Statistical Modeling and Causal Inference with R

Week 10: Moderation and heterogeneous effects

Manuel Bosancianu

Max Schaub

November 16, 2020

Hertie School of Governance

Today's focus

- ✓ Motivation
- ✓ Moderation vs. mediation
- ✓ Understanding treatment heterogeneity
- ✓ Estimating heterogeneous treatment effects
- ✓ Caveats

Motivation

Motivation

1. Philosophical
   - ✓ What sets the social sciences apart from many hard sciences is that we are population sciences is that we are dealing with individuals who differ in manifold ways from each other (Xie, 2013).
   - ✓ One individual cannot 'stand in' for the next, like e.g. molecules in chemistry. This is the reason why we have to make our inferences at the group level: while groups can be counterfactuals for each other, individuals rarely can.
   - ✓ Still it is somewhat paradoxical that in causal inference we often only estimate the average effect for all individuals, notwithstanding the fact that the recognition of individual differences are at the heart of our science. We would therefore clearly expect a treatment to have different effects for different individuals.
   - ✓ The explicit modeling of heterogeneity in treatment effects for subgroups addresses this tension between the necessity of having to do inference at the group level, and the recognition of individual differences.

Motivation

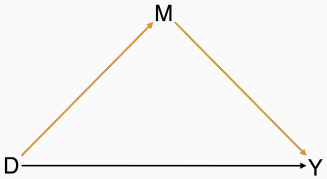
2. Inference:
   - ✓ We commonly invoke the constant effects assumption $E[Y_{1i}|D = 1] = E[Y_{0i}|D = 0] + \kappa$, e.g. when using simple additive regression models for estimating treatment effects
   - ✓ Where this assumption does not hold, results can be biased (Angrist, 1998; Elwert & Winship, 2010)
   - ✓ To obtain consistent results, we therefore have to include interaction terms/allow for heterogeneity.

Motivation

3. Policy relevance:
   - ✓ Analysis of heterogeneity can be very important for policy
   - ✓ The same treatment may be effective have (beneficial) effects for one subgroup, but be harmful for another, e.g. certain types of surgeries.
   - ✓ Resources can be much more efficiently spent if it is clear who reacts to a treatment, and who does not.

Moderation vs. mediation

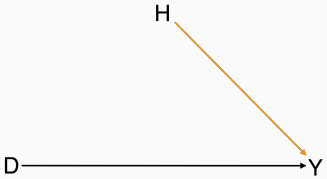
Mediator: variable that lies on the causal path from treatment D to outcome Y; explains (part of) the effect



Theme of next class!

Moderation vs. mediation

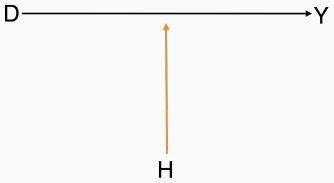
Moderator: variable that divides treated population in groups among which the treatment D has differential effects



DAGs not very well suited to depict moderation (Hernán & Robins, 2021, 80); H could also be additional predictor that is independent of X; here meant to indicate effect modification/heterogeneity

Moderation vs. mediation

Moderators sometimes depicted as an arrow aimed at the arrow linking D and Y



But note that this no longer is a DAG...

Moderation vs. mediation

Different terms used for moderation, depending on scientific field:

- ✓ Moderation
- ✓ Effect modification
- ✓ Interaction effects
- ✓ Conditional treatment effects
- ✓ Heterogeneous treatment effects
- ✓ Conditional average treatment effects (CATEs)

Understanding treatment heterogeneity

Understanding treatment heterogeneity

What is the issue? Consider the following two tables:

Constant treatment effects				
Individual	Y_{0i}	Y_{1i}	δ_i	Age group
A	4	7	3	Young
B	3	6	3	Young
C	-1	2	3	Young
D	11	14	3	Young
E	2	5	3	Old
F	1	4	3	Old
G	4	7	3	Old
H	0	3	3	Old
Av	3	6	3	

Homogeneous/constant treatment effect across subjects, as invoked by many proofs (e.g. use of OLS for estimating treatment effect)

Understanding treatment heterogeneity

What is the issue? Consider the following two tables:

Heterogeneous treatment effects				
Individual	Y_{0i}	Y_{1i}	δ_i	Age group
A	0	7	7	Young
B	1	6	5	Young
C	0	2	2	Young
D	4	14	10	Young
E	4	5	1	Old
F	4	4	0	Old
G	7	7	0	Old
H	4	3	-1	Old
Av	3	6	3	

Heterogeneous treatment effect, clearly patterned along a covariate. Effect only among young ($\delta_{Young} = 6$), zero among older.

Potentially highly policy relevant, e.g. if effect of vaccine.

Understanding treatment heterogeneity

In reality, like the average treatment effect, heterogeneous treatment effects have to be estimated because we only ever observe one potential outcome.

Heterogeneous treatment effects			
Individual	$Y_{0i}\|D=0$	$Y_{1i}\|D=1$	Age group
A	0		Young
B	1		Young
C		2	Young
D		14	Young
E	4		Old
F		4	Old
G	7		Old
H		3	Old
Av	3	5.75	2.75

Here the estimated treatment effect is 5.75, with that for the young being 7.5, and that for the old -2.

Descriptive vs. causal conditional treatment effects

- ✓ With randomization of D_i, the treatment is identified for all subgroups.
- ✓ However, unless the subgroups are the result of randomization, the factor along which heterogeneity is assessed does *not* have a causal interpretation.
- ✓ This is because this factor could be capturing the effect of another variable
- ✓ You *can* give a causal interpretation to your interaction effect if you are interacting two treatments that were both randomized.

Estimation of heterogeneous treatment effects

Estimation of heterogeneous treatment effects

Heterogeneous treatment effects are usually estimated with regression models that include an interaction between the treatment and the moderator:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 H_i + \beta_3 H_i \times D_i + \epsilon_i$$

Compare this to a conventional additive model in the form

$$Y_i = \gamma_0 + \gamma_1 D_i + \gamma_2 H_i + \mu_i$$

How are the coefficients in the different models to be interpreted?

Estimation of heterogeneous treatment effects

In the additive model, as you know, γ_1 is the marginal effect of D_i on Y_i:

$$\frac{\partial Y_i}{\partial D_i} = \gamma_1$$

and γ_2 is the marginal effect of H_i on Y_i:

$$\frac{\partial Y_i}{\partial H_i} = \gamma_2$$

i.e. the coefficients capture what happens to Y_i if D_i or H_i increase by one unit, respectively.

Estimation of heterogeneous treatment effects

In contrast, in the interaction model, the marginal effect of D_i on Y_i is:

$$\frac{\partial Y_i}{\partial D_i} = \beta_1 + \beta_3 H$$

meaning the effect of D_i depends on H_i – which is what we are after, of course!

For a binary moderating variable (that takes values 0 and 1), the marginal effect of D_i is $\beta_1 + \beta_3$ if the moderator is one.

The partial effect also implies that the marginal effect of D_i on Y_i is β_1 if $\beta_3 H$ is zero. In the case of a binary moderator, this coefficient is directly interpretable.

For a moderator that takes more than one value, plausible marginal effects should be calculated.

Estimation of heterogeneous treatment effects

To sum up, the parameters in an interaction model

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 H_i + \beta_3 H_i \times D_i + \epsilon_i$$

are interpreted as follows:

β_0: Constant
β_1: Effect of D_i on Y_i if H_i is zero.
β_2: Effect of H_i on Y_i if D_i is zero.
β_3: Difference in treatment effects of D_i depending on H_i.

In other words, in the binary case, β_3 is the difference-in-differences in means $((E[Y_i|D=1,H=1] - E[Y_i|D=0,H=1]) - (E[Y_i|D=1,H=0] - E[Y_i|D=0,H=0])$, and the difference in regression slopes where the interacted variable is continuous.

Calculating and plotting CATEs

An example: Recall study on right-wing support following expansion of broadband coverage (Schaub & Morisi, 2020).

Basic finding: broadband availability goes along with higher support for right-wing populists.

Does the finding hold for all demographics, e.g. age groups? In other words, is the effect of broadband access moderated by age?

We start with a binary moderator as in the following model

$$Y_i = \beta_0 + \beta_1 BB + \beta_2 Older_i + \beta_3 Older_i \times BB_i + \epsilon_i$$

where *BB* indicates broadband availability, which can be high (1) or low (0), and *Older* is an indicator that is 1 if an individual is 55 years or more, and 0 if else.

Calculating and plotting CATEs

Conditional effect of broadband access on right-wing support	
BB (β_1)	0.136***
	(0.05)
Older (β_2)	0.033
	(0.08)
Older × BB (β_3)	-0.149*
	(0.08)
Constant (β_0)	0.062
	(0.04)
Observations	1,158
R^2	0.03

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Calculating and plotting CATEs

What is the conditional average treatment effect for the non-old/the younger?

That's the constitutive term for the treatment indicator β_1 – which gives us the treatment indicator when the moderator is zero, i.e. $\beta_1 = 0.136$.

But is there an effect for the old at all (i.e. an effect different from zero)? The regression table cannot tell.

While we can calculate the treatment effect for the old from the table as

$(\beta_1 + \beta_3) = 0.136 - 0.149 = -0.013$,

we cannot tell if this effect is statistically significant/different from zero. This is because we lack the standard error for this quantity of interest i.e.

$\hat{\sigma}_{\frac{\partial Y_i}{\partial D_i}} = \sqrt{var(\hat{\beta_1}) + var(\hat{\beta_3}) + 2cov(\hat{\beta_1}\hat{\beta_3})}$.

Calculating and plotting CATEs

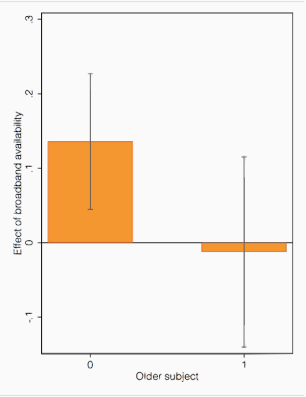
We can calculate this standard error with a Wald-Test of the hypothesis $H0 : \beta_1 + \beta_3 = 0$, or by having R calculate the marginal effects:

Marginal effects of broadband access on right-wing support	
if Older=0	0.136^{***}
	(0.05)
if Older=1	-0.013
	(0.07)
Observations	1,158

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Calculating and plotting CATEs

We can also plot the results:



Calculating and plotting CATEs

Finally, we may ask if the differences in heterogeneous treatment effects are different, i.e. test $H0 : \beta_1 - (\beta_1 + \beta_3) = 0$ – but that's β_3!

So we can just take this information from the regression table.

Here, despite the substantively big difference, this difference in treatment effect sizes is only marginally significant at p=0.06.

This is because information becomes sparse the moment we start subdividing our sample.

Interaction term vs. split sample

Sometimes you see researchers splitting their sample, i.e. calculate simple additive models for the different levels of the moderator variable. Is this legit?

The separate additive models and the interaction model are equivalent as long as only the treatment indicator D_i and the moderator H_i are included in the interaction model.

Once additional control variables are included, the interacted model and the split are *no longer* equivalent.

This is because in the interacted model, all control variables are constrained to have the same effect for the full sample.

In the split model, control variables can have sample-specific effects.

The 'split-model effect' can be emulated by including interaction terms for *all* included control variables.

Interaction term vs. split sample

Consider again our example.

Without additional covariates, the split sample models recover the marginal effects calculated with the interaction model

Split-sample regression without additional covariates		
	Older=0	Older=1
BB	0.136***	-0.013
	(0.05)	(0.07)
Constant	0.062	0.095
	(0.04)	(0.06)
Observations	581	577
R^2	0.01	0.00

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Interaction term vs. split sample

With additional covariates, the interaction and the split sample model are no longer equivalent because the treatment indicator interacts with the additional covariates in each subsample in different ways.

Split-sample regression with additional covariates

	Older=0	Older=1	Interaction	All interacted
BB	0.108**	-0.001	0.131***	0.108**
	(0.05)	(0.07)	(0.05)	(0.05)
Share w/ degree	0.260	-0.141	0.049	0.260
	(0.19)	(0.10)	(0.10)	(0.19)
Older=1			0.031	0.170
			(0.08)	(0.10)
Older=1 × BB			-0.147*	-0.108
			(0.08)	(0.08)
Older=1 × Share w/ degree				-0.401*
				(0.21)
Constant	-0.023	0.147**	0.046	-0.023
	(0.07)	(0.07)	(0.05)	(0.07)
Observations	581	577	1,158	1,158
R^2	0.01	0.00	0.03	0.03

Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Interaction term vs. split sample

This coefficients of the split model + covariates can be recovered with the fully interacted model. This is rarely used in practice, however.

Marginal effects after split-sample regression with additional covariates				
	Older=0	Older=1	Interaction	All interacted
BB	0.108**	-0.001		
	(0.05)	(0.07)		
BB if Older=0			0.131***	0.108**
			(0.05)	(0.05)
BB if Older=1			-0.017	-0.001
			(0.07)	(0.07)
Observations	581	577	1,158	1,158

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Calculating and plotting CATEs

In treatment-by-treatment interactions (i.e. where we manipulated two treatments simultaneously), while all the above applies, we are usually interested in the marginal effect of all the interaction terms. Consider the model
$Y_i = \beta_0 + \beta_1 D_i^1 + \beta_2 D_i^2 + \beta_3 D_i^1 \times D_i^2 + \epsilon_i$

with the two treatments D_i^1 and D_i^2.

These define four outcomes:

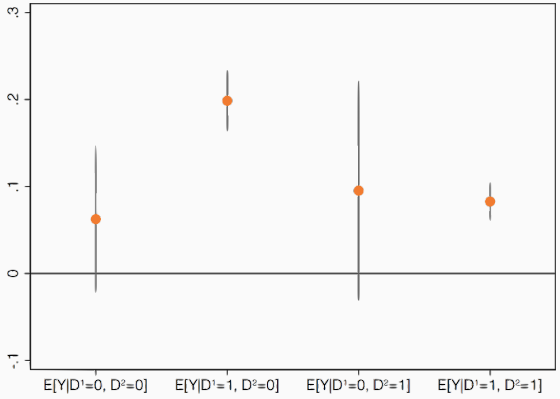
$$\begin{array}{ll} (E[Y_i|D_i^1 = 0, D_i^2 = 0] & \beta_0 \\ (E[Y_i|D_i^1 = 1, D_i^2 = 0] & \beta_0 + \beta_1 \\ (E[Y_i|D_i^1 = 0, D_i^2 = 1] & \beta_0 + \beta_2 \\ (E[Y_i|D_i^1 = 1, D_i^2 = 1] & \beta_0 + \beta_1 + \beta_2 + \beta_3 \end{array}$$

(Remember this from DiD?)

Note that only for β_0 the regression table provides the relevant standard error to gage the statistical significance of these quantities of interest. Calculating and plotting the margins thus becomes essential.

Calculating and plotting CATEs

Plot of marginal effect for experiment with 2x2 treatments:



Calculating and plotting CATEs

In treatment-by-treatment interactions, we will usually also be interested in pairwise comparisons and their significance, i.e. the question whether the estimated effects just shown are significantly different from each other:

Pairwise comparison	Associated hypothesis test
$(E[Y_i\|D_i^1 = 0, D_i^2 = 0]$ vs. $(E[Y_i\|D_i^1 = 1, D_i^2 = 0]$	$H0 : \beta_0 - (\beta_0 + \beta_1) = 0$
$(E[Y_i\|D_i^1 = 0, D_i^2 = 0]$ vs. $(E[Y_i\|D_i^1 = 0, D_i^2 = 1]$	$H0 : \beta_0 - (\beta_0 + \beta_2) = 0$
$(E[Y_i\|D_i^1 = 0, D_i^2 = 0]$ vs. $(E[Y_i\|D_i^1 = 1, D_i^2 = 1]$	$H0 : \beta_0 - (\beta_0 + \beta_1 + \beta_2 + \beta_3) = 0$
$(E[Y_i\|D_i^1 = 1, D_i^2 = 0]$ vs. $(E[Y_i\|D_i^1 = 0, D_i^2 = 1]$	$H0 : \beta_0 + \beta_1 - (\beta_0 + \beta_2) = 0$
$(E[Y_i\|D_i^1 = 1, D_i^2 = 0]$ vs. $(E[Y_i\|D_i^1 = 1, D_i^2 = 1]$	$H0 : \beta_0 + \beta_1 - (\beta_0 + \beta_1 + \beta_2 + \beta_3) = 0$
$(E[Y_i\|D_i^1 = 0, D_i^2 = 1]$ vs. $(E[Y_i\|D_i^1 = 1, D_i^2 = 1]$	$H0 : \beta_0 + \beta_2 - (\beta_0 + \beta_1 + \beta_2 + \beta_3) = 0$

Note: Once we do these multiple comparisons, we have to start worrying about 'fishing' for statistically significant differences (more below)

Calculating and plotting CATEs

What if the modifying variable is continous?

For example, consider the model:

$$Y_i = \beta_0 + \beta_1 BB + \beta_2 Age_i + \beta_3 Age_i \times BB_i + \epsilon_i$$

where *BB* indicates broadband availability, which can be high (1) or low (0), and *Age* is a continuous variable recording an individual's age.

Calculating and plotting CATEs

The corresponding regression table looks as follows:

Conditional effect of broadband access on right-wing support	
BB (β_1)	0.327***
	(0.09)
Age (β_2)	0.002
	(0.00)
BB × Age (β_3)	-0.005**
	(0.00)
Constant (β_0)	-0.039
	(0.08)
Observations	1,158
R^2	0.03

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note that here the constitutive term β_1 indicates the effect of broadband access when age is zero – it basically has no substantively interesting interpretation: calculation of marginal effects is mandatory in this case.

Calculating and plotting CATEs

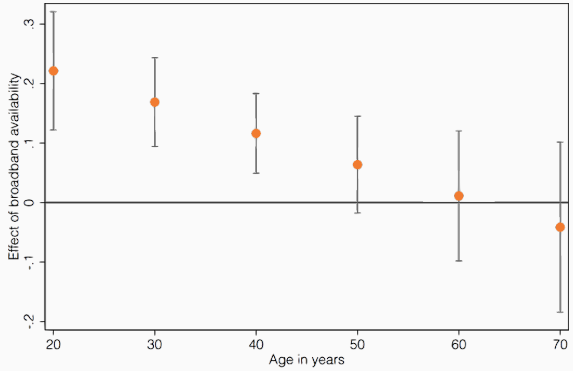
The table below shows marginal effects for individuals aged 20, 30, 40, 50, 60, and 70.

Marginal effects of broadband access on right-wing support	
Age = 20	0.221***
	(0.05)
Age = 30	0.169***
	(0.04)
Age = 40	0.116***
	(0.03)
Age = 50	0.064
	(0.04)
Age = 60	0.011
	(0.06)
Age = 70	-0.041
	(0.07)
Observations	1,158

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Calculating and plotting CATEs

For effective communication, the best practice is to plot the results:



Some caveats with regard to moderation analysis

Heterogeneous treatment effects can convey very important information that is often highly policy relevant.

However, they are often seen with skepticism by the research community.

This is because of the problem of 'fishing': with the number of heterogeneous treatment effects calculated, the likelihood of finding a significant effect grows.

Assuming the conventional p-value of 0.05 used to mark findings as 'significant', 1/20 coefficients will be statistically significant *by chance alone*.

This opens the possibility for researchers to calculate many heterogeneous effects, and to seek out and report 'significant' ones (Humphreys, de la Sierra, & van der Windt, 2013).

Some caveats with regard to moderation analysis

Several solutions to the 'multiple-comparison' problem have been proposed:

1. Statistical control, e.g. Bonferroni correction: for each comparison you do, the critical value against which a result is found to be significant is reduced.

   For example, if 4 values are reported, coefficients should meet the critical p-value of 0.05/4=0.0125 to be considered statistically significant. While rigorous, this test is often considered too conservative.

Some caveats with regard to moderation analysis

2. Automate the search for treatment effects using vector machines (R::FindIt), Bayesian additive regression trees (R::BayesTree), classification and regression trees (R::causalTree), random forests, and kernel regularized least squares (R::KRLS);

   These methods can 1) help to find the most important splits in the data, and 2) take discretion away from the researcher, thereby alleviating concerns with 'fishing' – however, they are atheoretical and may 'miss' finding heterogeneity – or the absence of it! – in groups that are important for policy makers.

Some caveats with regard to moderation analysis

3. Preregistration: researchers pre-register the comparisons and splits of the data they intend to investigate *before* collecting and/or analyzing the data; i.e. they 'bind their hands' as to which results they will present.

   This idea has become very popular. Also, writing a preregistration plan is very good practice for laying out your hypotheses *before* you start collecting data.

Some caveats with regard to moderation analysis

Moderation analysis is often used as a way to test for causal mechanisms.

The idea is that if the treatment effect varies strongly along an interacted variable, that variable must have *something* to do with the underlying causal mechanism.

As stated, this type of logic is at best highly informal, and at worst throughly misleading.

The only case where moderation can be used in the analysis of mechanisms in a rigorous fashion is when the stipulated mechanism is turned into a treatment, and treatment-by-treatment interactions are calculated.

Otherwise, the formal analysis of causal mechanisms is the domain of mediation analysis.

Thank you for watching, and see you next Monday!

References i

Angrist, J. (1998). Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants. *Econometrica*, *66*(2), 249–288.

Elwert, F., & Winship, C. (2010). Effect heterogeneity and bias in main-effects-only regression models. In R. Dechter, H. Geffner, & J. Y. Halpern (Eds.), *Heuristics, probability and causality: A tribute to Judea Pearl* (pp. 327–36). College Publications.

Hernán, M., & Robins, J. M. (2021). *Causal inference*.

Humphreys, M., de la Sierra, R. S., & van der Windt, P. (2013). Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration. *Political Analysis*, *21*(1), 1–20.

References ii

Schaub, M., & Morisi, D. (2020). Voter Mobilisation in the Echo Chamber: Broadband Internet and the Rise of Populism in Europe. *European Journal of Political Research*, *59*(4).

Xie, Y. (2013, April). Population heterogeneity and causal inference. *Proceedings of the National Academy of Sciences*, *110*(16), 6262–6268.