Moderation and Mediation Effects

(taken from the previous slides of Giacomo Lemoli)

2024-04-26

Today's plan

- Mediation
- Moderation

Mediation: concepts review

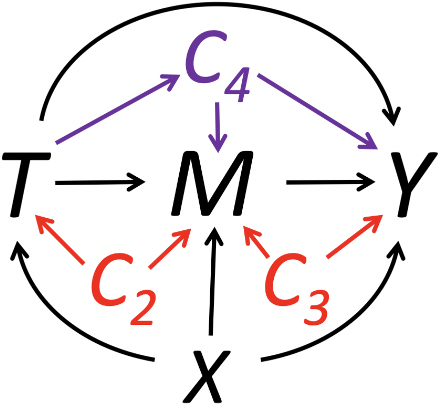
- Total Effect: $\tau_i = Yi(1, Mi(1)) - Yi(0, Mi(0))$
- Natural Direct Effect: $\zeta_i(t) = Y_i(1, M_i(t)) - Y_i(0, M_i(t))$
- Natural Mediation Effect: $\delta_i(t) = Y_i(t, M_i(1)) - Y_i(t, M_i(0))$
- $\tau_i = \delta_i(t) + \zeta_i(1 - t)$
- ID assumption for δ and ζ: sequential ignorability

Sequential ignorability

$$T_i \perp (Yi(t', m), M_i(t))|X_i = x$$
$$M_i(t) \perp Y_i(t', m)|T_i, X_i = x$$

- First part: CIA, satisfied in a randomized experiment
- Second part: no omitted post-treatment confounder or other mediator causally connected to M

Sequential ignorability



Causal mediation analysis

- R package `mediation`: causal mediation analysis under sequential ignorability
- Identify effect of T on M given X and the effect of M on Y given T and X
- With them compute the direct/mediation effects
- In the special case of linear models one can multiply the coefficients

Causal mediation analysis

- Working example from the `mediation` package: Brader et al (2008)
- T: Media stories about immigration
- Y: Letter about immigration policy to representative in Congress
- M: Anxiety
- X: Age, education, gender, income

Casual mediation analysis

```
library(mediation)

data(framing)

set.seed(2014)

# Model for the mediator (T + X)
med.fit <- lm(emo ~ treat + age + educ + gender + income, data = framing)

# Model for the outcome (M + T + X)
out.fit <- glm(cong_mesg ~ emo + treat + age + educ + gender + income, data = framing,
               family = binomial("probit"))

# Compute the mediation effects
med.out <- mediate(med.fit, out.fit, treat = "treat", mediator = "emo",
                   robustSE = TRUE, sims = 100)
```

Causal mediation analysis

```
summary(med.out)
```

```
##
## Causal Mediation Analysis
##
## Quasi-Bayesian Confidence Intervals
##
##                          Estimate 95% CI Lower 95% CI Upper p-value
## ACME (control)             0.0791       0.0351         0.15  <2e-16 ***
## ACME (treated)             0.0804       0.0367         0.16  <2e-16 ***
## ADE (control)              0.0206      -0.0976         0.12    0.70
## ADE (treated)              0.0218      -0.1053         0.12    0.70
## Total Effect               0.1009      -0.0497         0.23    0.14
## Prop. Mediated (control)   0.6946      -6.3109         3.68    0.14
## Prop. Mediated (treated)   0.7118      -5.7936         3.50    0.14
## ACME (average)             0.0798       0.0359         0.15  <2e-16 ***
## ADE (average)              0.0212      -0.1014         0.12    0.70
## Prop. Mediated (average)   0.7032      -6.0523         3.59    0.14
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Sample Size Used: 265
##
##
## Simulations: 100
```

Controlled Direct Effect

- Controlled Direct Effect: $\kappa_i(m) = Y_i(1, m) - Y_i(0, m)$
- Effect of T on Y when M has the same value for all units.
- Relative to NDE and NME, identified also in presence of intermediate confounders
- A natural approach to close M channel: include M as control in the regression
- In presence of intermediate confounders, this introduces post-treatment bias
- CDE is an estimand that allows to overcome this issue

Sequential g-estimation

- Popularized in political science by Acharya, Blackwell, and Sen (2016)
- Procedure in two stages:
- Regress Y on M, T, pre-treatment and intermediate variables
- Subtract from Y the effect of M: get the "demediated" Y, $\tilde{Y} = Y - \hat{\beta_M} M$
- Regress $\tilde{Y}$ on T and pre-treatment variables
- Doing it by hand would ignore variability in $\tilde{Y}$, which is an estimated quantity, resulting in wrong SEs
- Use the package `DirectEffect` or bootstrap
- Center the mediator at the value you want to "fix" it at

Sequential g-estimation

- Why is it important?
- Research questions may involve comparisons that "hold fixed" things realized after the treatment
- Do natural shocks impact political development even in absence of physical destruction?
- Does ethnic diversity lead to conflict even in absence of government instability?
- We may also want to rule causal mechanisms alternative to our theory
  - Are the effects of slavery/famine just due to subsequent changes in racial/ethnic composition? (Acharya, Blackwell, and Sen (2016); Rozenas and Zhukov (2019) resp.)

Application

- Alesina, Giuliano, and Nunn (2013): data provided with the `DirectEffects` package
- Y: share of political positions held by women in 2000
- T_i: relative proportion of ethnic groups that traditionally used the plow within a country
- M_i: log GDP per capita in 2000, mean-centered
- Z_i: post-treatment, pre-mediator intermediate confounders
  - civil conflict, interstate conflict, oil, European descent, communist, polity2..)
- X_i: pre-treatment characteristics of the country
  - tropical climate, agricultural suitability, large animals, political hierarchies, economic complexity, rugged

Application

```r
library(DirectEffects)

data("ploughs")

## ATE
ate_mod <- lm(women_politics ~ plow + agricultural_suitability + tropical_climate +
                large_animals + political_hierarchies + economic_complexity + rugged,
              data = ploughs)

summary(ate_mod)[[4]]["plow",]
```

```
##   Estimate Std. Error    t value   Pr(>|t|)
## -2.1031536  2.1270350 -0.9887725  0.3244216
```

Application

```r
## Formula for sequential_g
form_main <- women_politics ~ plow + agricultural_suitability + tropical_climate +
  large_animals + political_hierarchies + economic_complexity + rugged | # pre-treatment vars
  years_civil_conflict + years_interstate_conflict  + oil_pc + european_descent +
  communist_dummy + polity2_2000 + serv_va_gdp2000 | # intermediate vars
  centered_ln_inc + centered_ln_incsq # mediating vars

## Sequential g-estimation
direct <- sequential_g(formula = form_main, data = ploughs)
```

Application

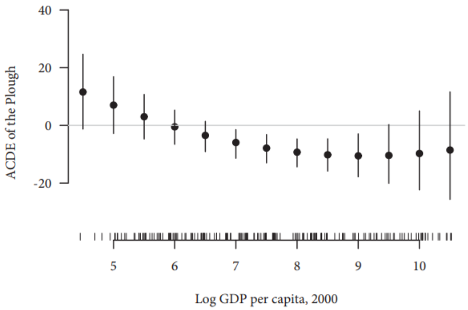
```
summary(direct)
```

```
##
## t test of coefficients:
##
##                          Estimate Std. Err. t value Pr(>|t|)
## (Intercept)              12.18450   3.64442  3.3433 0.001121 **
## plow                     -4.83879   2.34467 -2.0637 0.041312 *
## agricultural_suitability  4.57388   3.10477  1.4732 0.143458
## tropical_climate         -2.18919   2.10505 -1.0400 0.300554
## large_animals            -1.33001   3.40008 -0.3912 0.696401
## political_hierarchies     0.49575   1.09060  0.4546 0.650283
## economic_complexity      -0.10521   0.42973 -0.2448 0.807029
## rugged                   -0.30869   0.47821 -0.6455 0.519888
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

More on DirectEffects

- Sensitivity analysis using `cdesens` function
- Can center mediator at different values to see how CDE varies at different values of M

FIGURE 8. The ACDE of the Plough as a Function of the Fixed Level of Current-day Income



Note: vertical lines are 95% confidence intervals from 1,000 bootstrapped replications.

Moderation

- Characterize treatment effect heterogeneity
  - Why is it important?
  - Knowledge: going beyond the aggregation
  - Policy: on what sub-populations the intervention is more effective
  - Mechanisms: understanding what units drive the average effect gives insights about what the treatment is doing
  - Methodologically: regression-based methods vs non-parametric methods

Moderation in regression

- Classical approach: interaction terms. Let's start from the case of binary treatment D_i and binary moderator Z_i.

$$y_i = \alpha + \beta D_i + \gamma Z_i + \delta D_i * Z_i + \epsilon_i$$

- β: effect of D_i when $Z_i = 0$
- γ: effect of Z_i when $D_i = 0$
- δ: increase in the effect of D_i when Z_i goes from 0 to 1
- $\beta + \delta$: effect of D_i when $Z_i = 1$
- With continuous D_i and/or Z_i: restate in terms of marginal effects (increase the variable by 1 unit)

Moderation in regression

- Adding interaction term resembles the DiD methodology
- Important difference: in DiD the interaction Group $\times$ Post estimates the ATT under parallel trends
- In moderation, the interaction estimates the variation of ATE/ATT across strata of Z
- Careful about coefficients interpretation

Interpreting moderation in regression

- Continuous Z

$$y_i = \alpha + \beta D_i + \gamma Z_i + \delta D_i * Z_i + \epsilon_i$$

Recall:

- β: effect of D_i when $Z_i = 0$
- δ: increase in the effect of D_i when Z_i increases by 1
- $\beta + \delta * z$: average effect of D_i when $Z_i = z$

Note:

- β is an ATE for a sub-group without necessarily a substantive value: may not even exist in the data
- If center Z_i, e.g. interact with $\tilde{Z} = (Z_i - \bar{Z}_i)$ then β is the ATE at the mean of the moderator (interpretable as population ATE)

Moderation vs sub-group effects

- δ tells us by how much the ATE *varies* in a sub-group relative to a reference sub-group
- It is *not* the ATE for a subgroup. E.g. ATE(z) is given by $\beta + \delta * z$
- Standard packages compute the effect of D for sub-groups with different values of Z
  - Stata: `margins`. R: `margins` and `marginaleffects`

Issues with linear interaction terms

- Linear interaction terms used to study how the treatment effect evolves over the distribution of the moderator
- Hainmueller, Mummolo, and Xu (2019) point out that this practice relies on requirements which might be violated
- TE changes linearly in the moderator at any point of its distribution
  - May be non-linear or non-monotonic
- Common support between treatment and moderator
  - If not, the model relies on linear extrapolation

Working example

Slaveholding and state-building in the US South

Slavery, Reconstruction, and Bureaucratic Capacity in the American South

PAVITHRA SURYANARAYAN *Johns Hopkins University*
STEVEN WHITE *Syracuse University*

Conventional political economy models predict taxation will increase after franchise expansion to low-income voters. Yet, contrary to expectations, in ranked societies—where social status is a cleavage—elites can instead build cross-class coalitions to undertake a strategy of bureaucratic weakening *to limit future redistributive taxation. We study a case where status hierarchies were particularly extreme: the post-Civil War American South. During Reconstruction, under federal oversight, per capita taxation was higher in counties where slavery had been more extensive before the war, as predicted by standard theoretical models. After Reconstruction ended, however, taxes fell and bureaucratic capacity was weaker where slavery had been widespread. Moreover, higher intrawhite economic inequality was associated with lower taxes and weaker capacity after Reconstruction in formerly high-slavery counties. These findings on the interaction between intrawhite economic inequality and pre-War slavery suggest that elites built cross-class coalitions against taxation where whites sought to protect their racial status.*

interflex

- `interflex` package (in both R and Stata) proposes a more flexible procedure to moderation, proposed by Hainmueller, Mummolo, and Xu (2019)

Binning estimator

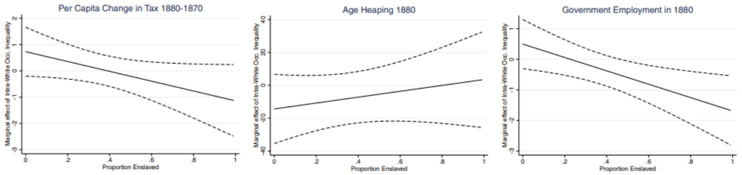
- Divide the support of Z into j bins (e.g. terciles), indicated by G_j, and estimate

$$y_{ij} = \sum_{j=1}^{J} \{\alpha_j + \beta_j D_{ij} + \gamma_j (Z_{ij} - Z_j^M) + \delta_j (Z_{ij} - Z_j^M) D_{ij}\} G_j + \psi X_{ij} + \epsilon_{ij}$$

- Z_j^M is the median value of Z inside bin j. Given the specification, β_js are the conditional ATEs at the center of each bin.

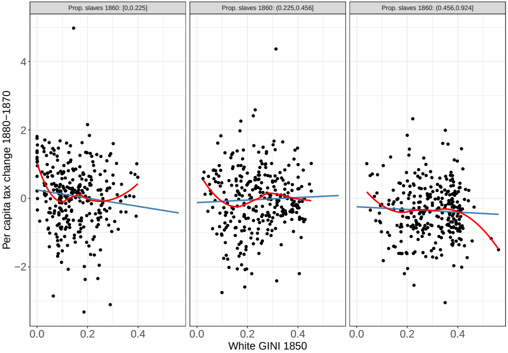
Moderation with linear estimator

FIGURE 7. Marginal Effect of Intrawhite Inequality on Taxation and Bureaucratic Quality



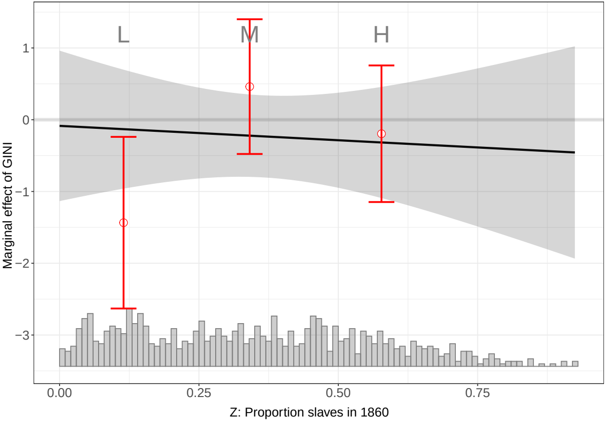
Moderation using binning estimator

```
library(interflex); library(haven); library(tidyverse)

# Import the data
d <- read_dta("suri_white_preprocessed.dta") %>% as.data.frame()

# Raw data plot
# Note we are not including controls used in the paper
interflex(estimator = "raw", data = d, Y = "tax_diff", D = "county_sei_gini_whitemale_1850",
                X = "pslave1860", ylab = "Marginal effect of GINI",
                xlab = "Z: Proportion slaves in 1860", theme.bw = T, ncols=3,
          Dlabel = "White GINI 1850", Ylabel = "Per capita tax change 1880-1870",
          Xlabel = "Prop. slaves 1860")
```



Moderation using the binning estimator

```
# Heterogeneous TE with interflex (note in this notation Z and X are inverted)
out <- interflex(estimator = "binning", data = d,
                 Y = "tax_diff", D = "county_sei_gini_whitemale_1850",
                 X = "pslave1860", ylab = "Marginal effect of GINI",
                 xlab = "Z: Proportion slaves in 1860", theme.bw = T)

out$figure
```



Kernel estimator

- Allow TE to vary over the whole distribution of the moderator, estimating the following semiparametric model

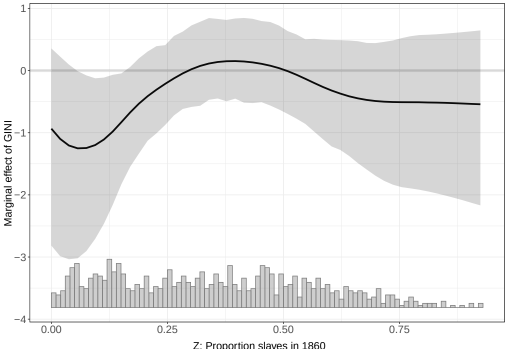
$$y_i = f(Z_i) + g(Z_i)D_i + h(Z_i)X_i + \epsilon_i$$

Moderation with the kernel estimator

```
set.seed(123)
outk <- interflex(estimator = "kernel", data = d,
                  Y = "tax_diff", D = "county_sei_gini_whitemale_1850",
                  X = "pslave1860", ylab = "Marginal effect of GINI",
                  xlab = "Z: Proportion slaves in 1860", theme.bw = T)
```

```
## Cross-validating bandwidth ...
## Parallel computing with 4 cores...
## Optimal bw=0.1222.
## Number of evaluation points:50
## Parallel computing with 4 cores...
##
```

```
outk$figure
```



Diagnostic tools

- `interflex` also gives diagnostic tools for model specification
- E.g. Wald tests for the hypothesis that the simple linear interaction is correct
- With a slight reparametrization, the null hypothesis is that the coefficients within each bin but one are jointly 0, i.e. constant coefficients

Diagnostic tools

- `interflex` also gives diagnostic tools for model specification
- E.g. Wald tests for the hypothesis that the simple linear interaction is correct
- With a slight reparametrization, the null hypothesis is that the coefficients within each bin but one are jointly 0, i.e. constant coefficients

```
out$tests$p.wald
```

```
## [1] "0.151"
```