POIR 611: Estimating the Effects of Human Rights Organizations

Suzie Mulesky

March 27, 2020

1 Introduction

1.1 Outline

- **interaction effects** - how to estimate and properly interpret them
- **marginal effects** - how to plot and properly interpret them
- **regression weights** and **leverage** - how they can help us interpret the generalizability/external validity of coefficient estimates

1.2 What you will get out of this lecture

- a step-by-step guide of how to replicate and diagnose an existing published paper in a top political science journal
- how to estimate and visualize interaction terms in R
- how to calculate and interpret multiple regression weights and leverage statistics in R

2 Literature: human rights advocacy

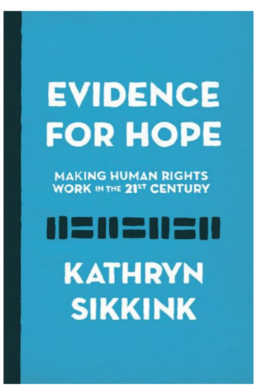
- Hundreds of qualitative case studies
- Dozens of quantitative papers
- **12 statistical papers that specifically estimate the effect of human rights NGOs on respect for human rights**

Authors	Year	Title	Journal
Hafner-Burton	2008	Sticks and Stones: Naming and Shaming the Human Rights Enforcement Problem	International Organization
Franklin	2008	Shame on You: The Impact of Human Rights Criticism on Political Repression in Latin America	International Studies Quarterly
Cardenas	2010	Conflict and Compliance: Responses to International Human Rights Pressure	University of Pennsylvania Press
Murdie and Davis	2012	Shaming and blaming: Using events data to assess the impact of human rights INGOs	International Studies Quarterly
Krain	2012	J'accuse! Does Naming and Shaming Perpetrators Reduce the Severity of Genocides or Politicides?	International Studies Quarterly
DeMeritt	2012	International Organizations and Government Killing: Does Naming and Shaming Save Lives?	International Interactions
Bell, Clay, and Murdie	2012	Neighbordhood Watch: Spatial Effects of Human Rights INGOs	Journal of Politics
Hendrix and Wong	2013	When Is the Pen Truly Mighty? Regime Type and the Efficacy of Naming and Shaming in Curbing Human Rights Abuses	British Journal of Political Science
Murdie and Hicks	2013	Can International Nongovernmental Organizations Boost Government Services? The Case of Health	International Organization
Hill and Jones	2014	An Empirical Evaluation of Explanations for State Repression	American Political Science Review
Murdie	2014	Help or Harm: The Human Security Effects of International NGOs	University of Stanford Press
Murdie and Peksen	2015	Women's Rights INGO Shaming and the Government Respect for Women's Rights	Review of International Organizations

2.1 What is the received wisdom?

> "Multiple studies, using a wealth of the best data on the topic, have shown that we can be cautiously optimistic about the impact of the work of human rights INGOs, but that for the greatest success, information politics need to be combined with efforts to build strong domestic advocacy sectors within states, while also bringing pressure to bear from outside."

- Kathryn Sikkink (2017) *Evidence for Hope*



2.2 Research Question

Does the statistical evidence support the conclusion that human rights NGO advocacy has had a significant impact on reducing state repression?

2.3 Let's look at the most-cited paper

- Amanda Murdie and David R. Davis (2012) "Shaming and Blaming: Using Events Data to Assess the Impact of Human Rights INGOs" *International Studies Quarterly*
- Key findings:

> "Mere shaming is not enough. Improvements in human rights practices result from the interaction of shaming by human rights organizations (HROs) with **a domestic presence of HROs within the targeted state.**"

In mathematical notation...

Y Physical Integrity Rights = α + β_1 HRO Shaming + β_2 Domestic HRO Presence + β_3 (HRO Shaming X Domestic HRO Presence) + ϵ

2.4 What are their reported results?

Variable	*Model 1*	*Model 2*
HRO Shaming	0.0101 (0.0388)	−0.458 (0.211)*
HRO Shaming × HRO Presence (ln)		0.112 (0.0474)*
HRO Shaming × Indirect Targeting		
HRO Shaming × (HRO Presence (ln) + Indirect Targeting)		
Indirect Targeting		
HRO Presence (ln)		0.0562 (0.134)
(HRO Presence (ln) + Indirect Targeting)		
Population (ln)	−0.576 (0.0352)**	−0.560 (0.0478)**
GDP per capita (ln)	0.371 (0.0430)**	0.393 (0.0505)**
Polity score (−10 to 10)	0.0755 (0.00762)**	0.0708 (0.00947)**
War (interstate or intrastate)	−1.93 (0.205)**	−2.01 (0.202)**
Intercept	4.95 (0.0620)**	4.74 (0.433)**
N	1371	1175
F statistic	181.76**	121.24**

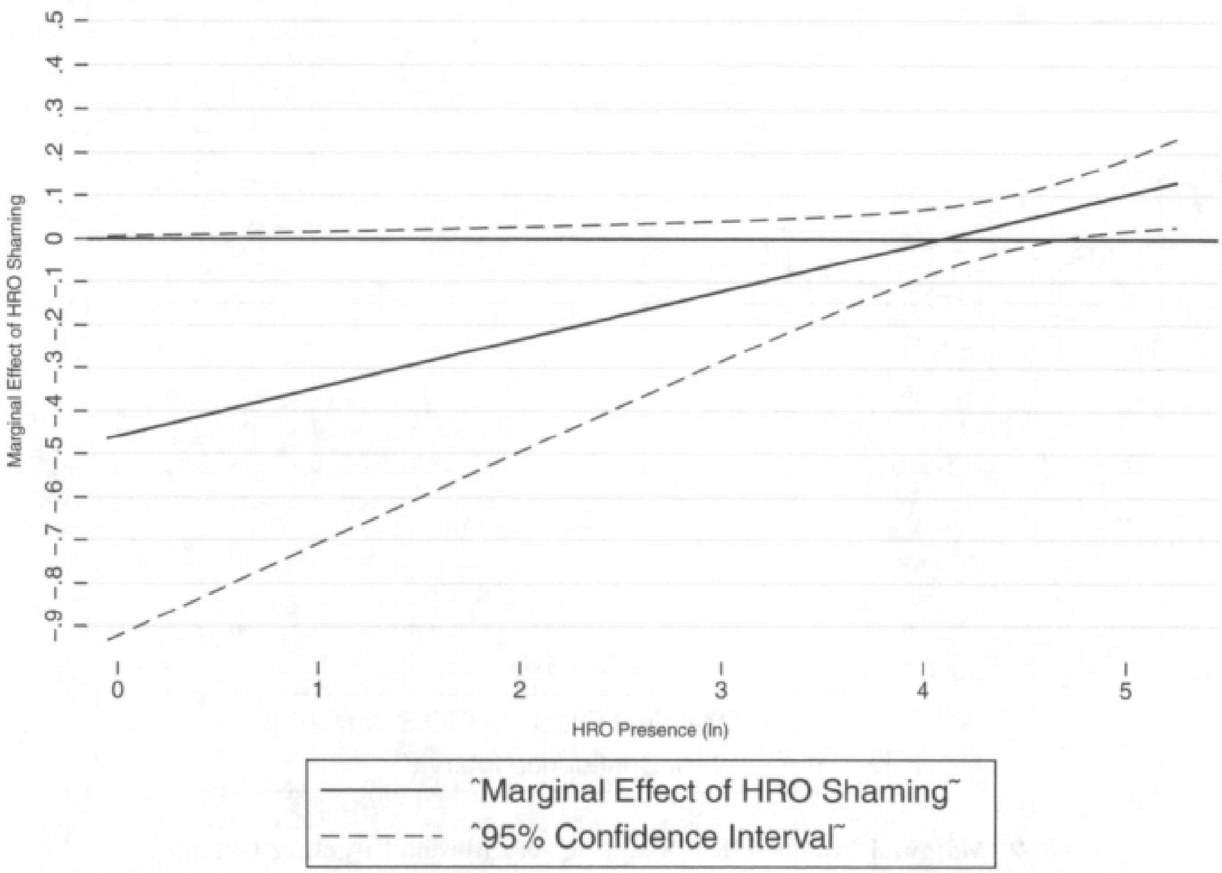


FIG 1. Marginal Effects: HRO Shaming As HRO Presence (ln) Changes

3 Let's replicate!

Download their replication data here.

3.1 Read data

```
# Load packages
library(readstata13)
library(dplyr)
library(ggplot2)
library(plotMElm)

# Set your working directory first
murdie_davis <- read.dta13("ISQ 2010 Murdie Davis final to ISQ.dta")
```

3.2 A little bit of data cleaning using the `dplyr` package

Note:

- We have to omit observations in which respect for physical integrity rights is highest (`physint == 8`) because Murdie and Davis (2012) omit these observations from their model.
- `na.omit()` performs list-wise deletion

```r
df <- murdie_davis %>%
  select(year, cowcode, NAMES_STD, physint, Fphysint, Lphysint,
         HRnc2gcnc2, lnhrfilled, normaledlnpop, normaledlngdp,
         polity2, WARonlocation) %>%
  filter(Fphysint < 8 & physint < 8 & Lphysint < 8) %>%
  na.omit()
```

3.3 Variable names and descriptions

- **physint:** Respect for physical integrity rights (ordinal, 0-8)
- **HRnc2gcnc2:** Number of HRO shaming events (count)
- **lnhrfilled:** Domestic HRO presence (natural log)
- **normaledlnpop:** Population (natural log)
- **normaledlngdp:** GDP (natural log)
- **polity2:** Regime type
- **WARonlocation:** Participation in war (binary)

3.4 Descriptive statistics

```r
df %>%
  select(physint, HRnc2gcnc2, lnhrfilled, normaledlnpop, normaledlngdp, polity2, WARonlocation) %>%
  summary()
```

```
##     physint        HRnc2gcnc2        lnhrfilled    normaledlnpop
##  Min.   :0.000   Min.   : 0.0000   Min.   :0.000   Min.   :-2.06751
##  1st Qu.:3.000   1st Qu.: 0.0000   1st Qu.:3.029   1st Qu.: 0.09215
##  Median :4.000   Median : 0.0000   Median :3.434   Median : 0.94348
##  Mean   :4.219   Mean   : 0.5104   Mean   :3.395   Mean   : 1.07752
##  3rd Qu.:6.000   3rd Qu.: 0.0000   3rd Qu.:3.829   3rd Qu.: 1.89410
##  Max.   :7.000   Max.   :27.0000   Max.   :4.934   Max.   : 5.73355
##  normaledlngdp        polity2        WARonlocation
##  Min.   :-3.0498   Min.   :-10.000   Min.   :0.0000
##  1st Qu.:-1.4310   1st Qu.: -4.000   1st Qu.:0.0000
##  Median :-0.4068   Median :  5.000   Median :0.0000
##  Mean   :-0.3248   Mean   :  2.475   Mean   :0.0665
##  3rd Qu.: 0.6165   3rd Qu.:  8.000   3rd Qu.:0.0000
##  Max.   : 3.0630   Max.   : 10.000   Max.   :1.0000
```

3.5 Estimate the OLS regression model

Note: Murdie and Davis (2012) estimate Newey-West robust standard errors. We're going to ignore that for the purpose of this demonstration.

```r
model <- lm(Fphysint ~ HRnc2gcnc2*lnhrfilled + normaledlnpop + normaledlngdp + polity2 +
              WARonlocation, data = df)
summary(model)
```

```
##
## Call:
## lm(formula = Fphysint ~ HRnc2gcnc2 * lnhrfilled + normaledlnpop +
##     normaledlngdp + polity2 + WARonlocation, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -5.3520 -0.9771  0.1193  1.0943  5.1748
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            4.665521   0.331058  14.093  < 2e-16 ***
## HRnc2gcnc2            -0.471562   0.168390  -2.800  0.00519 **
## lnhrfilled             0.075272   0.103336   0.728  0.46650
## normaledlnpop         -0.555692   0.037607 -14.776  < 2e-16 ***
## normaledlngdp          0.385477   0.036825  10.468  < 2e-16 ***
## polity2                0.069586   0.008133   8.556  < 2e-16 ***
## WARonlocation         -2.030347   0.188291 -10.783  < 2e-16 ***
## HRnc2gcnc2:lnhrfilled  0.114731   0.039830   2.881  0.00404 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.585 on 1195 degrees of freedom
## Multiple R-squared:  0.4257, Adjusted R-squared:  0.4223
## F-statistic: 126.5 on 7 and 1195 DF,  p-value: < 2.2e-16
```

An alternative way to represent an interaction term, using the colon rather than asterisk:

```
model.int <- lm(Fphysint ~ HRnc2gcnc2 + lnhrfilled + HRnc2gcnc2:lnhrfilled +
                  normaledlnpop + normaledlngdp + polity2 +
                  WARonlocation, data = df)
model.int
```

```
##
## Call:
## lm(formula = Fphysint ~ HRnc2gcnc2 + lnhrfilled + HRnc2gcnc2:lnhrfilled +
##     normaledlnpop + normaledlngdp + polity2 + WARonlocation,
##     data = df)
##
## Coefficients:
##           (Intercept)             HRnc2gcnc2             lnhrfilled
##               4.66552               -0.47156                0.07527
##         normaledlnpop          normaledlngdp                polity2
##              -0.55569                0.38548                0.06959
##         WARonlocation  HRnc2gcnc2:lnhrfilled
##              -2.03035                0.11473
```

A third way to represent an interaction term by creating a variable that stores the interaction:

```
df$shame_HROs <- df$HRnc2gcnc2 * df$lnhrfilled
model.int2 <- lm(Fphysint ~ HRnc2gcnc2 + lnhrfilled + shame_HROs +
                  normaledlnpop + normaledlngdp + polity2 +
                  WARonlocation, data = df)

model.int2
```

```
##
## Call:
## lm(formula = Fphysint ~ HRnc2gcnc2 + lnhrfilled + shame_HROs +
##     normaledlnpop + normaledlngdp + polity2 + WARonlocation,
##     data = df)
##
## Coefficients:
##   (Intercept)     HRnc2gcnc2     lnhrfilled     shame_HROs  normaledlnpop
```

```
##       4.66552       -0.47156        0.07527        0.11473       -0.55569
## normaledlngdp        polity2  WARonlocation
##       0.38548        0.06959       -2.03035
```

3.6 Interpreting the continuous-by-continuous interaction

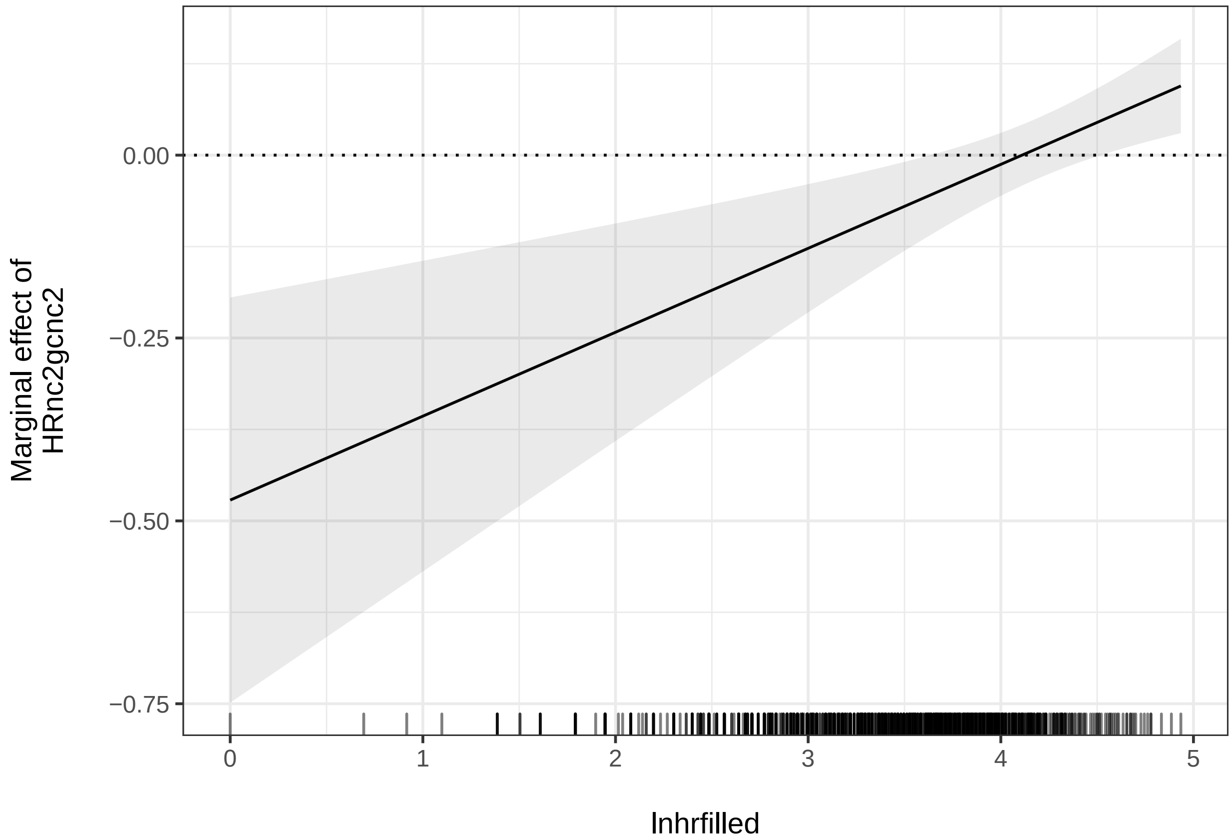
- This interaction is statistically significant with a p-value less than 0.01.
- The slope is positive. What does that mean?
- As the domestic presence of HROs increases by one unit, the change in respect for physical integrity rights associated with a one-unit increase of NGO naming and shaming increases by 0.115.
- A one-unit increase in HRO shaming * domestic HRO presence leads to a 0.115 unit increase in respect for physical integrity rights (holding all other variables constant).
- This isn't very intuitive. It helps to visualize the relationship.

3.7 Marginal Effects

- One way to explain a continuous-by-continuous interaction is to explore how the average effect of NGO naming and shaming changes as a function of the domestic presence of HROs.
- We can do this by calculating the slope of the treatment variable (HRO shaming) at different values of the moderator variable (domestic HRO presence).
- To do this, we can plot the **marginal effect** of HRO shaming across different values of domestic HRO presence. This allows us to visualize how the slope changes at low and high levels of domestic HRO presence.

3.7.1 Plot the marginal effect of HRO Shaming across the observed range of HRO Presence

```r
# plotMElm::plot_me
plot_me(model, 'HRnc2gcnc2', 'lnhrfilled', ci = 95)
```



3.7.2 Interpreting the marginal effects plot

- When values of `lnhrfilled` are greater than or equal to 4.5 (ln), `HRnc2gcnc2` has a statistically significant and positive effect on respect for physical integrity rights. For reference, $e^{4.5} \approx 90$.
- When values of `lnhrfilled` are between 3.5 (ln) and 4.5 (ln), `HRnc2gcnc2` does not have a statistically significant effect on respect for physical integrity rights. ($e^{3.5} \approx 33$)
- When values of `lnhrfilled` are less than 3.5 (ln), `HRnc2gcnc2` has a statistically significant and negative effect on respect for phyisical integrity rights.

3.7.3 Exercise

1. What is the estimated effect of NGO naming and shaming when there are only 7 domestic human rights organizations (about 2 ln)?

Answer: The slope is about -0.25, which means that a one unit increase in HRO shaming leads to a 0.25 unit decrease in respect for physical integrity rights when `lnhrfilled` equals 2 (ln).

The marginal effects plot automatically generates a rug plot, which displays the distribution of `lnhrfilled`. This helps us understand how the impact of NGO naming and shaming is manifested in the sample.

2. Are most countries positively affected, negatively affected, or not affected by human rights NGOs?

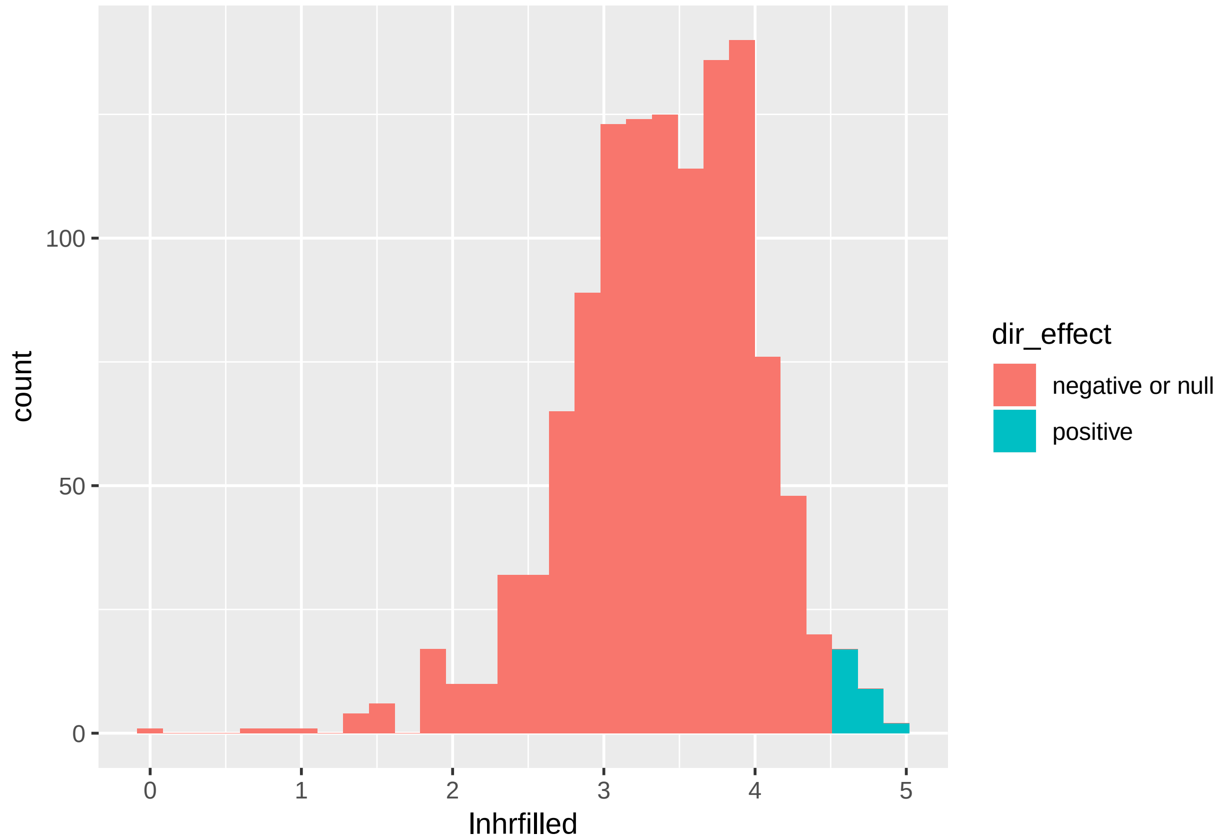
Answer: Based on the rug plot, we can see that there are very few observations with a value of `lnhrfilled` greater than 4.5.

3. What proportion of the estimation sample *is* positively affected by NGO naming and shaming?

```
df %>%
  mutate(dir_effect = ifelse(lnhrfilled >= 4.5,
```

```
                              "positive", "negative or null")) %>%
  ggplot(aes(x = lnhrfilled, fill = dir_effect)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



4. Which observations are positive?

```
df %>%
  filter(lnhrfilled >= 4.5) %>%
  group_by(NAMES_STD) %>%
  summarise(n_obs = n())
```

```
## # A tibble: 6 x 2
##   NAMES_STD      n_obs
##   <chr>          <int>
## 1 Canada             2
## 2 France             5
## 3 Italy              4
## 4 Spain              3
## 5 United Kingdom     7
## 6 United States      7
```

Yet Murdie and Davis claim that the critical level of HRO presence is within one standard deviation of the mean (around 50 HROs).

"Having a greater-than-average domestic presence of HROs within a targeted state is crucial for shaming to be an effective strategy for human rights promoters. **This level of HRO domestic presence was seen**

in Brazil, Senegal, and Sri Lanka, among many others, during this time period."

```r
summary(df$lnhrfilled)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   3.029   3.434   3.395   3.829   4.934
```

```r
sd(df$lnhrfilled)
```

```
## [1] 0.6167454
```

```r
mean(df$lnhrfilled) + sd(df$lnhrfilled)
```

```
## [1] 4.011995
```

```r
exp(4.011995)
```

```
## [1] 55.257
```

```r
exp(mean(df$lnhrfilled) + sd(df$lnhrfilled))
```

```
## [1] 55.25698
```

```r
df %>%
  filter(NAMES_STD %in% c("Brazil", "Senegal", "Sri Lanka")) %>%
  select(NAMES_STD, year, lnhrfilled)
```

```
##    NAMES_STD year lnhrfilled
## 1     Brazil 1992   3.864232
## 2     Brazil 1993   3.951244
## 3     Brazil 1994   4.007333
## 4     Brazil 1995   4.060443
## 5     Brazil 1996   4.151040
## 6     Brazil 1997   4.234107
## 7     Brazil 1998   4.262680
## 8     Brazil 1999   4.290460
## 9     Brazil 2000   4.290460
## 10    Brazil 2001   4.313034
## 11    Brazil 2002   4.335110
## 12   Senegal 1992   3.536117
## 13   Senegal 1993   3.688879
## 14   Senegal 1994   3.772761
## 15   Senegal 1995   3.850147
## 16   Senegal 1996   3.881564
## 17   Senegal 1997   3.912023
## 18   Senegal 1998   3.912023
## 19   Senegal 1999   3.912023
## 20   Senegal 2000   4.143135
## 21   Senegal 2001   4.066174
## 22   Senegal 2002   3.982792
## 23   Senegal 2003   3.891820
## 24 Sri Lanka 1992   3.455265
## 25 Sri Lanka 1993   3.583519
## 26 Sri Lanka 1994   3.725693
## 27 Sri Lanka 1995   3.850147
## 28 Sri Lanka 1996   3.941582
## 29 Sri Lanka 1997   4.025352
## 30 Sri Lanka 1998   4.007333
## 31 Sri Lanka 1999   3.988984
```

```
## 32 Sri Lanka 2000   4.007333
## 33 Sri Lanka 2001   4.019382
## 34 Sri Lanka 2002   4.031286
## 35 Sri Lanka 2003   4.043051
```

So far, we have used common sense and qualitative judgment to infer what % of the sample is affected in different ways by the independent variable. But, there are more systematic methods we can use.

- multiple regression weights
- leverage

4 Multiple regression weights

- **Multiple regression weights** measure the extent to which each observation contributes to the effect estimate of the treatment variable. By doing so, multiple regression weights allow us to study the representativeness of our regression estimators.
- If we have a model ($Y = \alpha + \beta_1 + \beta_2 + \epsilon$), and we want to know whether different observations contribute differentially to the construction of β_1, we can calculate the multiple regression weights of β_1.
- In multiple regression models, observations are weighted differently when constructing the coefficient estimates. This happens when the control variables can explain the variance in the treatment variable better for some observations than others. The *unexplained variation* in the treatment variable is used to predict values for the outcome variable.
- When observations have multiple regression weights of zero, this means the covariates completely explain their treatment condition. When observations have multiple regression weights with higher values, this means there is unexplained variation left in the treatment condition even after controlling for other factors.
- When this happens, the effect estimate of the treatment variable generalizes to the smaller sample of observations that contributed more weight to the estimate - not to the full **nominal sample** or population of interest. And while the nominal sample may be representative of the underlying population of interest, the coefficient estimate is not.
- We can reweight the nominal sample using multiple regression weights, which can then tell us the **effective sample** that was actually used to estimate the effect.

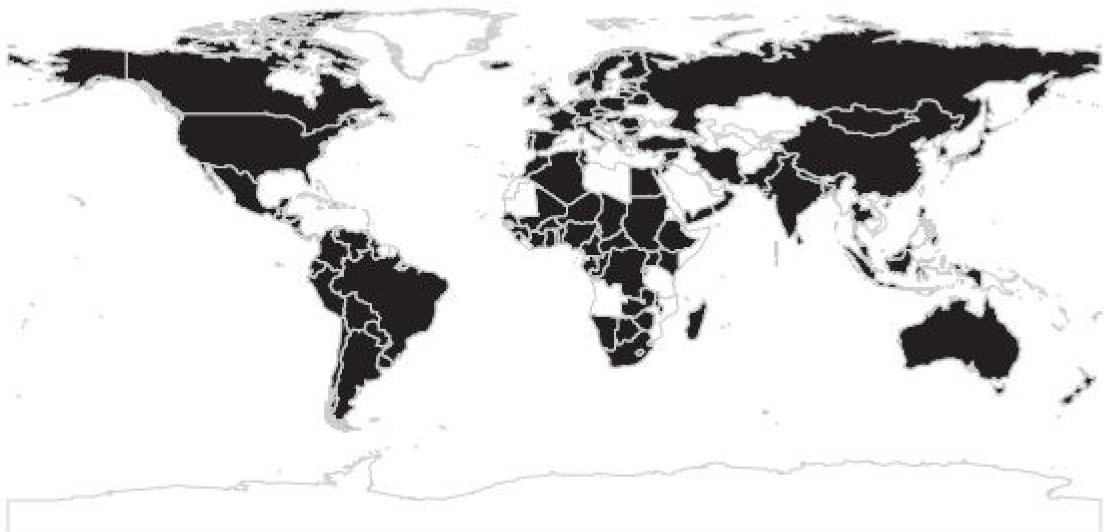
Note: The code for calculating multiple regression weights comes from Aronow and Samii (2015) "Does Regression Produce Representative Estimates of Causal Effects?" *American Journal of Political Science*. You can find their replication data here.

4.1 Example 1

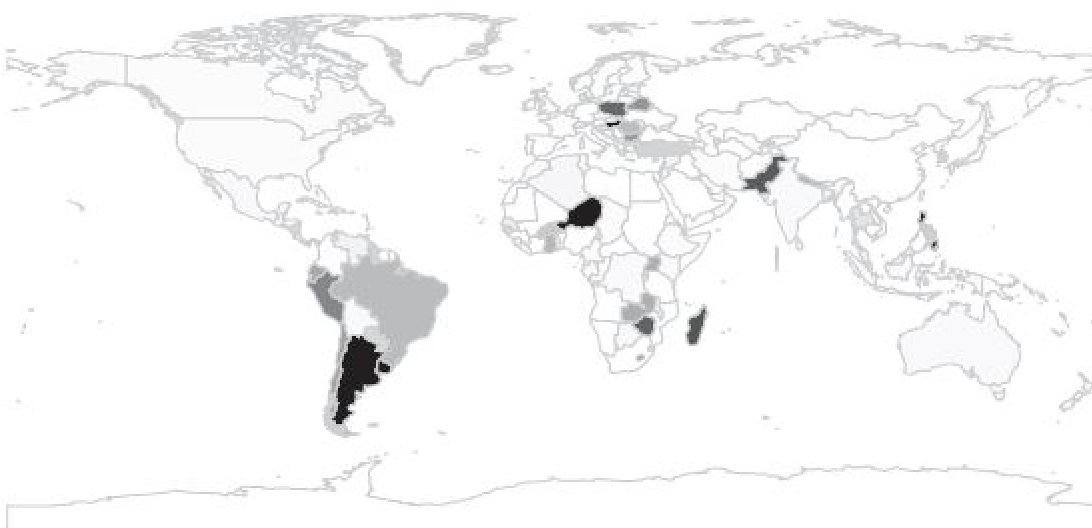
From Aronow and Samii (2015)

- The effect of **regime type** on **FDI** from Jensen (2003): "democratic political institutions are associated with higher levels of FDI inflows" (588).
- Jensen (2003) finds that a one-unit increase in the Polity score corresponds to a 0.02 increase in net FDI inflows as a percentage of GDP ($p < 0.001$).
- The control variables: lagged FDI, GDP, GDP per capita, growth, trade, budget deficit, government consumption, and country and decade fixed effects.

Nominal Sample



Effective Sample



4.2 How to calculate weights

Note: We don't want to use multiple regression weights for interaction terms because the covariates are mechanically related to the interaction term (i.e. two of the covariates are components of the interaction term). To interpret generalizability of effect estimates for interactions, we'll look at leverage in the next section.

For the sake of demonstration, let's get rid of the interaction term and focus on calculating multiple regression weights for the HRO shaming variable (`HRnc2gcnc2`).

First, fit the original linear model.

```
fit.y <- lm(Fphysint ~ HRnc2gcnc2 + lnhrfilled + normaledlnpop +
              normaledlngdp + polity2 + WARonlocation, data = df)
```

Second, fit a linear model with the treatment variable as the outcome variable.

```
fit.d <- lm(HRnc2gcnc2 ~ lnhrfilled + normaledlnpop +
              normaledlngdp + polity2 + WARonlocation, data = df)
```

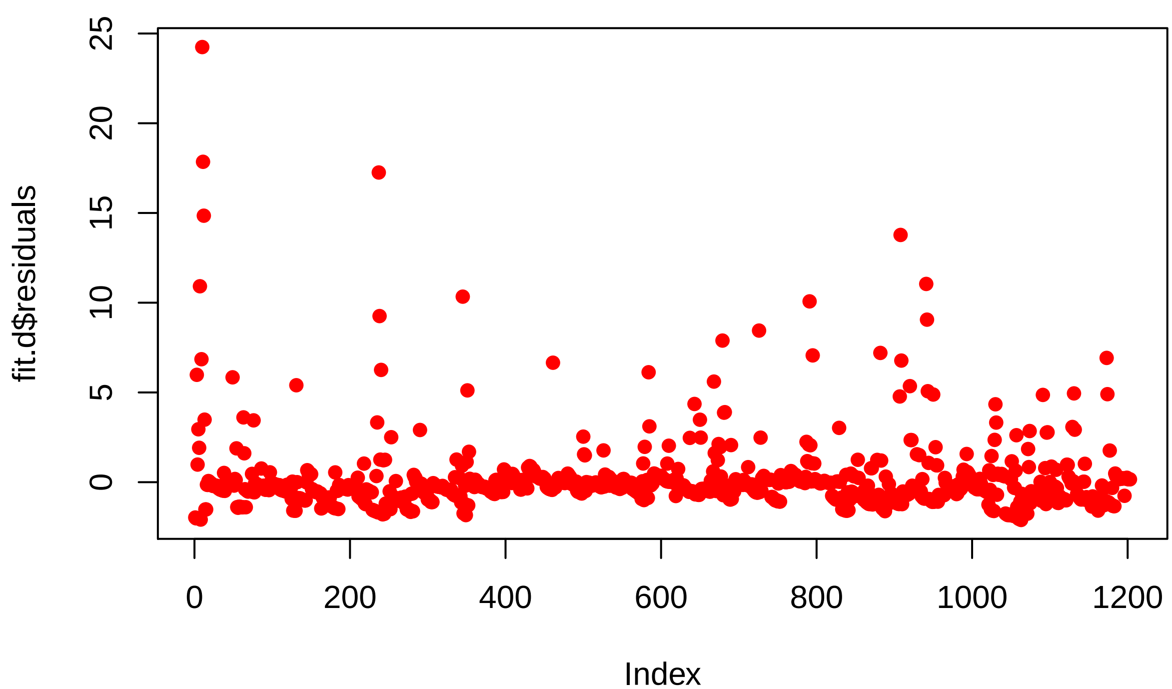
Calculate the residuals.

```
# First method
df$d.tilde <- residuals(fit.d)

# Second method
df$d.pred <- predict(fit.d)
df$d.res <- df$HRnc2gcnc2 - df$d.pred
```

Visualize the residuals.

```
plot(fit.d$residuals, pch = 16, col = "red")
```

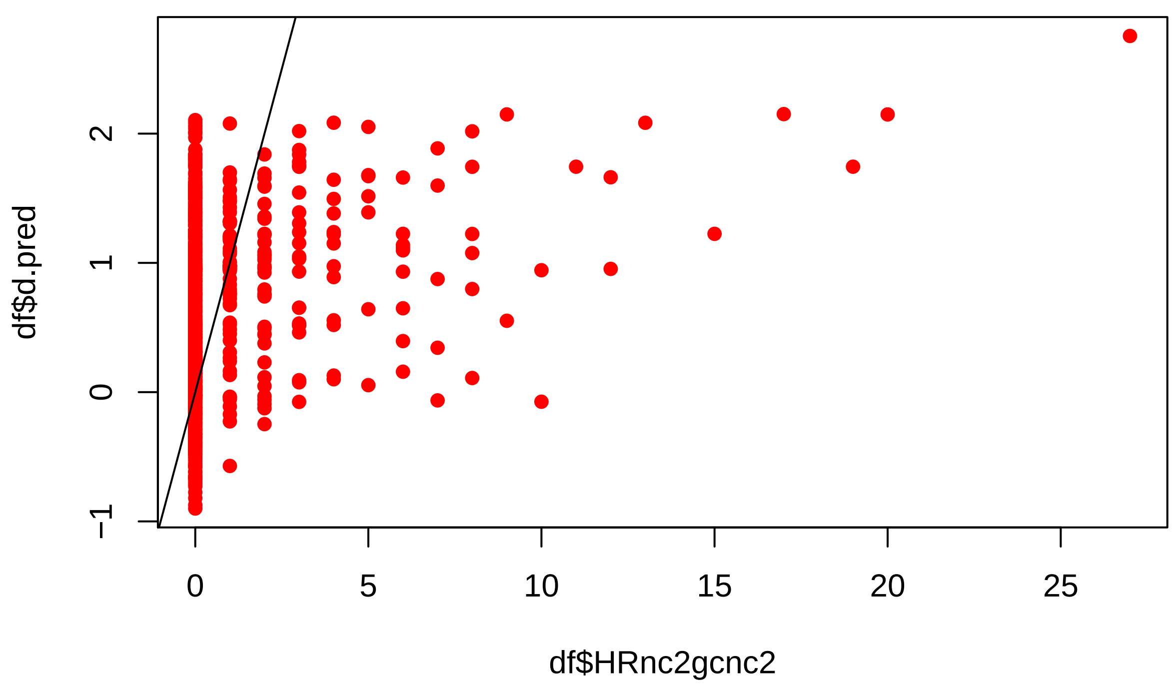


What does the graph tell us? The control variables do a relatively good job of estimating the values for `HRnc2gcnc2`, except for some observations. Notice one of the observations with a residual value near 25?

That observation is going to contribute more weight to the coefficient estimate of `HRnc2gcnc2` in the `fit.y` model.

To calculate the weights, we want to know which observations were poorly predicted by the control variables (i.e. the residuals). But we don't care if the model under- or over-estimated values for `HRnc2gcnc2`.

```
plot(df$HRnc2gcnc2, df$d.pred, pch = 16, col = "red")
abline(a = 0, b = 1)
```



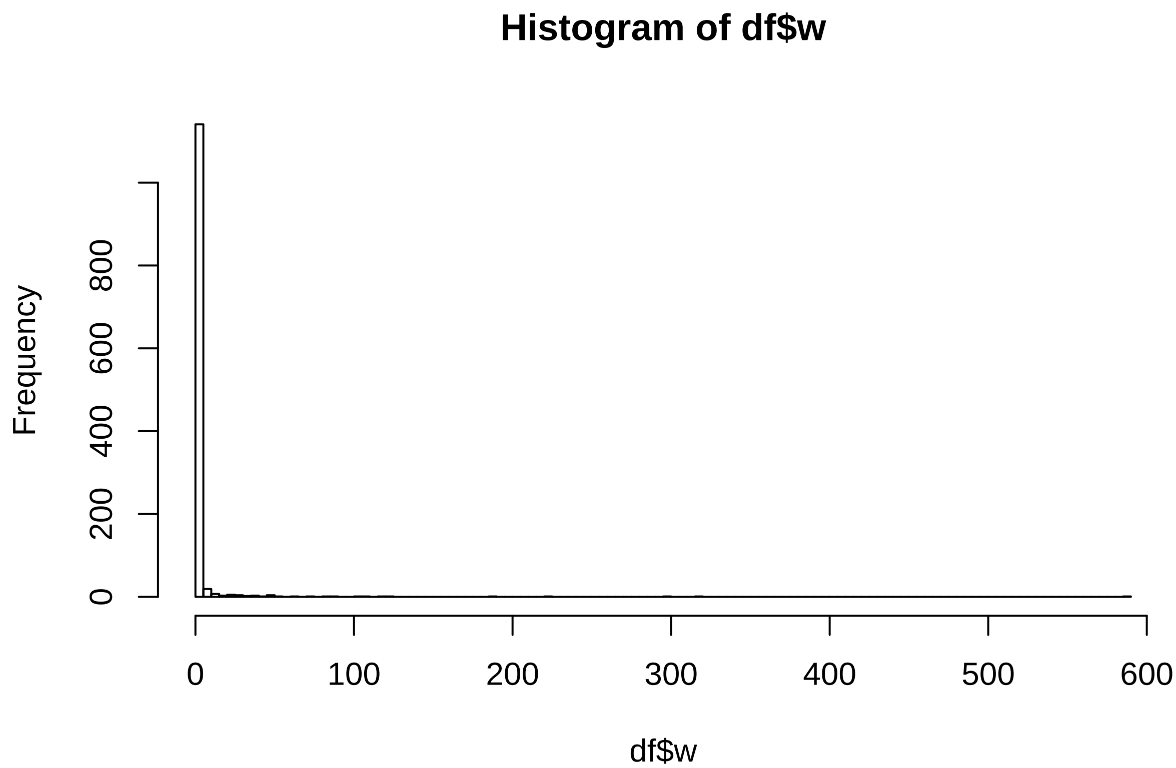
So we need to estimate the weights by squaring the residuals.

```
df$w <- df$d.tilde^2
```

In summary, we calculate multiple regression weights by squaring the residuals from the model in which the treatment variable is regressed on the control variables.

Visualize the distribution of weights.

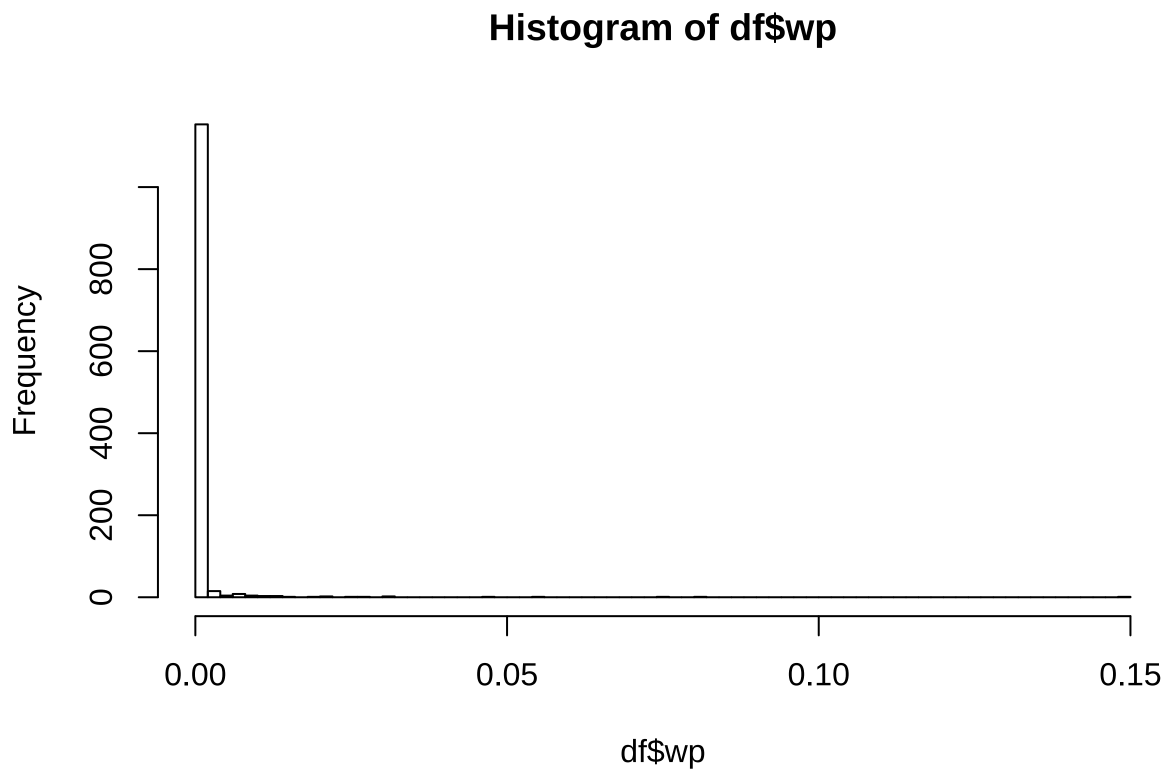
```
hist(df$w, breaks = 100)
```



High values indicate more weight; low values indicate less weight. But what does it mean when an observation has a multiple regression weight of 600? We can standardize the weights to aid interpretation by calculating the **proportion** of the weight contributed by each observation. This will tell us to what extent an observation is over- or under-represented in the construction of the coefficient estimate of `HRnc2gcnc2`. We do this by dividing an observation's multiple regression weight by the sum of all of the weights. If an observation has a value of 0.25, then we can say that it contributes 25% of the weight used to construct the estimate of `HRnc2gcnc2`.

```
df$wp <- df$w/sum(df$w)

hist(df$wp, breaks = 100)
```



Which observations contribute the most weight?

```
df %>%
  arrange(-wp) %>%
  select(NAMES_STD, year, wp) %>%
  slice(1:10)
```

```
##          NAMES_STD year         wp
## 1    United States 2001 0.14845212
## 2    United States 2002 0.08048424
## 3   United Kingdom 2000 0.07520240
## 4    United States 2003 0.05568636
## 5            Egypt 2001 0.04792509
## 6           Israel 2000 0.03081851
## 7    United States 1998 0.03009533
## 8           Russia 1995 0.02699016
## 9        Swaziland 1997 0.02563233
## 10  United Kingdom 2001 0.02163722
```

The coefficient estimate of `HRnc2gcnc2` is highly unrepresentative. One single observation - the USA in 2001 - contributes nearly *15% of the weight* used to construct the estimate of naming and shaming. The top 10 observations contribute over *50% of the weight*. Most of the estimated effect of NGO naming and shaming applies to a very small number of observations. The effective sample is much, much smaller than the nominal sample.

```
df %>%
  arrange(-wp) %>%
  select(wp) %>%
```

```r
  slice(1:10) %>%
  sum()
```

```
## [1] 0.5429238
```

What would the weight-proportion be if every observation contributed equally?

```r
1/nrow(df)
```

```
## [1] 0.0008312552
```

What proportion of the weight does each observation contribute relative to the expectation or average?

```r
df$wpx <- df$wp/(1/nrow(df))

df %>%
  arrange(-wpx) %>%
  select(NAMES_STD, year, wpx) %>%
  slice(1:10)
```

```
##         NAMES_STD year       wpx
## 1   United States 2001 178.58790
## 2   United States 2002  96.82254
## 3  United Kingdom 2000  90.46848
## 4   United States 2003  66.99069
## 5           Egypt 2001  57.65388
## 6          Israel 2000  37.07467
## 7   United States 1998  36.20468
## 8          Russia 1995  32.46916
## 9       Swaziland 1997  30.83570
## 10 United Kingdom 2001  26.02957
```

The USA in 2001 contributes over *178 times* the average weight.

We can also look at the bottom contributing countries.

```r
df %>%
  mutate(wpx = wpx*100) %>%
  group_by(NAMES_STD) %>%
  summarise(wpx = mean(wpx)) %>%
  arrange(wpx) %>%
  slice(1:10)
```

```
## # A tibble: 10 x 2
##    NAMES_STD       wpx
##    <chr>         <dbl>
##  1 Mauritius    0.0497
##  2 Madagascar   0.179
##  3 Togo         0.217
##  4 Benin        0.223
##  5 Mauritania   0.315
##  6 Botswana     0.402
##  7 Sierra Leone 0.440
##  8 Nicaragua    0.460
##  9 Guinea       0.540
## 10 Gabon        0.558
```

4.3 Exercises

1. How many of the highest-weighted observations combine to contribute 50% of the weight used to construct the coefficient estimate for HRO naming and shaming (HRnc2gcnc2)?

2. How many of the lowest-weighted observations combine to contribute 1% of the weight used to construct the effect estimate?

3. What proportion of the weight does the United States alone contribute?

```
# 1 - 9 observations or 0.75% of the sample
df %>%
  arrange(-wp) %>%
  select(NAMES_STD, year, wp) %>%
  mutate(wp_cum = cumsum(wp)*100) %>%
  slice(1:10)
```

```
##         NAMES_STD year         wp   wp_cum
## 1   United States 2001 0.14845212 14.84521
## 2   United States 2002 0.08048424 22.89364
## 3  United Kingdom 2000 0.07520240 30.41388
## 4   United States 2003 0.05568636 35.98251
## 5           Egypt 2001 0.04792509 40.77502
## 6          Israel 2000 0.03081851 43.85687
## 7   United States 1998 0.03009533 46.86640
## 8          Russia 1995 0.02699016 49.56542
## 9       Swaziland 1997 0.02563233 52.12865
## 10 United Kingdom 2001 0.02163722 54.29238
```

```
# 2 - 613 observations or 51% of the sample
df2 <- df %>%
  arrange(wp) %>%
  select(NAMES_STD, year, wp) %>%
  mutate(wp_cum = cumsum(wp)*100)

# 3 - 34% of the weight
df %>%
  group_by(NAMES_STD) %>%
  summarise(twp = sum(wp)) %>%
  filter(NAMES_STD == "United States")
```

```
## # A tibble: 1 x 2
##   NAMES_STD       twp
##   <chr>         <dbl>
## 1 United States 0.342
```

5 Leverage

Y Physical Integrity Rights = α + β_1 HRO Shaming + β_2 Domestic HRO Presence + β_3 (HRO Shaming X Domestic HRO Presence) + ϵ

- We can't calculate the multiple regression weight for interaction terms (β_3) using the method above because the coefficient estimate of the interaction term is mechanically related to the control variables, β_1 and β_2 (i.e. two of the control variables are components of the interaction term itself). Instead, we can combine our understanding of **leverage** and **multiple regression weights** to calculate an observation's contribution to both β_1 and β_3.

5.1 What is leverage?

- Leverage is similar to multiple regression weights, but they're not quite the same.
- Whereas multiple regression weights measure the contribution of an observation to the construction of a coefficient estimate, the leverage measures the effect of dropping an observation on the overall coefficient vector.
- So think of multiple regression weights as measuring an observation's contribution to one coefficient estimate. Think of leverage as measuring an observation's contribution to the entire model.

5.2 How to calculate leverage

If we want to know an observation's contribution to the overall model, we calculate leverage.

```
# Fit the original model with the interaction term.
model <- lm(Fphysint ~ HRnc2gcnc2 + lnhrfilled + shame_HROs +
              normaledlnpop + normaledlngdp + polity2 +
              WARonlocation, data = df)

# Calculate the leverage using hatvalues().
df$lev <- hatvalues(model)
```

Leverage is bound between 0 and 1, with larger numbers indicating higher leverage. Which observations have the highest leverage?

```
df %>%
  select(NAMES_STD, year, lev) %>%
  arrange(-lev) %>%
  slice(1:10)
```

```
##         NAMES_STD year        lev
## 1   United States 2001 0.28193650
## 2  United Kingdom 2000 0.15698547
## 3       Swaziland 1997 0.14725558
## 4   United States 2002 0.14697412
## 5   United States 2003 0.10397729
## 6          Russia 1995 0.10082759
## 7       Swaziland 2001 0.07139530
## 8            Iran 2001 0.06553238
## 9           Egypt 2001 0.05854218
## 10 United Kingdom 2001 0.05043551
```

Another way to calculate leverage, which will come in handy later:

```
# First, assign the values from the covariates to a matrix, X.
X <- cbind(1, df$HRnc2gcnc2, df$lnhrfilled, df$shame_HROs,
      df$normaledlnpop, df$normaledlngdp, df$polity2,
      df$WARonlocation)
# Second, calculate the hat matrix, H.
H <- X %*% solve(t(X) %*% X) %*% t(X)
# Third, extract the diagonal elements of the hat matrix, which tell us the leverage.
df$lev2 <- diag(H)

# We can see that df$lev and df$lev2 are the same.
df %>%
  select(lev, lev2) %>%
  slice(1:10)
```

```
##            lev        lev2
```

```
## 1  0.009996342 0.009996342
## 2  0.009879756 0.009879756
## 3  0.019327609 0.019327609
## 4  0.009191435 0.009191435
## 5  0.011862785 0.011862785
## 6  0.010609067 0.010609067
## 7  0.050402295 0.050402295
## 8  0.010484855 0.010484855
## 9  0.032023985 0.032023985
## 10 0.281936500 0.281936500
```

5.3 How to use leverage to calculate multiple regression weights.

But, for the purpose of this exercise, we're not interested in evaluating an observation's contribution to the entire model. Instead, we want to know how much an observation contributes to the conditional effect of `HRnc2gcnc2`. To accomplish this, we can combine our understanding of leverage and multiple regression weights. You'll see below that the leverage of the residuals is equivalent to the multiple regression weight.

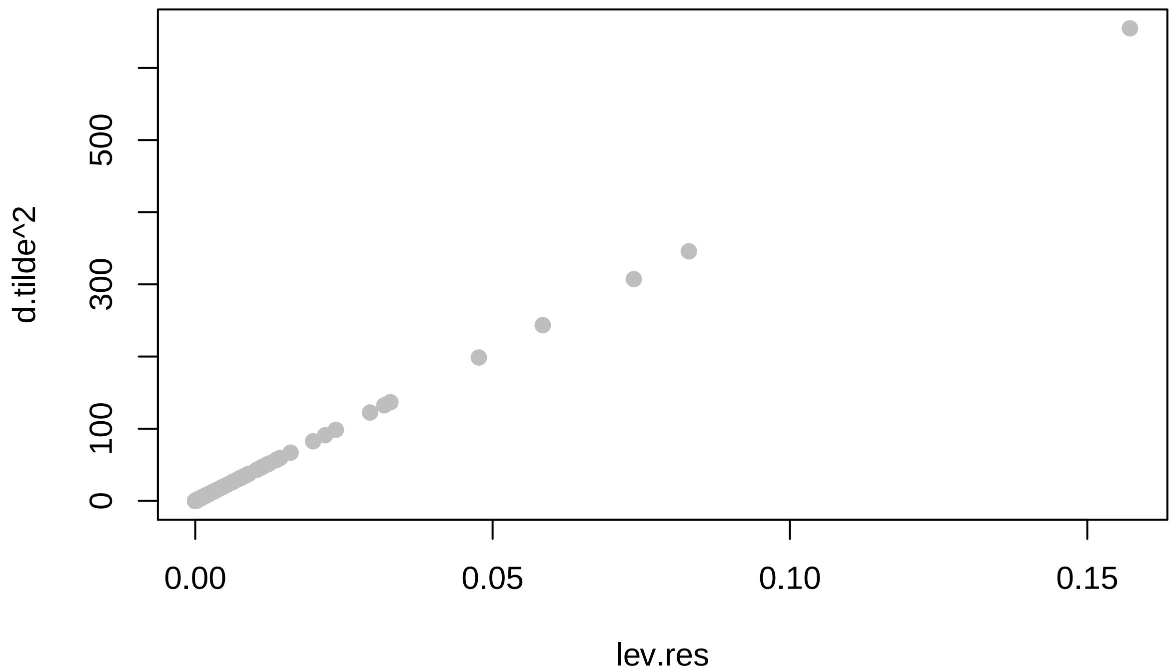
To illustrate how this will work, let's look at a toy example.

```r
# Based on code developed by Cyrus Samii, NYU.

# Fit a linear model.
fit.y <- lm(Fphysint ~ HRnc2gcnc2 + lnhrfilled, data = df)
# Fit a linear model with the treatment variable as the outcome variable.
fit.d <- lm(HRnc2gcnc2 ~ lnhrfilled, data = df)
# Extract the residuals.
d.tilde <- residuals(fit.d)
# Assign the residuals to a matrix.
X.res <- as.matrix(d.tilde)
# Extract the diagonal elements of the hat matrix.
lev.res <- diag(X.res%*%solve(t(X.res)%*%X.res)%*%t(X.res))
```

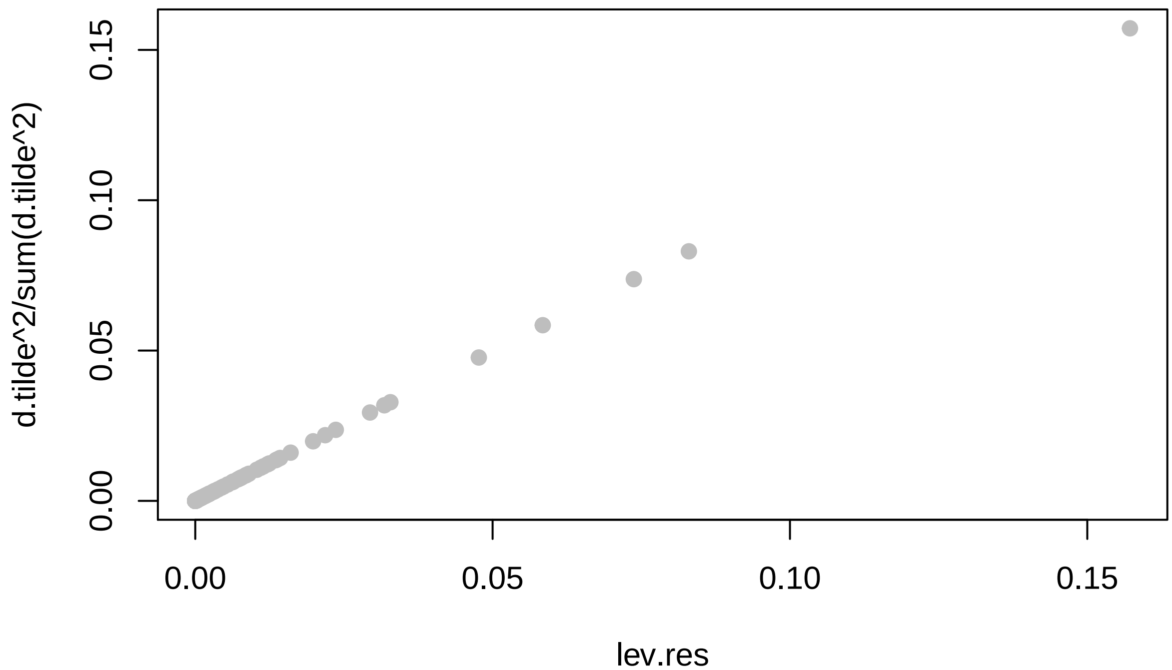
The leverage of the residuals from fit.d are equivalent to the multiple regression weights. They measure the same thing.

```r
plot(lev.res, d.tilde^2, pch = 19, col = "grey")
```



But they don't look like they're measuring the same thing even though they are perfectly correlated. The y-axis and x-axis have different values. Remember from earlier when we standardized the multiple regression weights by measuring the weight-proportion (i.e. the proportion of the weight contributed by each observation)? It turns out that *is* the leverage for the residual.

```
plot(lev.res, d.tilde^2/sum(d.tilde^2), pch = 19, col = "grey")
```



Pretty cool, huh?

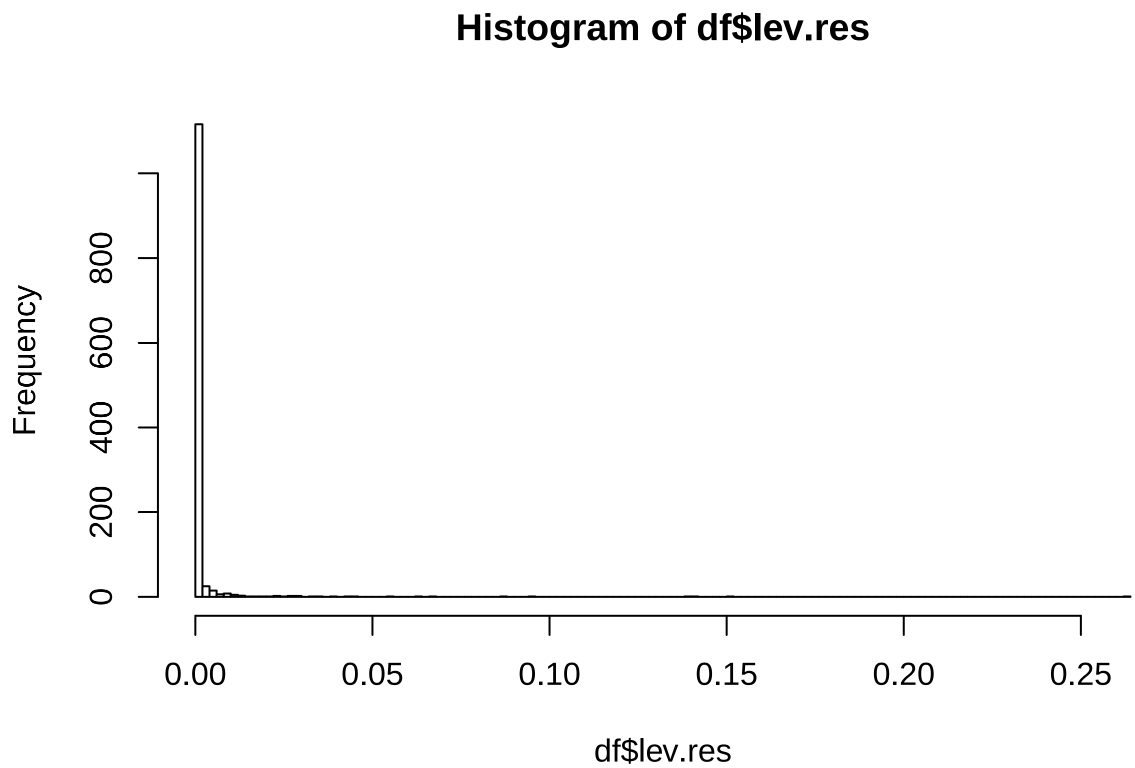
We can use the concept of measuring the leverage for the residuals of one regression to calculate the leverage for the residuals of multiple regressions. We'll have to do this in order to calculate the multiple regression weights for two coefficient estimates.

```
# y = d + z + dz + X
fit.y <- lm(Fphysint ~ HRnc2gcnc2 + lnhrfilled + shame_HROs +
              normaledlnpop + normaledlngdp + polity2 +
              WARonlocation, data = df)
fit.d.xz <- lm(HRnc2gcnc2 ~ lnhrfilled + normaledlnpop +
              normaledlngdp + polity2 + WARonlocation, data = df)
fit.dz.xz <- lm(shame_HROs ~ lnhrfilled + normaledlnpop +
              normaledlngdp + polity2 + WARonlocation, data = df)

d.tilde <- residuals(fit.d.xz)
dz.tilde <- residuals(fit.dz.xz)

X.res <- cbind(1, d.tilde, dz.tilde)
df$lev.res <- diag(X.res%*%solve(t(X.res)%*%X.res)%*%t(X.res))

hist(df$lev.res, breaks = 100)
```




```
df %>%
  select(NAMES_STD, year, lev.res) %>%
  arrange(-lev.res) %>%
  slice(1:10)
```

```
##         NAMES_STD year    lev.res
## 1   United States 2001 0.26236887
## 2  United Kingdom 2000 0.15120083
## 3       Swaziland 1997 0.14199322
## 4   United States 2002 0.13850031
## 5   United States 2003 0.09544352
## 6          Russia 1995 0.08650674
## 7       Swaziland 2001 0.06636383
## 8            Iran 2001 0.06224288
## 9           Egypt 2001 0.05445794
## 10 United Kingdom 2001 0.04468254
```

5.4 Exercises

1. What proportion of the weight does the United States contribute to the conditional effect of HRO shaming?
2. How many countries are over-represented in the effective sample?
3. What happens to the interaction term when we omit the United States?

```
# 1
df %>%
```

```r
  filter(NAMES_STD == "United States") %>%
  select(lev.res) %>%
  sum()
```

```
## [1] 0.5868192
```

```r
# 2
df %>%
  mutate(wpx = lev.res/(1/nrow(df)),
         wp_high = ifelse(wpx > 2, "over", "under")
         ) %>%
  select(wp_high) %>%
  table()
```

```
## .
##  over under
##   119  1084
```

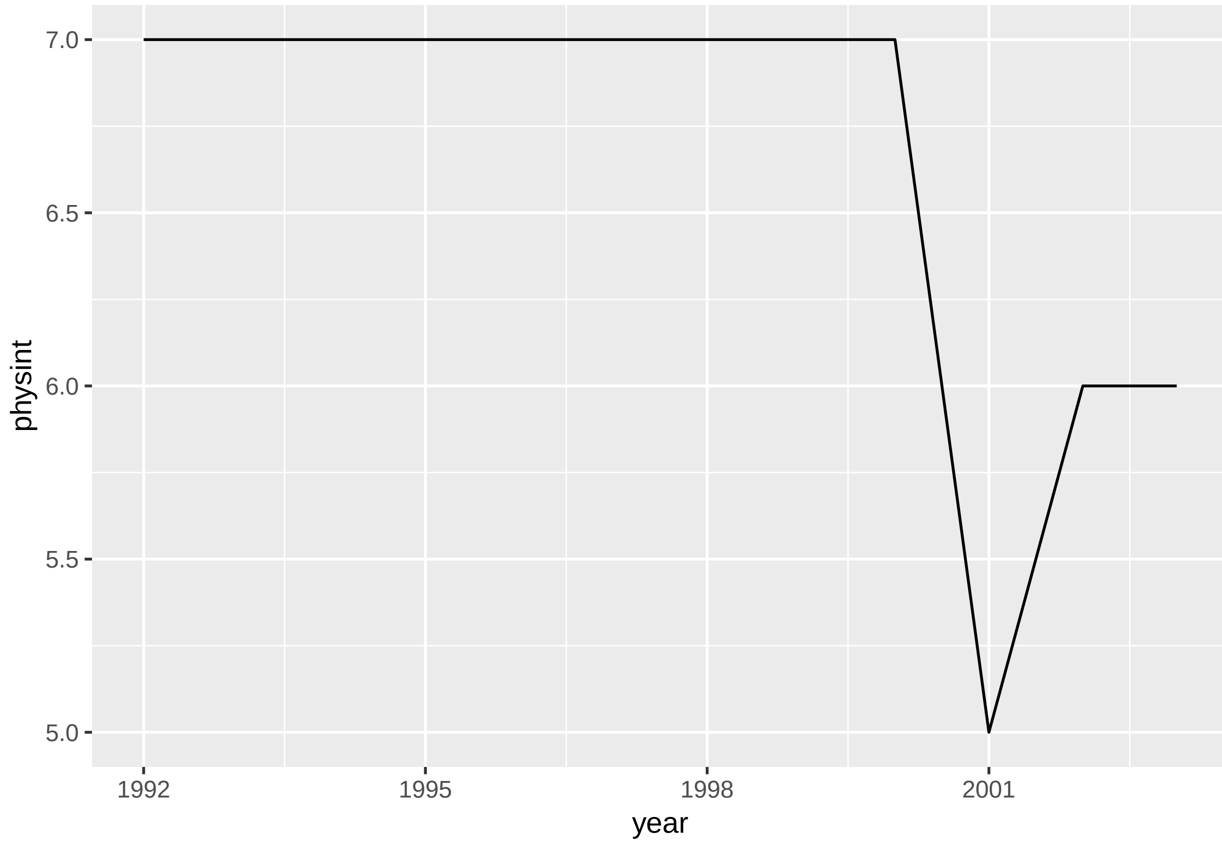
```r
# 3
model2 <- lm(Fphysint ~ HRnc2gcnc2 + lnhrfilled + shame_HROs +
              normaledlnpop + normaledlngdp + polity2 +
              WARonlocation, data = subset(df, NAMES_STD != "United States"))
summary(model2)
```

```
##
## Call:
## lm(formula = Fphysint ~ HRnc2gcnc2 + lnhrfilled + shame_HROs +
##     normaledlnpop + normaledlngdp + polity2 + WARonlocation,
##     data = subset(df, NAMES_STD != "United States"))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.3644 -0.9860  0.0964  1.0859  5.1596
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.569676   0.331347  13.791   <2e-16 ***
## HRnc2gcnc2    -0.294215   0.198725  -1.481    0.139
## lnhrfilled     0.106469   0.103465   1.029    0.304
## shame_HROs     0.062436   0.050463   1.237    0.216
## normaledlnpop -0.577908   0.037892 -15.251   <2e-16 ***
## normaledlngdp  0.362351   0.037103   9.766   <2e-16 ***
## polity2        0.068496   0.008102   8.454   <2e-16 ***
## WARonlocation -2.023082   0.188475 -10.734   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.577 on 1183 degrees of freedom
## Multiple R-squared:  0.429,  Adjusted R-squared:  0.4256
## F-statistic:   127 on 7 and 1183 DF,  p-value: < 2.2e-16
```

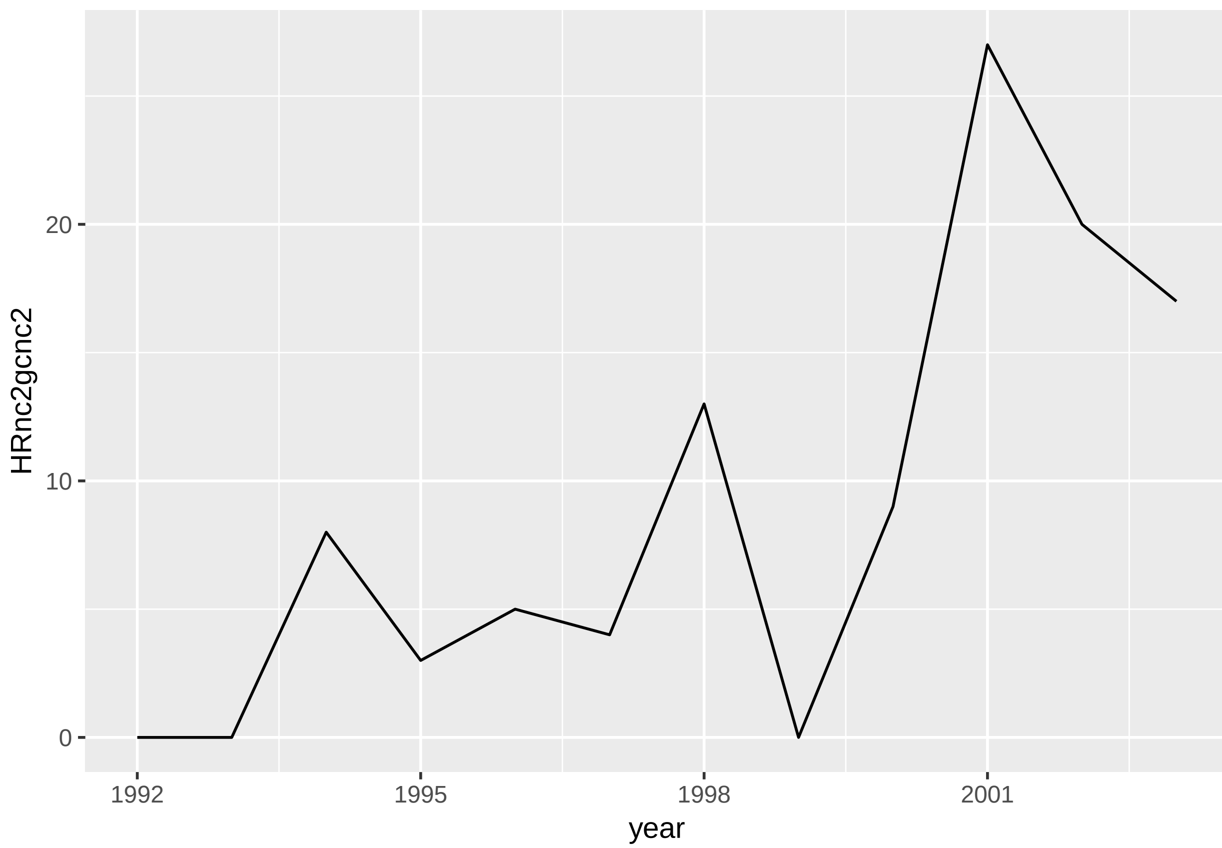
4. Discussion question: are we facing an outlier problem or an external validity problem?

```r
df %>%
  filter(NAMES_STD == "United States") %>%
  ggplot(aes(x = year, y = physint)) +
```

```r
  geom_line()
```




```r
df %>%
  filter(NAMES_STD == "United States") %>%
  ggplot(aes(x = year, y = HRnc2gcnc2)) +
  geom_line()
```



6 Summary

- **Multiple regression weights:** calculate these if you want to know an observation's contribution to the coefficient estimate of a treatment variable.
- **Leverage:** calculate this if you want to know an observation's contribution to the model.
- **Leverage for the residuals:** calculate this if you want to know an observation's contribution to multiple coefficient estimates.

7 Resources

7.1 Marginal effects in R

- R package, plotMElm, by Christopher Gandrud
- Also check out his book, Reproducible Research with R and RStudio

7.2 Multiple regression weights

- Aronow, Peter M. and Cyrus Samii (2015) "Does Regression Produce Representative Estimates of Causal Effects?" *American Journal of Political Science* 60(1): 250-267
- Replication data for Aronow and Samii