Understanding linear and nonlinear interactions

Jeremy Dawson

CARMA webcast lecture

15 September 2023

Lecture Overview

- Testing and interpreting linear 2-way interactions
  - Testing interaction effects
  - Interpreting effects
  - Post-hoc testing
- Extensions to generalized linear models
- Extensions to quadratic regression models
- Other extensions

Two-way linear interactions

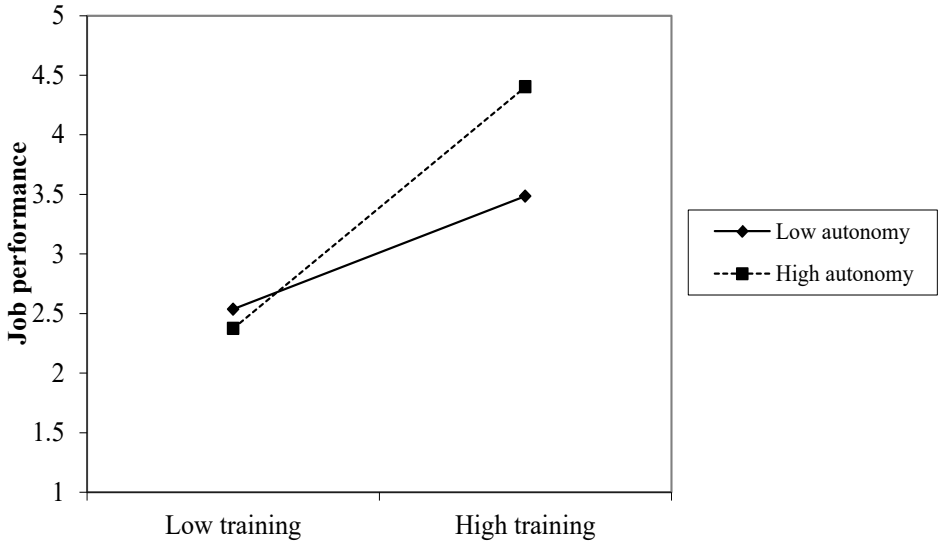
Moderation

- Often used synonymously with interaction effects
- Simple case: where the relationship between two variables (x and y) depends on the value of a third variable (z)
- For example:

  Training → Job performance

  Autonomy is a moderator

  Effect is stronger (more positive) when autonomy is higher



A simple **moderated** regression model

Equation: $y = a + b_1x + b_2z + b_3xz + \varepsilon$

b_3 represents the moderation effect

- Note: b_3 cannot be interpreted in isolation; also need to consider b_1 and b_2 to get full effect!
- If b_3 significant, we then interpret the model

Why is this a moderated model?

- Can be rewritten as

$$y = (a + b_2 z) + (b_1 + b_3 z)x + \varepsilon$$

- Therefore both the intercept and the slope (of x) vary as a linear function of z
- Note that b_1 is the slope of x when z = 0

Testing two-way interactions

- IV: TRAIN (extent of training)
- Moderator: AGE (age)
- DV: JOBSAT (job satisfaction)

In SPSS:

```
compute TRAXAGE = TRAIN*AGE.
regression /statistics = r coeff bcov
 /dependent = JOBSAT
 /method = enter TRAIN AGE TRAXAGE.
```

In R:

```
intmod1 <- lm(JOBSAT ~ TRAIN*AGE, data=PDW)
summary(intmod1)
```

In Stata:

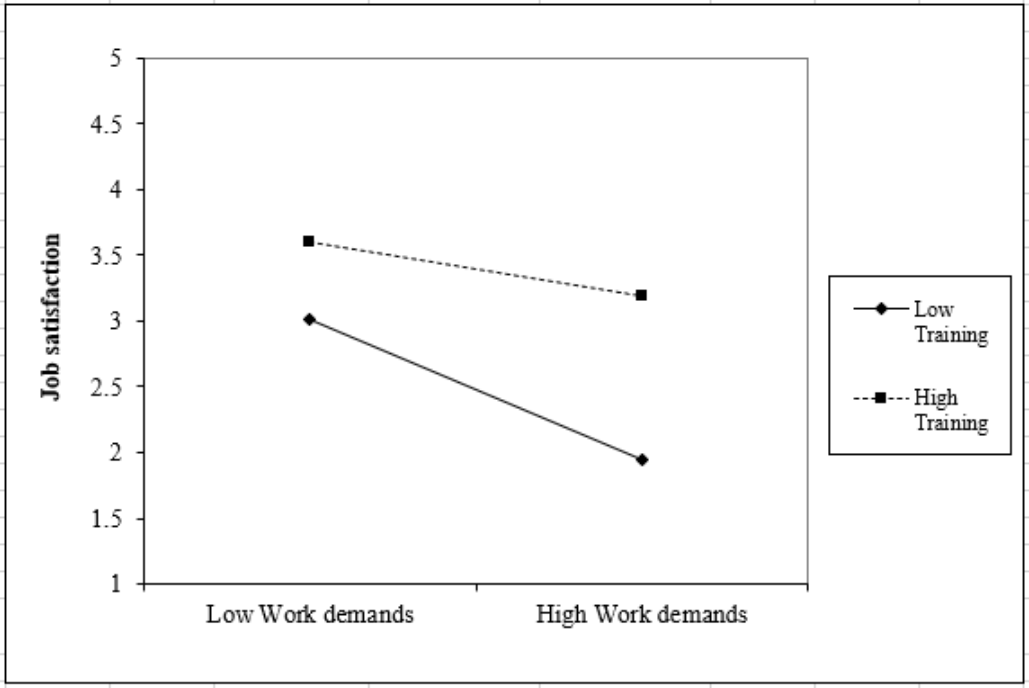
```
regress JOBSAT c.TRAIN##c.AGE
```

Plotting two-way interactions

- Some software will do this automatically
- e.g. in R, use the “marginaleffects” package and the “plot_predictions” command
- In Stata, can use “margins” with “plot”
- If not (e.g. in SPSS), can use Excel tools at www.jeremydawson.com/slopes.htm

2-way linear interactions template

Variable names:	
Name of independent variable:	Work demands
Name of moderator:	Training
Name of dependent variable:	Job satisfaction
Unstandardised Regression Coefficients:	
Independent variable:	-0.69378
Moderator:	0.172
Interaction:	0.1368
Intercept / Constant:	3.107
Means / SDs of variables:	
Mean of independent variable:	0
SD of independent variable:	1
Mean of moderator:	0
SD of moderator:	1
Values of variables at which to plot slopes*:	
Low value of IV:	1.333333333
High value of IV:	4.166666667
Low value of moderator:	2.333333333
High value of moderator:	4
(If left blank, this will automatically be done at one standard deviation above and below mean)*	
Optional alternative legend:**	
Low value of independent variable:	
High value of independent variable:	
Low value of moderator:	
High value of moderator:	
*(** Leave these cells blank to get the normal "low/high" legend)*	



Key questions along the way…

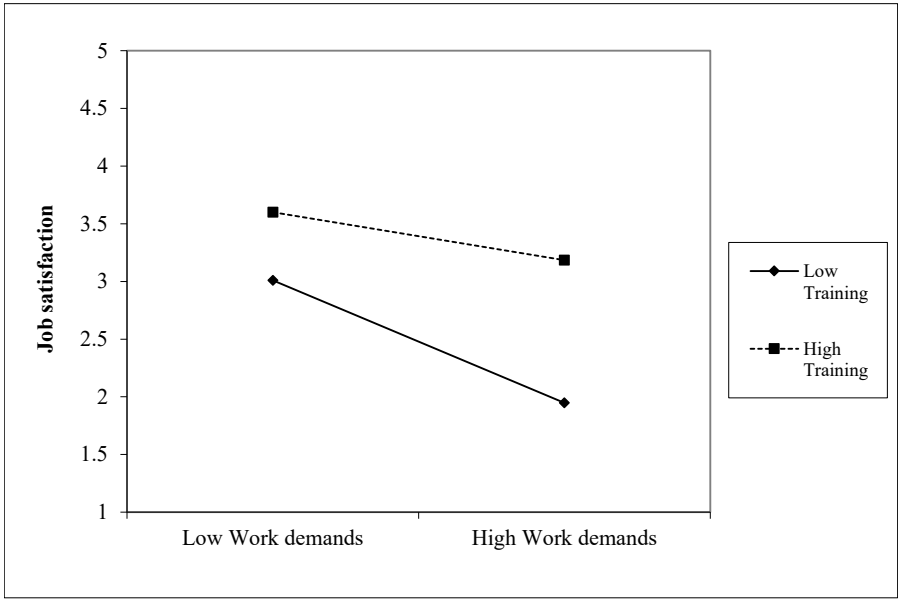
- Should you center variables before the analysis?
  - Doesn't matter in terms of the findings, but some interpretation is easier if you do
- What values of the moderator should the effect be plotted at?
  - Should be "typical" low and high values – not necessarily 1SD below and above the mean

Further probing of the effect

- A good plot is often sufficient if the moderator is continuous!
  - Especially if it also contains confidence bands (e.g. using "marginaleffects" package in R)
- Simple slope tests: specific conditional tests of the x-y relationship at particular values of the moderator
  - Will say more about these in a moment
- "Regions of significance" (Johnson-Neyman technique) – not recommended

Simple slope tests

- Common question: is the relationship between work demands and job satisfaction still significant when training is high, or is it only when training is low?



Simple slope tests: two methods

- Direct method
  - Uses elements of the coefficient covariance matrix to calculate the standard error of the slope at particular value, and test for significance
- Indirect method
  - Center the moderator around the value you want to test at – then re-run analysis. Coefficient of IV is the simple slope (with corresponding p-value, confidence interval etc.)

Simple slope tests: a warning!

- These only tell us whether there is evidence of a non-zero effect at a **specific** value of the moderator
- Often this is entirely arbitrary – and thus the results are not very informative (look at what happens if we change the high value of the moderator in our example)
- Better to choose **meaningful** values of the moderator, not just 1 SD above and below the mean

Other features to describe

- Use the coefficient b_3 to describe the effect: this is the change in the x-y slope as z increases by one unit
- Is the x-y effect always positive, mostly positive, a crossover effect etc.?
- Is the effect a disordinal one - where the lines cross?
- If they do, the lines cross at value of X where the value of Z makes no difference to y. This can be calculated at

$$x = -\frac{b_2}{b_3}$$

(Equation: $y = a + b_1x + b_2z + b_3xz$)

Interactions with generalized linear models

Generalized linear models

- Apply a link function to the linear combination of x and z to convert it to the required metric
- Instead of

$$y = a + b_1x + b_2z + b_3xz + \varepsilon$$

- We have

$$E[y] = g^{-1}(a + b_1x + b_2z + b_3xz)$$

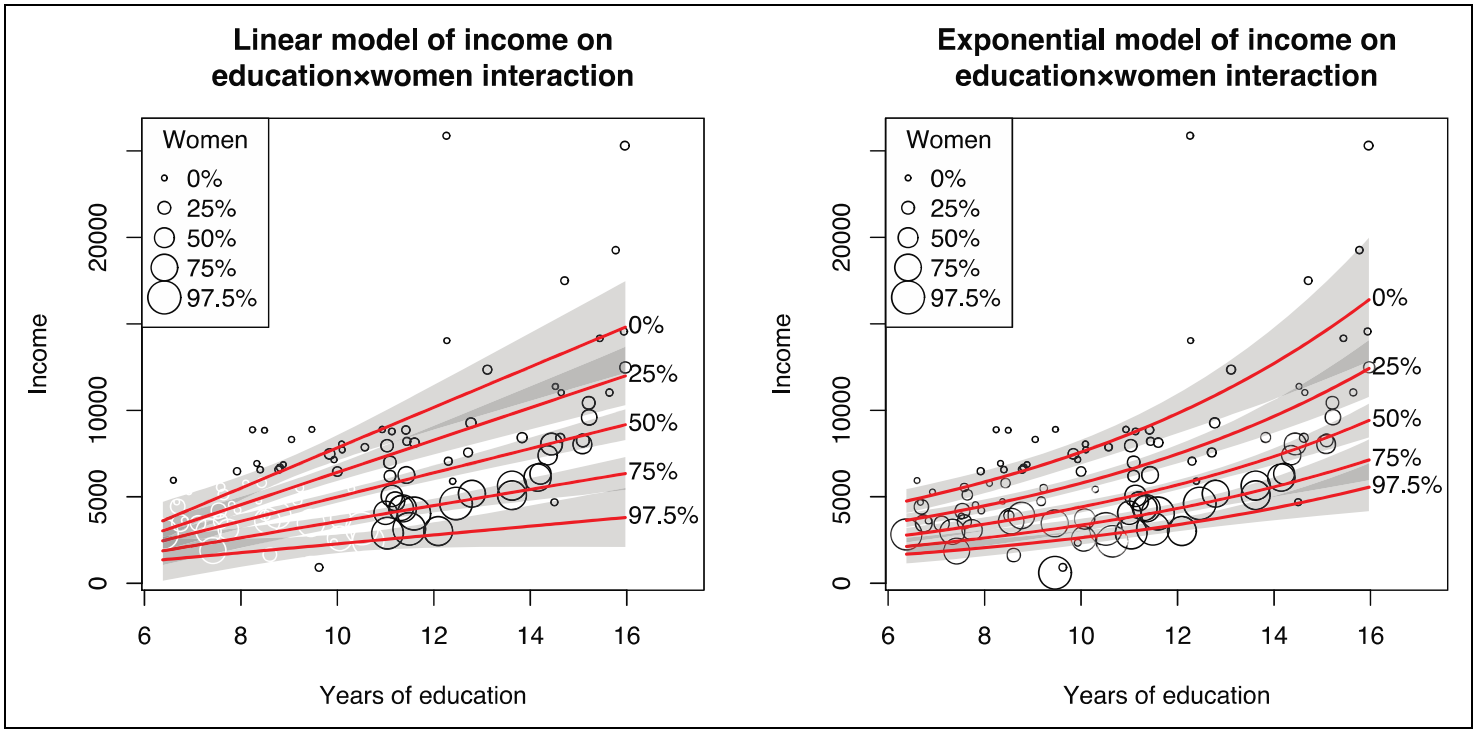
- For example:
  - Binary logistic regression (logit link function)
  - Poisson regression (log link function)
  - Ordinal logistic regression (ordered logit link)

Moderation with generalized linear models

- In some senses this is easy: if you know how to used the generalized linear model, and know how test linear interactions, the two are just combined!

- However, it becomes more difficult in the interpretation

Example (Rönkkö et al., 2022)



Rönkkö, M., Aalto, E., Tenhunen, H., & Aguirre-Urreta, M. I. (2022). Eight simple guidelines for improved understanding of transformations and nonlinear effects. *Organizational Research Methods*, *25*(1), 48–87. https://doi.org/10.1177/1094428121991907

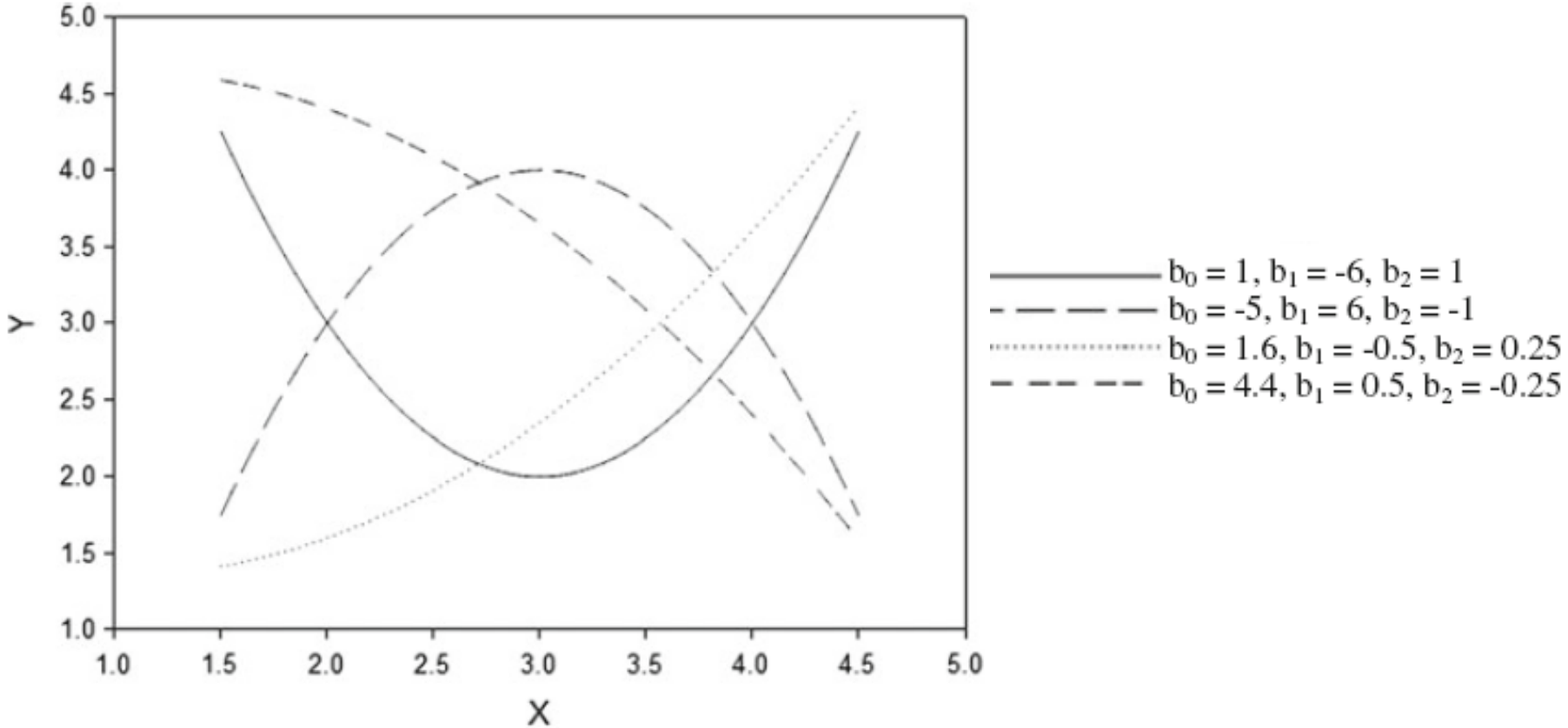
Key issues to consider

- Is the moderation desired of the linear effect, or of the proportion effect?
- Are you using the right model for this?
- Three possibilities to interpret differences:
  - Testing the linear predictor metric
  - Testing the average marginal effect
  - Testing differences between four points

Interactions with quadratic regression models

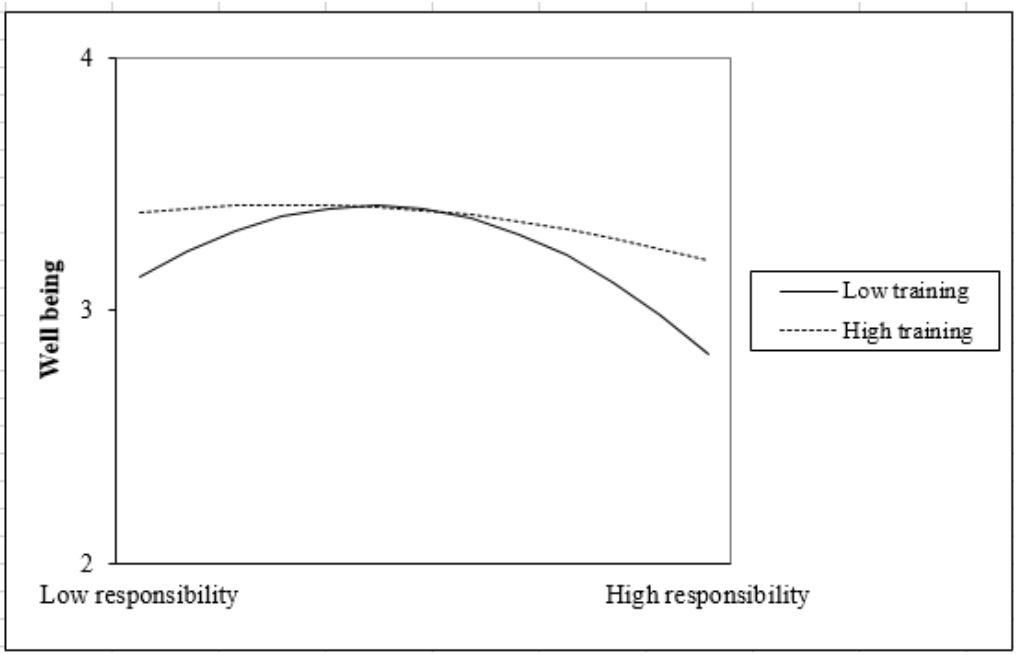
Quadratic models

- Quadratic equation: $y = a + b_1x + \mathbf{b_2x^2}$



Quadratic interactions

- $y = a + b_1x + b_2x^2 + b_3z + b_4xz + b_5x^2z$
  - If b_5 significant, this means the curvature is moderated
  - If b_4 significant, this means the location of the curve is moderated



Probing curvilinear interactions

- Most important thing: plot the effect!
  - Quadratic two-way interactions sheet
- Identify where maximum/minimum point occurs
  - It occurs where $x = -b_1/2b_2$, where b_1 is X coefficient and b_2 is X^2 coefficient
  - Use indirect method to center around moderator value and find this for each curve plotted
- Are "simple slope" tests necessary?

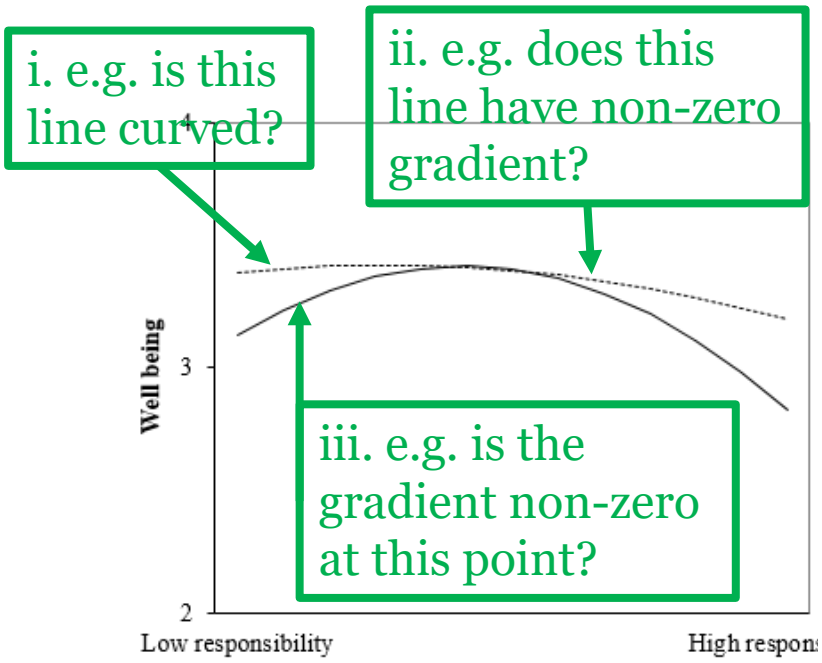
Variations on simple slope tests

- Three versions:

  i. Testing whether there is a *curvilinear* effect at a particular value of the moderator

  ii. Testing whether there is *any* effect at a particular value of the moderator

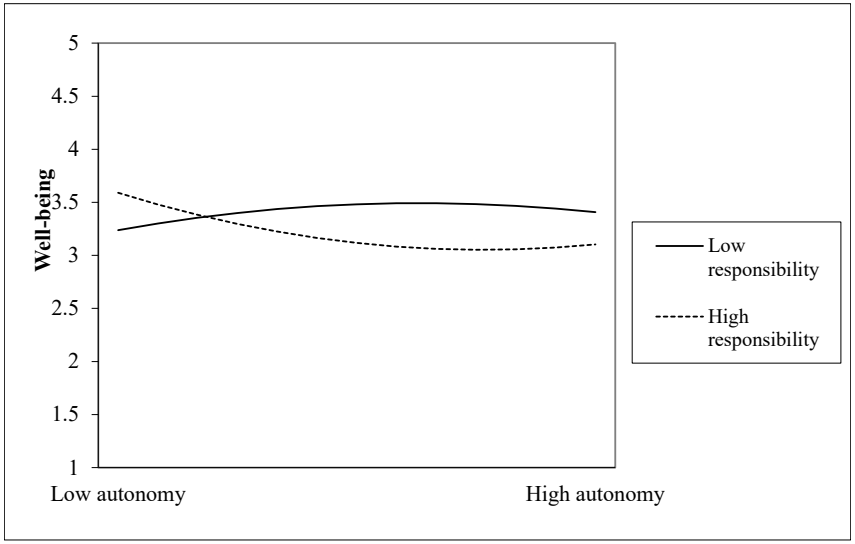
  iii. Testing whether there is any effect at a particular value of the moderator *and* a particular value of the independent variable



Where do the curves cross?

- $y = a + b_1x + b_2x^2 + b_3z + b_4xz + b_5x^2z$
- (Maximum of) two crossing points for curves:

$$x = \frac{-b_4 \pm \sqrt{{b_4}^2 - 4b_3b_5}}{2b_5}$$



Should you do these tests?

- Only if there's a specific need to!

Other extensions

Multilevel models

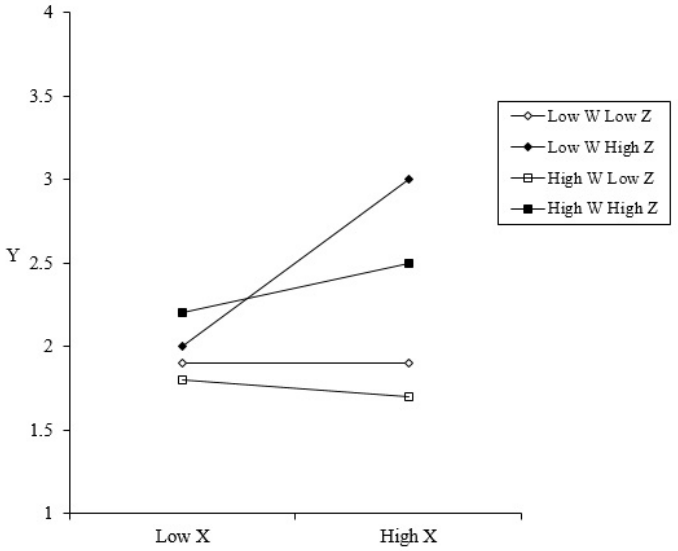
- Interactions can be plotted using the same template as relevant for single-level interactions
  - Estimates produced in output are equivalent to unstandardized coefficients in ordinary regression
  - Care is needed over mean & SD of variables
- However, in general, simple slope & slope difference tests in SPSS need to use the indirect method

Three-way linear interactions

- Doubly-moderated models

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_1x_2 + b_5x_1x_3 + b_6x_2x_3 + b_7x_1x_2x_3$$

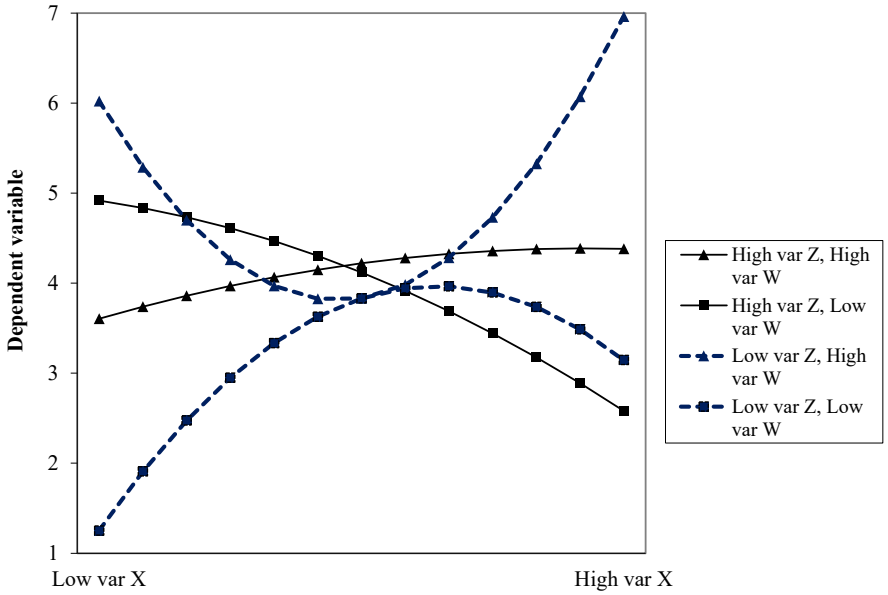
- As before, b_7 cannot be interpreted in isolation: needs plotting
- Can also be supplemented by slope difference test (Dawson & Richter, 2006)



Three-way quadratic interactions

$$y = a + b_1x + b_2x^2 + b_3z + b_4w + b_5xz + b_6x^2z + b_7xw + b_8x^2w + b_9zw + b_{10}xzw + b_{11}x^2zw$$

- No direct equivalent to slope difference test: but can test whether maximum/minimum points, or the curvature, are the same on two curves



Moderation and mediation

- Needs a whole other session!
- However, PROCESS is a good start
- Worth working out all equations (see e.g. Edwards & Lambert, 2007)
- Most specific hypotheses can be tested using bootstrapping (or delta method, e.g. in Stata)

Interactions within structural equation models

- In principle, all techniques that apply to regression apply to SEMs
  - Specialized packages and functions, e.g. semTools::probe2WayMC
  - Note: specialized packages can be black boxes
- Simple slopes tests can be done with Wald tests
  - require understanding the scales of the latent variables to choose meaningful values of moderators

Questions?