  .1229889     0     1  Is tenured?
female      2797      2  .3775474     0     1  Scientist is female?
articles    2797     48  7.050411     0    73  Total number of articles.
prestige    2797     98  2.646591   .65   4.8  Prestige of department.
-------------------------------------------------------------------------------


. sum tenure female articles prestige

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      tenure |       2797    .1229889    .3284832          0          1
      female |       2797    .3775474    .4848602          0          1
    articles |       2797    7.050411    6.575682          0         73
    prestige |       2797    2.646591    .7769724        .65        4.8
```

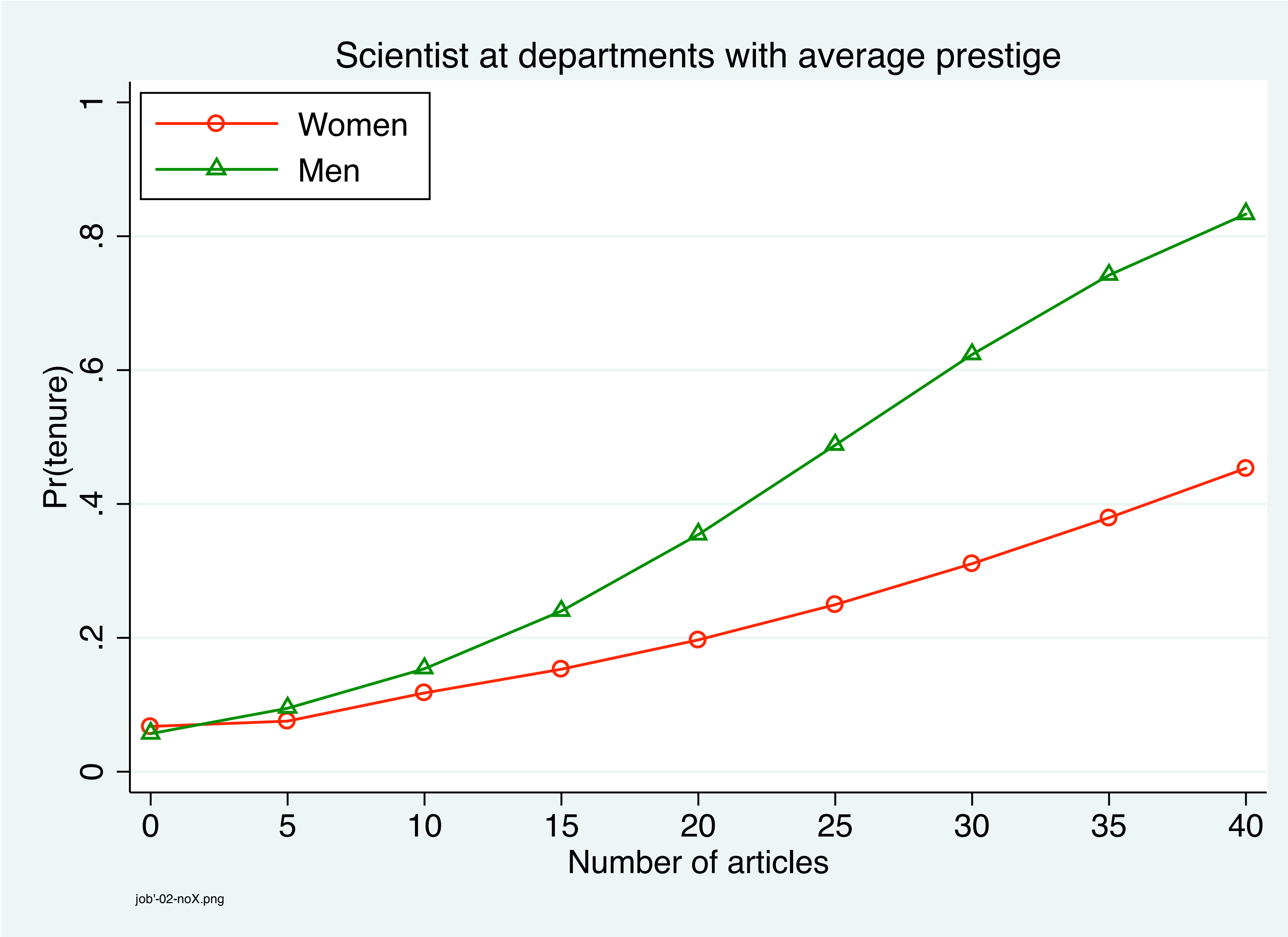
Step 0: Run the Model

```
. logit tenure i.female##c.articles prestige , nolog

Logistic regression                               Number of obs   =       2945
                                                  LR chi2(4)      =     144.19
                                                  Prob > chi2     =     0.0000
Log likelihood = -1027.4788                       Pseudo R2       =     0.0656

------------------------------------------------------------------------------
      tenure |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |
     Female  |   .1785884   .1789557     1.00   0.318    -.1721584    .5293352
    articles |   .1038231   .0100759    10.30   0.000     .0840748    .1235714
             |
      female#|
  c.articles |
     Female  |  -.0496277   .0141527    -3.51   0.000    -.0773666   -.0218889
             |
    prestige |  -.3610535   .0792061    -4.56   0.000    -.5162946   -.2058125
       _cons |  -1.809395   .2203708    -8.21   0.000    -2.241314   -1.377476
------------------------------------------------------------------------------
```

Step 1: Plot!



Scientist at departments with average prestige
Women
Men
Pr(tenure)
0
.2
.4
.6
.8
1
0
5
10
15
20
25
30
35
40
Number of articles
job'-02-noX.png

Step 2: Compute Second Differences

(a) Generate predicted probabilities

```
. mtable, at(female=(0 1) articles=(10 25)) post

Expression: Pr(tenure), predict()

          |    female   articles      Pr(y)
----------+-----------------------------
        1 |         0         10      0.152
        2 |         0         25      0.459
        3 |         1         10      0.115
        4 |         1         25      0.227
```

(b) Generate first differences

```
. *Effect of more papers for men (first difference)
. mlincom 2-1, stat(est se p)

             |    lincom          se     pvalue
-------------+------------------------------
           1 |     0.307       0.039      0.000


. *Effect of more papers for women (first difference)
. mlincom 4-3, stat(est se p)

             |    lincom          se     pvalue
-------------+------------------------------
           1 |     0.112       0.029      0.000
```

(c) Generate second difference

```
. *Difference between effect for men & women (second difference)
. mlincom (4-3) - (2-1), stat(est se p)

             |    lincom         se     pvalue
-------------+--------------------------------
           1 |    -0.196      0.048      0.000
```

**How do we interpret this?

What about other types of interactions?

Nominal x nominal: A table might suffice?

Table 6. Predictions and Tests of Second Differences from Random Intercept Multilevel Models Predicting Challenging Behavior: Study 2 ($N_{level\ 1}$ = 2,490; $N_{level\ 2}$ = 83).

	Pr(Challenging)	Test of First Difference	Test of Second Difference
Low cost			
Not aggressive	.85 (.04)	.85 – .70 = .15***	.15 – .05 = .10*
Aggressive	.70 (.06)		
Medium cost			
Not aggressive	.60 (.06)	.60 – .55 = .05	
Aggressive	.55 (.06)		

Note: Probabilities (Pr) of challenging are presented, with standard errors in parentheses. A Level 2 random intercept for the participant is included to account for clustering. All models include controls for reputation of partner and time block of the experiment.
$^{*}p < .05$. $^{***}p < .001$.

Table from Benard, Berg & Mize, SPQ, 2017

Unless you have lots of categories; then may want to plot…

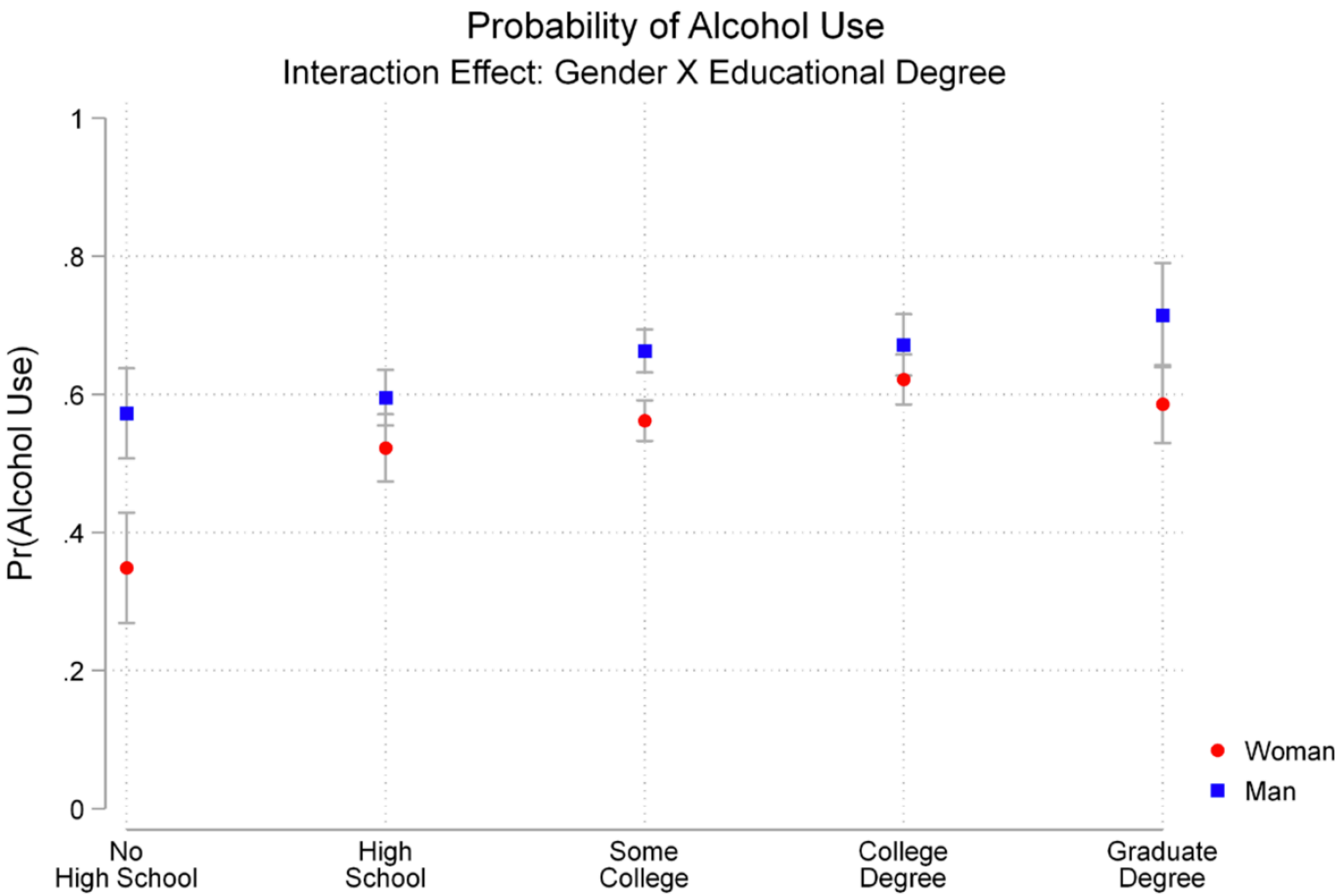


Table 1: Probability of Alcohol Use by Gender and Education

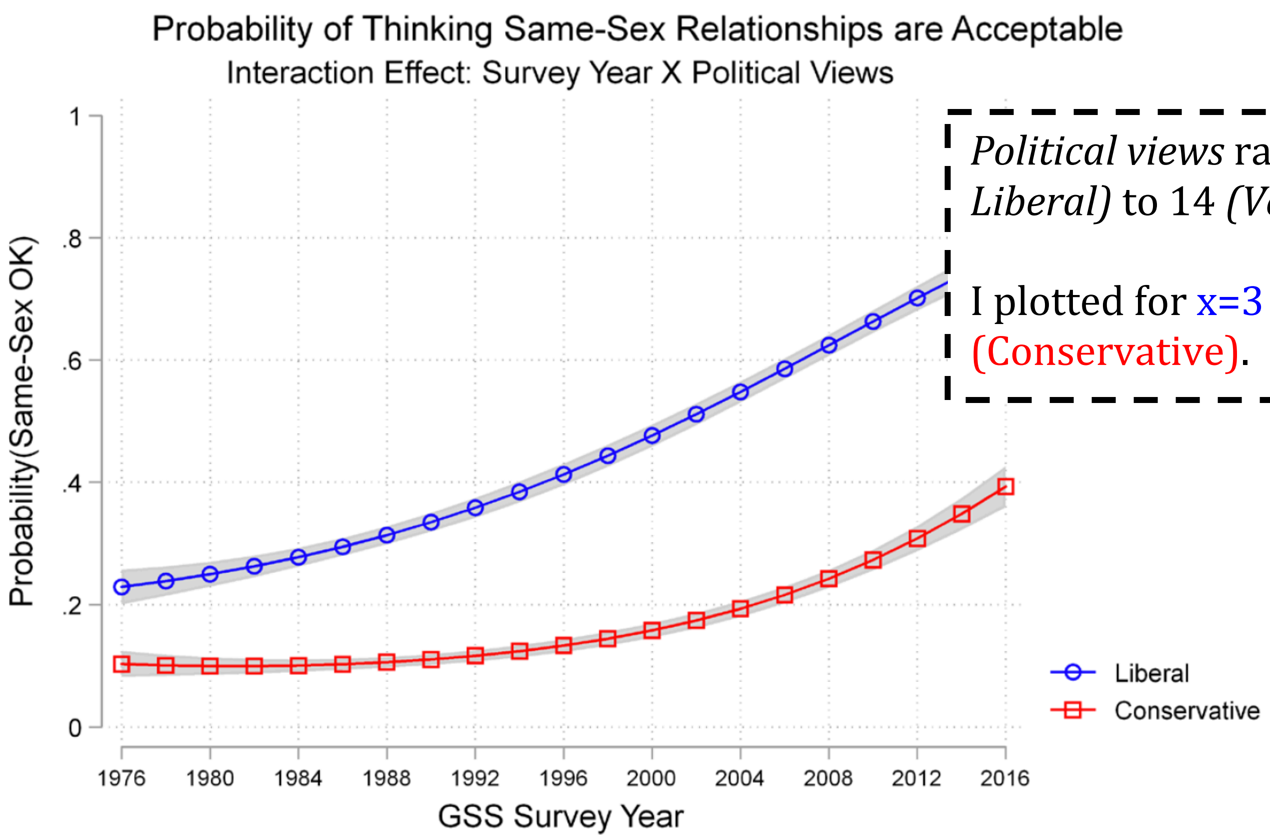
	Women	Men	Gender Gap	Contrasts
a No High School	0.349	0.572	0.223***	*b, c, d*
b High School	0.522	0.595	0.073*	*a*
c Some College	0.562	0.663	0.101***	*a*
d College Degree	0.622	0.672	0.050	*a*
e Graduate Degree	0.586	0.715	0.129**	

Notes: (1) All education categories refer to highest degree completed. (2) Statistic for "gender gap" is the difference in the effect of education between men and women. (3) Contrasts column reports which gender gaps are significantly different (second differences).

**How would you interpret this?

Continuous x continuous: Always need figures, but which values to present?

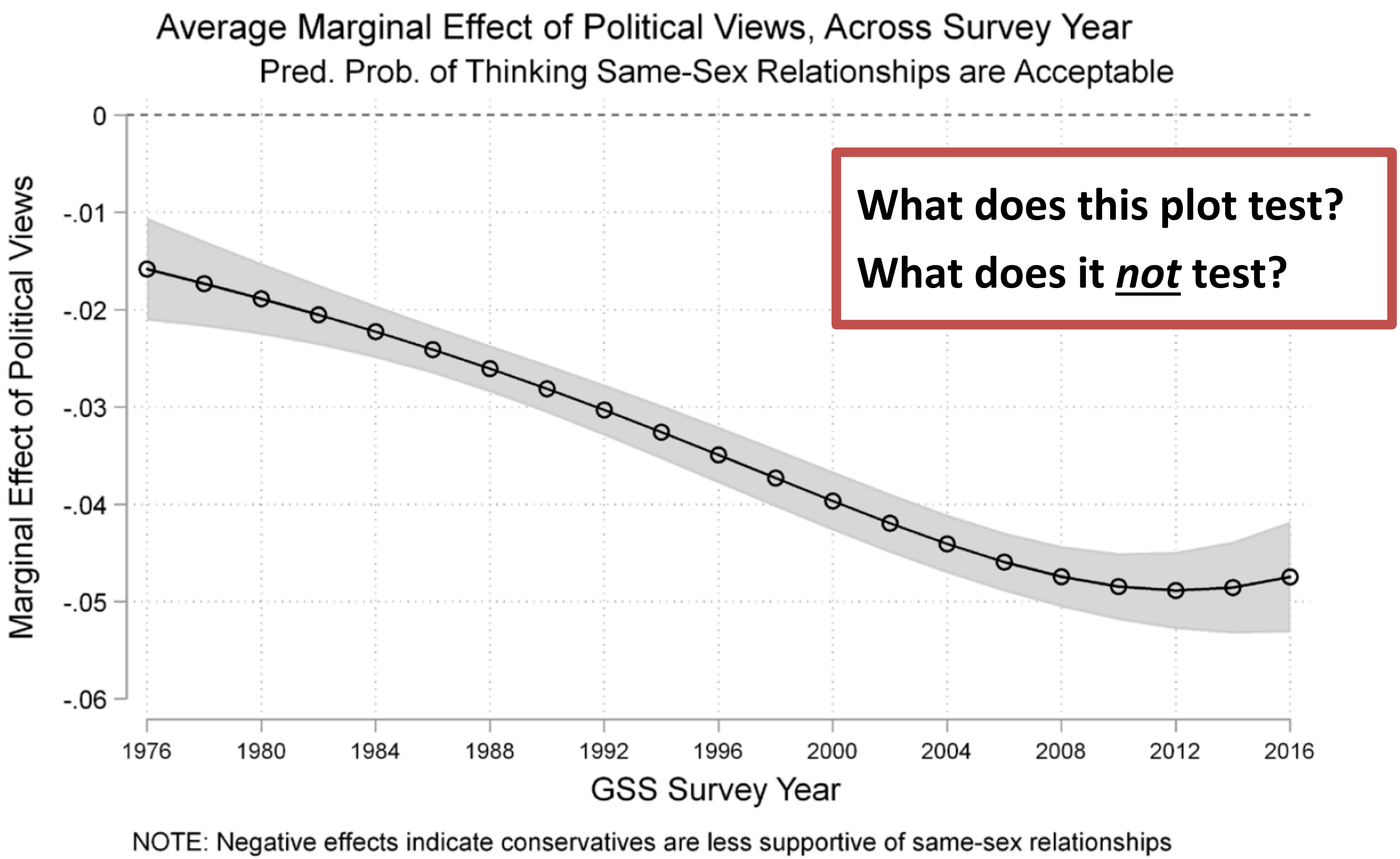
i. **Ideal types: Plot certain values for 1st continuous variable across range of 2nd**



Political views range from 0 *(Very Liberal)* to 14 *(Very Conservative)*.

I plotted for x=3 (Liberal) & x=11 (Conservative).

ii. Plot the average marginal effect of one variable across the range of another



Compute first differences

```
. mtable, dydx(polscale) at(year=(1980 2016)) post

| yearcat    d Pr(y)
--------+-------------------
1    |    1980     -0.019
2    |    2016     -0.047
```

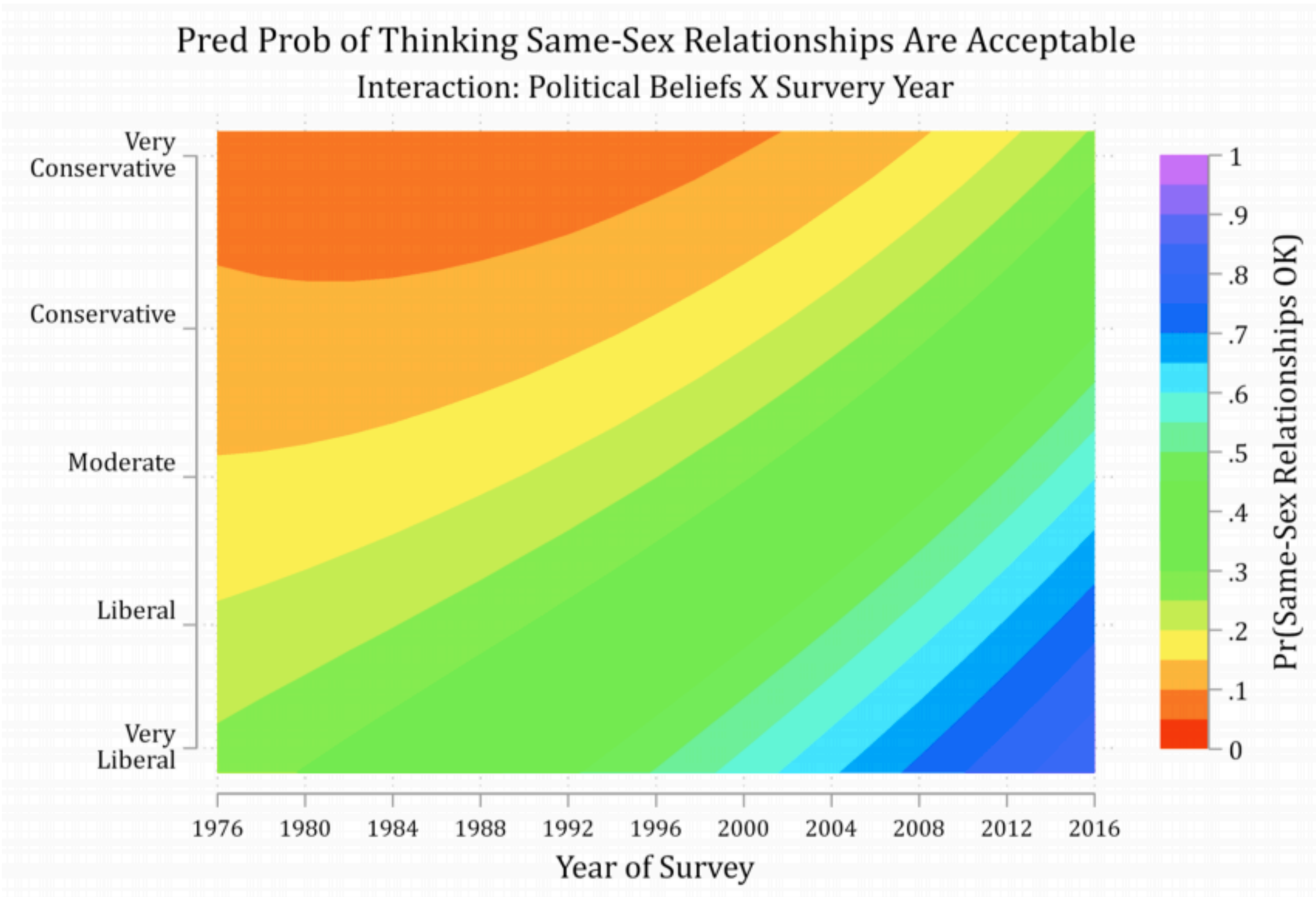
Compute second difference

```
. mlincom 2 - 1, stat(est se p)

        | lincom         se       pvalue
--------+--------------------------------------
   1    | -0.029      0.003      0.000
```

**How do we interpret?

iii. Contour plots can help with understanding change in both dimensions



Summary of nonlinear product term recommendations

1. **Include a product term** if you are interested in whether the effect of one variable is contingent on the level of a second variable.
   - ***OR*** if you want to disrupt the default compression interactions native to the BRM functional form.

2. **Ignore the coefficient**. Seriously. SERIOUSLY!

3. **Plot the predictions**. How is the product term impacting your model?

4. **Use marginal effects** (changes in Pr(Y=1) to determine the size & significance of effects of interest.

5. **Test for second differences** (equivalence of two marginal effects) to determine whether your interaction terms is significant.

For more, see Mize 2019.

Part 3: Cross-model comparisons in BRMs

Part 3: Cross-model comparisons in BRM

Main points:

- Setting the variance of the errors for the BRM means that β **coefficients are not individually identified**
    - Differences in unobserved heterogeneity are also reflected in βs.
- Since that can vary from model to model (or group to group), the scale represented by a particular β is *specific to that model* & **βs cannot be compared directly across models.**
    - Two applications:
        - Evaluating change in the effect of a variable when other variables are added (*cf.* mediation)
        - Comparing the effect of the same variable across two groups

- **Valid cross-model comparisons must either:**
    - Rescale coefficients
    - Compare marginal effects

Example 1: Change in effects of a variable

```
. usecda cda_standardized

. qui logit ybinary x1
. estimates store m1

. qui logit ybinary x2
. estimates store m2

. logit ybinary x1 x2
. estimates store m3
```

```
. esttab m1 m2 m3, bic

------------------------------------------------------------
                      (1)             (2)             (3)   
                  ybinary         ybinary         ybinary   
------------------------------------------------------------
ybinary                                                     
x1                  0.739***                        1.789***
                  (10.13)                          (9.81)   

x2                                  0.489***        1.173***
                                  (10.13)          (9.71)   

_cons             -0.0530         -0.0724          -0.214   
                  (-0.50)         (-0.68)         (-1.32)   
------------------------------------------------------------
N                     500             500             500   
BIC                 543.5           544.9           268.1   
------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

Question (for you)

If we found this in the context of the LRM, what would we assume about x1 and x2?

```
. corr ybinary x1 x2, means
(obs=500)

             |  ybinary       x1       x2
-------------+---------------------------
     ybinary |   1.0000
          x1 |   0.5248   1.0000
          x2 |   0.5225   0.0000   1.0000
```

BUT…

x1 and x2 are completely uncorrelated!! What's going on??

{REMINDER}

In order for the model to be identified, we have to make an assumption about the variance of the errors.

This sets our metric.

Estimated SD of y*

$$SD(\widehat{y^*}) = \sqrt{\mathrm{var}(\mathbf{x}\hat{\boldsymbol{\beta}}) + \mathrm{var}(\varepsilon)}$$

Let's look at those model results one more time

	Model 1	Model 2	Model 3
x1	0.739		1.7892
x2		0.489	1.1731

Let's look at those model results one more time

	Model 1	Model 2	Model 3
x1	0.739		1.7892
x2		0.489	1.1731
SD(y*)	**2.3396**	**2.3322**	**5.3368**

…& then one more time

	Model 1 *(y-standardized)*	Model 2 *(y-standardized)*	Model 3 *(y-standardized)*
x1	0.316		**0.320**
x2		0.210	**0.212**
SD(y*)	2.3396	2.3322	5.3368

The "re-scaling problem"

β **coefficients are not individually identified**; rather they are identified only "<u>up to a scale factor</u>."

$$b = \frac{\beta}{s} \qquad\qquad s = \frac{\sigma_\varepsilon}{\omega}$$

The **<u>size of the coefficient (b) reflects three things</u>**:

- The *explained variation* for that variable (β)
- The *true residual variation* of the underlying model (σ_ε)
  - This changes as we change the variables included in the model; since we fix residual error variance, improvements in model fit must result in an improved ratio of explained to residual variance → increase in total variance!
- The assumption we make about the *variance of the errors (ω)*

***Breen, Karlson & Holm 2018 provide a great discussion of this issue.*

Example 2: Comparing across groups

Is the effect of a variable the same across two groups?

- E.g., Do women faculty get the same returns to publishing that men do?

LRM

Comparing groups

Estimate separate regression models for two (or more groups) and then compare coefficients across groups

$$y^m = \alpha^m + \beta^m_{articles} x_{articles}$$

$$y^w = \alpha^w + \beta^w_{articles} x_{articles}$$

Do men and women get the same returns for publishing?

$$H_0 : \beta^m_{articles} = \beta^w_{articles}$$

Testing using a Chow type test

$$z = \frac{\hat{\beta}^m_{articles} - \hat{\beta}^w_{articles}}{\sqrt{Var(\hat{\beta}^m_{articles}) + Var(\hat{\beta}^w_{articles})}}$$

BRM
Comparing groups

The above test assumes that the *scale factor* for the coefficients is the same.

i.e., it confounds:

- Group differences in the effect of a predictor
- Group differences in the true residual variation *(unobserved heterogeneity;* σ_ε*)*

Structural model

For men: $$y^* = \alpha^m + \beta^m_{articles} x_{articles} + \varepsilon^m$$

For women: $$y^* = \alpha^w + \beta^w_{articles} x_{articles} + \varepsilon^w$$

Estimated model (for probit)

For men: $$\frac{y^*}{\sigma^m} = \frac{\alpha^m}{\sigma^m} + \frac{\beta^m_{articles}}{\sigma^m} x_{articles} + \frac{\varepsilon^m}{\sigma^m}$$

For women: $$\frac{y^*}{\sigma^w} = \frac{\alpha^w}{\sigma^w} + \frac{\beta^w_{articles}}{\sigma^w} x_{articles} + \frac{\varepsilon^w}{\sigma^w}$$

Tests of group differences

We want to test:

$$H_0 : \beta^m_{articles} = \beta^w_{articles}$$

Can only test:

$$H_0 : \frac{\beta^m_{articles}}{\sigma^m} = \frac{\beta^w_{articles}}{\sigma^w}$$

These are only equivalent if:

$$\sigma^m = \sigma^w$$

But

If in fact:

$$\sigma^m \neq \sigma^w$$

Then:

"… the standard tests for cross-group differences in the β coefficients **tell us nothing** about differences in the [true] coefficients" (Allison, 1999).

"[I]n the presence of ***even fairly small differences*** in residual variation, naive comparisons of coefficients [across groups] can ***indicate differences where none exist, hide differences that do exist, and even show differences in the opposite direction*** of what actually exists" (Hoetker, 2004)

In a nutshell

We could have this problem:

	Men	Women
True model	$y^* = \alpha + \beta_{articles} x_{articles} + \varepsilon$	$y^* = \alpha + 2\beta_{articles} x_{articles} + 2\varepsilon$
Estimated model	$y^* = \alpha \ \ + \beta_{articles} x_{articles} + \varepsilon$	$y^* = \alpha \ \ + \beta_{articles} x_{articles} + \varepsilon$

Hide differences that do exist!

Or, alternatively, this problem:

	Men	Women
True model	$y^* = \alpha \ \ + \beta_{articles} x_{articles} + \varepsilon$	$y^* = \alpha \ \ + \beta_{articles} x_{articles} + 2\varepsilon$
Estimated model	$y^* = \alpha \ \ + \beta_{articles} x_{articles} + \varepsilon$	$y^* = \alpha \ \ + .5\beta_{articles} x_{articles} + \varepsilon$

Indicate differences where none exist!

So what do we do about it?

Solution 1: Rescale the coefficients

Karlson, Holm & Breen (*SMR 2012; also see Breen et al., SMR 2018)* propose a solution that allows for rescaling coefficients so they can be directly compared.

Implementable via `khb` command in Stata; similar commands in R.

Limitations:

- Largely limited to looking at coefficients in original metric (e.g., log odds)
- Will examine average marginal effects *(average partial effect)*, but does not provide a statistical test

Solution 2: Examine predictions/marginal effects

Mize, Doan & Long (SM, 2019) propose a general framework that allows for cross-model comparisons by:

- Combining multiple model estimates using **S**eemingly **U**nrelated **Est**imation (SUEST)
- Testing cross-model differences in predictions/marginal effects using Wald test

$$H_0: ME_1 = ME_2$$

$$z = \frac{\widehat{ME}_1 - \widehat{ME}_2}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2 - 2\hat{\sigma}_{1,2}}}$$

**Implementable in Stata via `mecompare`; see www.trentonmize.com for more details & examples

Example application

Table 3: Effect of having a college education on happiness using average marginal effects (AMEs) from binary logit model (N=9,216).

		Model 1: College only	Model 2: + Controls	Model 3: + Wages	Model 4: + Prestige
Panel A: Average Marginal Effects					
College Degree		0.072***	0.060***	0.036**	0.019
		(0.010)	(0.011)	(0.011)	(0.012)
Panel B: Cross Model Differences					
$AME_{Model\,1} - AME_{Model\,2}$	$0.072 - 0.060 =$		0.012**		
			(0.004)		
$AME_{Model\,2} - AME_{Model\,3}$	$0.060 - 0.036 =$		0.024***		
			(0.004)		
$AME_{Model\,4} - AME_{Model\,3}$	$0.036 - 0.019 =$		0.017***		
			(0.004)		

Limitations

- Requires decisions about which marginal effects to look at.

Wrapping up

Wrapping Up

In BRMs, that β coefficient isn't what you think it is!

- The identification assumptions + the functional form prevent us from using β coefficients to understand &/or appropriately test interactions or cross-model comparisons.
- Applying methods learned from the LRM context to these coefficients can lead to incorrect conclusions.

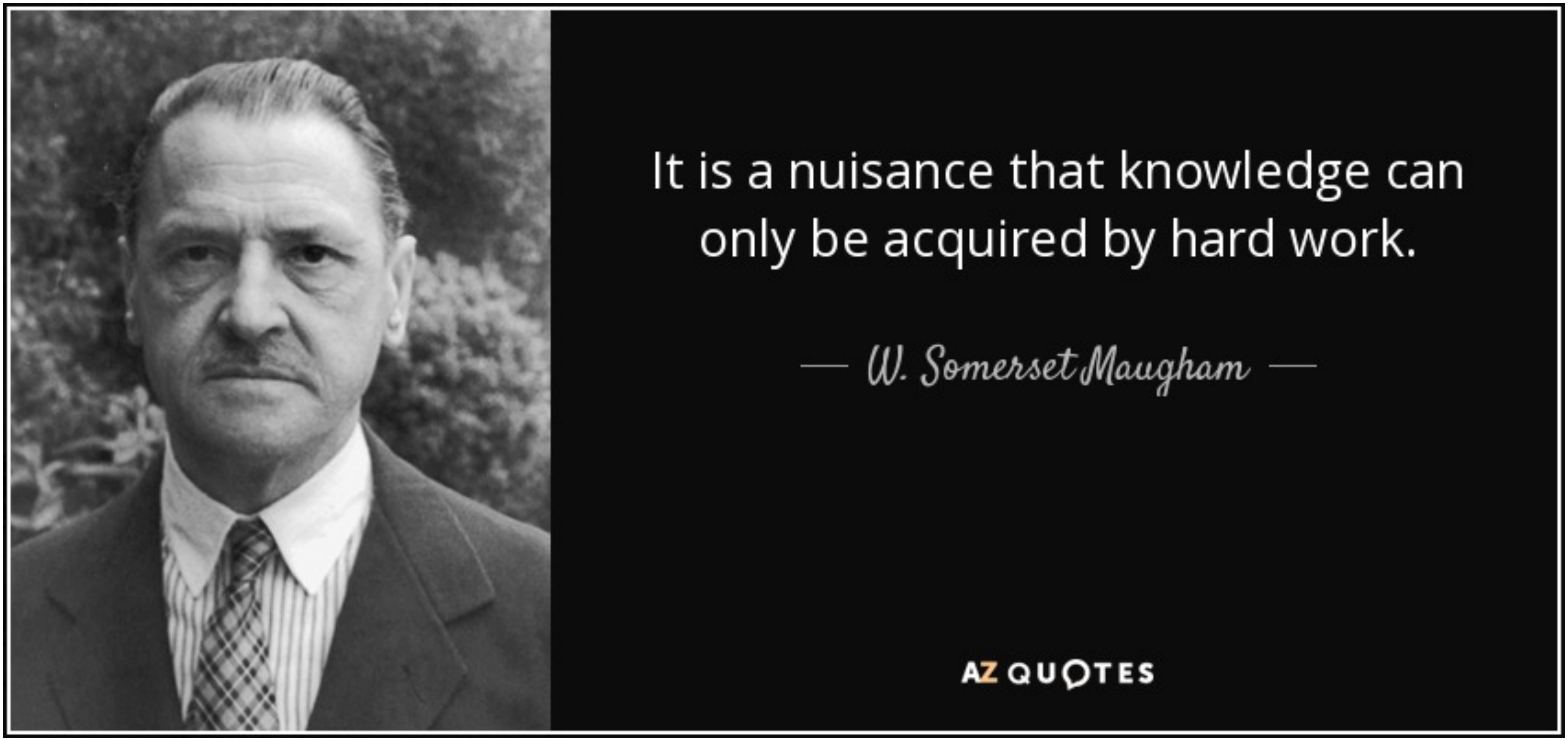
Safer bet: think of β as a nuisance parameter

- First step on the path, but requires more work to understand
  - **Interactions:** Visualize, test second-order differences
  - **Cross-model comparisons:** Rescale, test marginal effects
- Since marginal effects are not impacted by identification assumptions, they are always a safer bet

Don’t want to do the work? Reconsider that LRM (LPM)

- In RCTs, some recommendations to use LPM to determine Average Treatment Effect (Greene, 2011; Gomila, 2019)
- Even in non-trial settings though, using an LRM can be a good way to get a gut-check on your product terms & cross-group comparisons
    - This is particularly true if most of your predictions fall in the linear portion of the curve (e.g., between 30-70%)
    - Less helpful if interested in true conditional predictions or high/low probabilities
- But recognize that reviewers may want you to do the work anyway—e.g., confirm that results don’t change when you move to BRM…

Thank you!



shawnana@umich.edu
@notthatdrsmith // @icpsrcda

Resources

Ai, C. & Norton, E.C. 2003. Interaction terms in logit and probit models. Economics Letters 80: 123–129.

Allison, P. D. 1999. Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods and Research*. 28: 186-208.

Berry, W.D., DeMeritt, J.H. & Esarey, J., 2010. Testing for interaction in binary logit and probit models: is a product term essential?. *American Journal of Political Science*, 54: 248-266.

Breen, R., Karlson, K.B. & Holm, A. 2018. Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models. *Annual Review of Sociology*, 44: 39-54.

Breen, R., Karlson, K. B., & Holm, A. 2018. A Note on a Reformulation of the KHB Method. *Sociological Methods & Research*, online only.

Gomila, R. 2019. Logistic or linear? Estimating causal effects of treatments on binary outcomes using regression analysis. Online at https://psyarxiv.com/4gmbv *(pre-print)*.

Greene, W.H. 2010. Testing hypotheses about interaction terms in non-linear models. Econ Lett 107, 291–296.

Hoetker, G.P., 2004. Confounded coefficients: Extending recent advances in the accurate comparison of logit and probit coefficients across groups. *Available at SSRN 609104*.

Karlson, Holm, and Breen. 2012. Comparing regression coefficients between same-sample nested models using logit and probit: A new model. *Sociological Methodology*, 42: 286-313.

Mize, Trenton D. 2019. Best practices for estimating, interpreting, and presenting non-linear interaction effects. *Sociological Science*, 6: 81-117.

Mize, Trenton D., Long Doan, J. Scott Long. 2019. A General Framework for Comparing Predictions and Marginal Effects Across Models. *Sociological Methodology*.

Rainey, C., 2016. Compression and conditional effects: A product term is essential when using logistic regression to test for interaction. *Political Science Research and Methods*, *4*: 621-639.

Extra slides

Impact of assumption

With probit, ε is rescaled so that:

$$Var(\varepsilon) = Var\left(\frac{\varepsilon}{\sigma}\right) = 1$$

With logit, ε is rescaled so that:

$$Var(\varepsilon) = \frac{\pi}{\sqrt{3}} \frac{\varepsilon}{\sigma} = \frac{\pi^2}{3}$$

Possible interpretation…

“There is a significant gender gap in alcohol use across most levels of education, with women reporting lower levels of alcohol use across all education levels except among those with a college degree (see column 4 in Table 1). The gender gap is largest among those without a high school degree (significantly larger than the gender gap for those with: a high school degree, some college, and a college degree (all $p < 0.05$; see column 5 in Table 1).”