Does personal violence have an effect on attitudes towards peace in Darfur?: Reproducing results from Cinelli and Hazlett [2020]

Anuradha Ramachandran

Section 1: Introduction

This report is reproducing some results from Cinelli and Hazlett(2020) and Hazelett (2012) paper. In this data analysis project, a dataset on attitude of Darfurian refugees in eastern Chad is considered. The objective of this project is to estimate the effect of violence on individuals and their attitude towards seeking peace. The treatment variable is "directlyharmed (D_i)" which indicates whether an individual was physically harmed during attacks on villages in Darfur between 2003 and 2004.
The response variable of interest is "peacefactor" that measures the pro-peace attitudes.
Other features include age, female, past_voted, farmer_dar,herder_dar.
This dataset is found in the sensemakr package in R. It consists of 1275 observations with 15 features.
In this project let:

- Y_i = peacefactor (outcome)
- D_i = directlyharmed (binary indicator of "treatment")
- F_i = female (binary indicator)
- V_i = village (a matrix of binary village indicators)
- X_i = a matrix with a constant and the covariates herder dar, farmer dar, age, and past voted.

Section 2: Reasearch question

In 2003 and 2004, the Darfurian government orchestrated a horrific campaign of violence against civilians, killing an estimated two hundred thousand people.This project asks whether, on average, being directly injured or maimed in this episode (D_i) changed individuals attitudes towards peace (Y_i). Did exposure to violence make individuals more likely to feel "vengeful" and unwilling to make peace with those who perpetrated this violence, or, more likely to feel "weary," and motivated to see it end by making peace? More specifically, suppose we are interested in the average treatment effect of D_i on Y_i (ATE).
Writing ATE in terms of counterfactuals:
D_i = "directlyharmed" binary variable. Let $D_i = 0$ denote individuals not harmed and $D_i = 1$ denote indivials that were harmed.
Potential outcomes Y_i be :
(i) when $D_i = d_i = 0 : Y_i(d = 0)$ represents the potential peacefactor if by intervention the individual "i" is not harmed. (Control)
(i) when $D_i = d_i = 1 : Y_i(d = 1)$ represents the potential peacefactor if by intervention the individual "i" is harmed. (Treatement)
Individual treatment effect $\tau_i = Y_i(1) - Y_i(0)$
Average treatment effect $E[\tau_i] = E[Y_i(1) - Y_i(0)] = E[Y_i(1)] - E[Y_i(0)]$

Section 3: Identification of ATE

The purpose of these attacks was to punish civilians from ethnic groups presumed to support the opposition, and to kill or drive these groups out so as to reduce support for the opposition. Violence against civilians included aerial bombardments by the government as well as ground assaults by the Janjaweed, a pro-government militia. Now suppose a researcher argues that, while some villages were more or less intensively attacked, within village violence was largely indiscriminate. The bombings could not be finely targeted owing to their crudeness, and there were not many reasons to target them. Similarly, the Janjaweed had no reason to target certain individuals rather than others, and no information with which to do so—with one major exception: women were targeted and often subjected to sexual violence.
Given these considerations, one may argue that adjusting for village V_i and female F_i is sufficient for control of confounding.
Under this consideration, it can be assumed that there are no unobserved counfounding variable of D_i and Y_i beyond V_i and F_i, from ignorability condition (conditional independence assumption) it can be stated that $\{Y_i(1), Y_i(0)\} \perp D_i|(V_i, F_i)$
Under the assumption of conditional ignorability and cosistency of potential outcomes, the ATE can be identified from the data non-paramterically from the data as:
$E[\tau_i] = E[Y_i(1)] - E[Y_i(0)] = E[E[Y_i|D_i = 1, V_i, F_i]] - E[E[Y_i|D_i = 0, V_i, F_i]]$
The other covariates such as age, past_voted, farmer_dar,herder_dar can be denoated as X_i. Since V_i and F_i are assumed to be sufficient control for confounding, X_i is not necessary to be included in the model, however X_i is included to improve the precision of the estimated coefficients.
Mathematically, we can state that village, female and remaining covariates are sufficient control for confounding by conditional ignorability condition:
$\{(Y_i(1) - Y_i(0))\} \perp D_i|V_i, F_i, X_i$

Section 4: Estimation of ATE

Section 4.1: Model 1
Assume that the conditional expectation function (CEF) be approximated with a linear function of the covariates:
Let model1 be $= E[Y_i|D_i, F_i, V_i, X_i] = \tau_1 D_i + \beta_{1f} F_i + V_i^T \beta_{1v} + X_i^T \beta_{1x}$
From section 3 , $ATE = E[\tau_i] = E[E[Y_i|D_i = 1, F_i, V_i, X_i] - E[Y_i|D_i = 0, F_i, V_i, X_i]]$
Substituing CEF in ATE equation yields,
$ATE = E[\tau_i] = E[\tau_1 + \beta_{1f} F_i + V_i^T \beta_{1v} + X_i^T \beta_{1x} - (\beta_{1f} F_i + V_i^T \beta_{1v} + X_i^T \beta_{1x})] = \tau_1$
The estimated ATE from regressing Y_i on D_i, V_i, F_i, X_i is 0.0973. On average it can be seen that there's a 10% increase in attitudes towards peace among those individuals who were directly harmed than those who were not.
The ATE can also be estimated by using Frisch–Waugh–Lovell (FWL) theorem, by regressing the residuals from regressing Y_i on V_i, F_i, X_i on residuals from regressing D_i on V_i, F_i, X_i
ATE identified from FWL theorem is identical to the estimated ATE from linear regression = 9.7 Hence we can conclude that there is a positive causal relationship between directlyharmed and peacefactor. The 95% confidence interval for estimated ATE using classical standard errors and robust standard errors found using non parametric bootstrap method using 1000 samples are similar.
95% confidence interval(CI) for "directlyharmed" = 95% CI for ATE using classical model is (0.052, 0.143)
95% confidence interval(CI) for "directlyharmed" = 95% CI for ATE using robust model is (0.052, 0.142)
Note: Code for all these estimates can be found in the appendix.

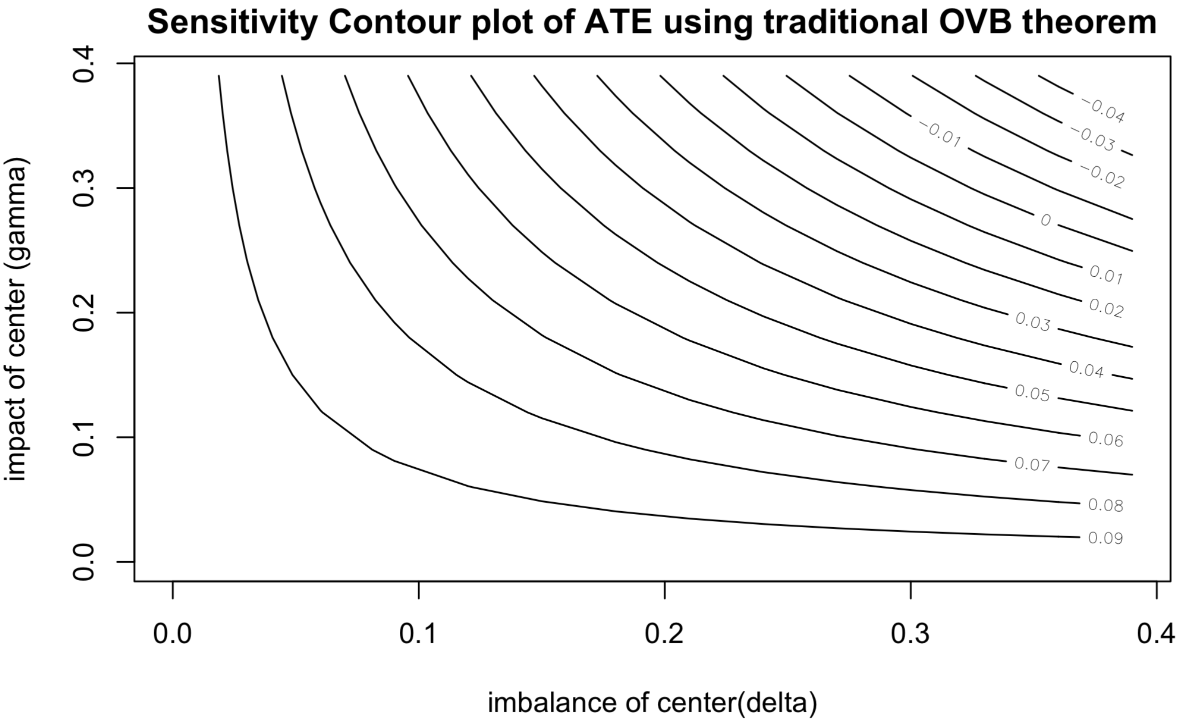
Section 4.2: Model 2
Consider an interaction term between D_i and F_i, as in:
Let model2 $= E[Y_i|D_i, F_i, V_i, X_i] = \tau_2 D_i + \beta_{2f} F_i + \beta_{2fd} D_i F_i + V_i^T \beta_{2v} + X_i^T \beta_{2x}$
From section 3, $ATE = E[\tau_i] = E[E[Y_i|D_i = 1, F_i, V_i, X_i] - E[Y_i|D_i = 0, F_i, V_i, X_i]]$
Substituing CEF in ATE equation yields,

$ATE = E[\tau_i] = E[\tau_2 + \beta_{2f}F_i + \beta_{2fd}F_i + V_i^T\beta_{2v} + X_i^T\beta_{2x} - (\beta_{2f}F_i + V_i^T\beta_{2v} + X_i^T\beta_{2x})] = \tau_2 + \beta_{2fd}E[F_i]$
Estimated ATE using plug in principle and regression coefficients = 0.0975
95% confidence interval for ATE using non parametric bootstrap is (0.056,0.139)
In this model too, on average it can be seen that there's a 10% increase in attitudes towards peace among those individuals who were directly harmed than those who were not.
A positive causal relationship between directlyharmed and peacefactor is identified.
Hence, including the interaction term of "direcltyharmed" and "female" did not alter the average treatment effect of "directlyharmed" on "peacefactor".
Note: Code for all these estimates can be found in the appendix.

Section 5: Sensitivity analysis

Section 5.1: Traditional Omitted Variable Bias
Consider model 1 to estimate ATE in this section
The causal interpretation of the previous estimates requires the assumption of no un- observed confounders. While this may be supported by the claim that there was no targeting of violence within village-gender strata, not all investigators may agree with this account. For example, a reasonable argument could be made that, although the bombing was crude, bombs were still more likely to hit the center of the village, and those in the center would also likely hold different attitudes towards peace.
Let C_i denote the unobserved binary confounder center. Thus, one could argue that unconfoundedness would hold only conditional on F_i, V_i and C_i, and, to estimate (approximate) the ATE, we should have instead run the regression (Model 3):
model 3: $E[Y_i|D_i, F_i, V_i, X_i, C_i] = \tau D_i + \beta_f F_i + V_i^T\beta_v + X_i^T\beta_x + \gamma C_i$
Under this assumption, the unconfoundedness assumption is modified as
$\{Y_i(1) - Y_i(0)\} \perp D_i|F_i, V_i, X_i, C_i$
Including the hypothetical confounder C_i in the regession model might impact the estimation of the coefficient of the treatment variable D_i.
Depending on the bias (negative or positive), inclusion of C_i might increase or decrease the estimated coefficient of the treatment variable D_i.
If the strength of the hypothetical confounder "C_i" is very high, it could also make the effect of "directly-harmed" on "peacefactor" not statistically significant.
Let
model1: $E[Y_i|D_i, F_i, V_i, X_i] = \tau_1 D_i + \beta_{1f}F_i + V_i^T\beta_{1v} + X_i^T\beta_{1x}$
model3: $E[Y_i|D_i, F_i, V_i, X_i] = \tau D_i + \beta_f F_i + V_i^T\beta_v + X_i^T\beta_x + \gamma C_i$
From OVB theorem,
$\tau_1 = \tau + \gamma\delta$
where $\delta = \frac{cov(D_i^{\perp W_i}, C_i^{\perp W_i})}{var(D_i^{\perp W_i})}$
$\implies, \tau_1 - \tau = \gamma\delta = bias$
Interpretation of each term:
τ_1: coefficient of treatment "directlyharmed" from model 1 regressing Y_i on D_i, V_i, F_i, X_i. Assuming only V_i, F_i, X_i, are good control for confounding.
τ: coefficient of treatment "directlyharmed" from model 3 regressing Y_i on D_i, V_i, F_i, X_i, C_i. Assuming V_i, F_i, X_i, C_i are good control for confounding.
$\tau_1 - \tau$: Bias in the estimation of coefficient of treatment "directlyharmed" when counfounder "center(C_i)" is omitted in the regression model1.
γ: Within the class of other covariates, average difference in peacefactor among those individuals living in center (C_i) compared to those not living at the center of the village (Cinelli 2020).
In other words, it measures the predictive effect of C_i on Y_i.
δ: Represents the conditional imbalance of the "center (C_i)" with respect to D_i.

Sensitivity contour plot of bias induced by the unobserved confounder C_i on ATE is shown below. The contour lines represent the adjusted ATE obtained from traditional OVB theorem at hypothesized values of the impact and imbalance parameters. As seen from the plot the imbalance of confounder "center" has to be > 0.3 and impact of center variable has to be > 0.3 in order to bring down the estimated effect of D_i to



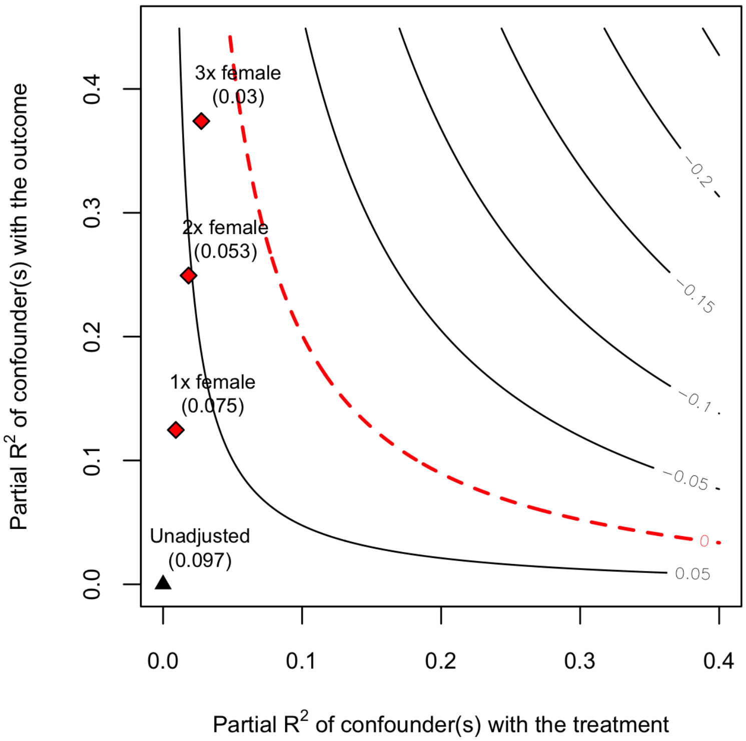
zero.

Suppose a researcher claims that the covariate Center could have a conditional impact of $\gamma = 0.2$ in attitudes towards peace, and conditional imbalance of $\delta = 0.2$, meaning that on average, those who were physically injured were also 20 percentage points more likely to live in the center of the village.
The estimate of τ using regression coefficient τ_1 of "directlyharmed" from model 1 and bias of 0.04 is 0.0573.
The estimate of τ using plug in principle with a bias of 0.04 is 0.0573.
The 95% confidence interval for τ using non parametric bootstrap is (0.01522134,0.09941030).
The 95% confidence interval for τ_1 in model1 is (0.05522134,0.13941030), the adjusted estimate τ is within this range. Hence a bias of 0.04 does not change the conclusions of the study.

Section 5.2: Partial R^2 parameterization

Beyond center, one may also argue that the Janjaweed may have observed signals that indicate, for example, the wealth of individuals. Or, perhaps, that an individual's prior political attitudes could have led them to take actions that exposed them to greater risks. To complicate things, all these factors could interact with each other or otherwise have non-linear effects, all acting as confounders. Let Zi denote a vector of all these covariates, including nonlinear transformations if necessary. The traditional OVB theorem falls short in able to account for multiple unobserved confounders. However partial R^2 parameterization is useful in assessing the sensitivity of an estimate to any number or even all confounders acting together, possibly non-linearly.
Omitted variable bias in terms of the partial R^2 of Z_i with Y_i and D_i can be written as: $\hat{bias} = \sqrt{\frac{R^2_{Y\sim Z|D,W} * R^2_{D\sim Z|W}}{1-R^2_{D\sim Z|W}}} \frac{sd(Y^{\perp W,D})}{sd(D^{\perp W})}$
where $W = [V, X, F]$
$\frac{sd(Y^{\perp W,D})}{sd(D^{\perp W})}$ can be estimated from the data.
$\sqrt{\frac{R^2_{Y\sim Z|D,W} * R^2_{D\sim Z|W}}{1-R^2_{D\sim Z|W}}}$ is limited by hypothesis regarding strength of confounding.

Sensitivity contour plot of bias induced by the unobserved confounder C_i on ATE is shown below. The contour lines represent the adjusted estimate that would be obtained for an unobserved confounder using partial R^2 parameterization. The strength of the unobserved confounder has to be greater than three times the strength of female F_i variable in order to bring down the estimated effect of D_i to zero.



Now suppose that Z_i can explain, collectively, at most 12% of the residual variation of the outcome, and 1% of the residual variation of the treatment. Assuming that the direction of the bias reduces the magnitude of τ, the point estimate of τ using plug in principle and "adjusted estimate" function in sensemakr package is 0.075.

95% confidence interval for τ using bootstrap method is (0.0325,0.117).

Since the robustness value at 5% significance is 7.6%. This means that if Z_i can explain 7.6% of the residual variation of both treatment and outcome then it can make the estimate not statistically significant at 5% significance level.

However, with the hypothesis that $R^2_{Y\ Z|D,W} = 12\%$ and $R^2_{D\ Z|W} = 1\%$, means that Z_i cannot explain 7.6% of the residual variation of both treatment and outcome.

Hence the strength of unobserved confounders Z_i is not strong enough to change the conclusion regarding the sign of the treatment effect.

Section 6 Appendix

```r
# loading data
data("darfur")
```

```r
# ATE estimate using BLP
ols = lm(peacefactor ~ directlyharmed + female + village + age +
    farmer_dar + herder_dar + pastvoted + hhsize_darfur, data = darfur)
est.ATE = coef(ols)["directlyharmed"]
est.ATE
```

Section 6.1: Code for Section 4.1

```
## directlyharmed
##     0.09731582
```

```r
# residual from regressing Y_i on F_i, V_i and X_i
myfvx = lm(peacefactor ~ female + village + age + farmer_dar +
    herder_dar + pastvoted + hhsize_darfur, data = darfur)
tilde.yi = resid(myfvx)

# residual from regressing D_i on F_i, V_i and X_i
mdfvx = lm(directlyharmed ~ female + village + age + farmer_dar +
    herder_dar + pastvoted + hhsize_darfur, data = darfur)
tilde.di = resid(mdfvx)

# using FWL theorem
tau1 = cov(tilde.yi, tilde.di)/var(tilde.di)
tau1
```

```
## [1] 0.09731582
```

```r
# regressing tilde.yi on tilde.di
residreg = lm(tilde.yi ~ tilde.di)
residreg
```

```
##
## Call:
## lm(formula = tilde.yi ~ tilde.di)
##
## Coefficients:
## (Intercept)     tilde.di
##  -7.889e-18    9.732e-02
```

```r
# 95% confidence interval using non parametric bootstrap
# and regression coefficient for ATE
set.seed(123)
B = 1000
n = nrow(darfur$peacefactor)
ate_boot = rep(NA, B)
for (i in 1:B) {
    idx_boot = sample(1:n, size = n, replace = T)
    data_boot = darfur[idx_boot, ]
    ols_boot = lm(peacefactor ~ directlyharmed + female + village +
        age + farmer_dar + herder_dar + pastvoted + hhsize_darfur,
        data = data_boot)
    ate_boot[i] = coef(ols_boot)["directlyharmed"]
}

se.boot = sd(ate_boot)
z_crit = qnorm(0.95)
low = est.ATE - z_crit * se.boot
up = est.ATE + z_crit * se.boot
# 95% CI using bootstrap method
c(low, up)
```

```
## directlyharmed directlyharmed
##     0.05522134     0.13941030
```

```r
# 95% confidence interval using non parametric bootstrap
# and plug in estimate for ATE
set.seed(123)
B = 1000
n = nrow(darfur$peacefactor)
ate_boot1 = rep(NA, B)
for (i in 1:B) {
    idx_boot = sample(1:n, size = n, replace = T)
    data_boot = darfur[idx_boot, ]
    ols_boot = lm(peacefactor ~ directlyharmed + female + village +
        age + farmer_dar + herder_dar + pastvoted + hhsize_darfur,
        data = data_boot)
    newdata1 = data.frame(directlyharmed = 1, female = data_boot$female,
        village = data_boot$village, age = data_boot$age, farmer_dar = data_boot$farmer_dar,
        herder_dar = data_boot$herder_dar, pastvoted = data_boot$pastvoted,
        hhsize_darfur = data_boot$hhsize_darfur)
    Ey1.boot = mean(predict(ols_boot, newdata = newdata1))
    newdata2 = data.frame(directlyharmed = 0, female = data_boot$female,
        village = data_boot$village, age = data_boot$age, farmer_dar = data_boot$farmer_dar,
        herder_dar = data_boot$herder_dar, pastvoted = data_boot$pastvoted,
        hhsize_darfur = data_boot$hhsize_darfur)
    Ey0.boot = mean(predict(ols_boot, newdata = newdata2))
    ate_boot1[i] = Ey1.boot - Ey0.boot
}

se.boot = sd(ate_boot1)
z_crit = qnorm(0.95)
low = est.ATE - z_crit * se.boot
up = est.ATE + z_crit * se.boot
# 95% CI using bootstrap method
c(low, up)
```

```
## directlyharmed directlyharmed
##     0.05522134     0.13941030
```

```r
# 95% confidence interval using classical model
confint(ols, "directlyharmed")
```

```
##                     2.5 %    97.5 %
## directlyharmed 0.05166327 0.1429684
```

```r
# 95% CI using robust std error
CI = Confint(ols, vcov. = vcovHC(ols, type = "HC0"))
```

```
## Standard errors computed by vcovHC(ols, type = "HC0")
```

```r
CI["directlyharmed", ]
```

```
##   Estimate      2.5 %     97.5 %
## 0.09731582 0.05246717 0.14216447
```

Section 6.2: Code for Section 4.2

```r
# Estimated ATE using plug in principle
ols2 = lm(peacefactor ~ directlyharmed + female + I(directlyharmed *
```

```r
    female) + village + age + farmer_dar + herder_dar + pastvoted +
    hhsize_darfur, data = darfur)
newdata1 = data.frame(directlyharmed = 1, female = darfur$female,
    village = darfur$village, age = darfur$age, farmer_dar = darfur$farmer_dar,
    herder_dar = darfur$herder_dar, pastvoted = darfur$pastvoted,
    hhsize_darfur = darfur$hhsize_darfur)
Ey1 = mean(predict(ols2, newdata = newdata1))
newdata2 = data.frame(directlyharmed = 0, female = darfur$female,
    village = darfur$village, age = darfur$age, farmer_dar = darfur$farmer_dar,
    herder_dar = darfur$herder_dar, pastvoted = darfur$pastvoted,
    hhsize_darfur = darfur$hhsize_darfur)
Ey0 = mean(predict(ols2, newdata = newdata2))
est.ATE2 = Ey1 - Ey0
est.ATE2
```

```
## [1] 0.09751186
```

```r
# Estimated ATE using regression formula
est.ATE3 = coef(ols2)["directlyharmed"] + coef(ols2)["I(directlyharmed * female)"] *
    mean(darfur$female)
est.ATE3
```

```
## directlyharmed
##     0.09751186
```

```r
# 95% confidence interval for ATE using bootstrap and
# regression coefficient
set.seed(123)
B = 1000
n = nrow(darfur$peacefactor)
ate_boot = rep(NA, B)
for (i in 1:B) {
    idx_boot = sample(1:n, size = n, replace = T)
    data_boot = darfur[idx_boot, ]
    ols_boot = lm(peacefactor ~ directlyharmed + female + I(directlyharmed *
        female) + village + age + farmer_dar + herder_dar + pastvoted +
        hhsize_darfur, data = data_boot)
    ate_boot[i] = coef(ols_boot)["directlyharmed"] + coef(ols_boot)["I(directlyharmed * female)"] *
        mean(data_boot$female)
}

se.boot = sd(ate_boot)
z_crit = qnorm(0.95)
low = est.ATE3 - z_crit * se.boot
up = est.ATE3 + z_crit * se.boot
# 95% CI using bootstrap method
c(low, up)
```

```
## directlyharmed directlyharmed
##     0.05556585     0.13945787
```

```r
# 95% confidence interval for ATE using bootstrap and plug
# in
set.seed(123)
B = 1000
n = nrow(darfur$peacefactor)
```

```r
ate_boot1 = rep(NA, B)
for (i in 1:B) {
    idx_boot = sample(1:n, size = n, replace = T)
    data_boot = darfur[idx_boot, ]
    ols_boot = lm(peacefactor ~ directlyharmed + female + I(directlyharmed *
        female) + village + age + farmer_dar + herder_dar + pastvoted +
        hhsize_darfur, data = data_boot)
    newdata1 = data.frame(directlyharmed = 1, female = data_boot$female,
        village = data_boot$village, age = data_boot$age, farmer_dar = data_boot$farmer_dar,
        herder_dar = data_boot$herder_dar, pastvoted = data_boot$pastvoted,
        hhsize_darfur = data_boot$hhsize_darfur)
    Ey1.boot = mean(predict(ols_boot, newdata = newdata1))
    newdata2 = data.frame(directlyharmed = 0, female = data_boot$female,
        village = data_boot$village, age = data_boot$age, farmer_dar = data_boot$farmer_dar,
        herder_dar = data_boot$herder_dar, pastvoted = data_boot$pastvoted,
        hhsize_darfur = data_boot$hhsize_darfur)
    Ey0.boot = mean(predict(ols_boot, newdata = newdata2))
    ate_boot1[i] = Ey1.boot - Ey0.boot
}

se.boot = sd(ate_boot1)
z_crit = qnorm(0.95)
low = est.ATE3 - z_crit * se.boot
up = est.ATE3 + z_crit * se.boot
# 95% CI using bootstrap method
c(low, up)
```

```
## directlyharmed directlyharmed
##     0.05556585     0.13945787
```

```r
# estimating tau using regression coefficients

# imbalance
delta = 0.2
# impact
gamma = 0.2
# tau : coefficient of regressing Yi on Di, Vi, Fi, Wi and
# Ci
tau = coef(ols)["directlyharmed"] - delta * gamma
tau
```

Section 6.3: Code for Section 5.1

```
## directlyharmed
##     0.05731582
```

```r
# estimating tau using FWL theorem

# from FWL theorem, tau1 = cov(tilde.yi,
# tilde.di)/var(tilde.di)
est_tau1 = cov(tilde.yi, tilde.di)/var(tilde.di)
# Estimated tau using a bias of 0.04 is
est.tau = est_tau1 - delta * gamma
est.tau
```

```
## [1] 0.05731582
```

```r
# estimating tau using plug in principle

# Since ATE for model 1 = (E[Y|D=1, V,F,X] - E[Y|D=0,
# V,F,X])/n (using plug in principle)
Ey1 = mean(predict(ols, newdata = newdata1))
Ey0 = mean(predict(ols, newdata = newdata2))
est.tau1 = Ey1 - Ey0
# Estimated tau using a bias of 0.04 is
est.tau = est.tau1 - delta * gamma
est.tau
```

```
## [1] 0.05731582
```

```r
# 95% Confidence interval for tau using bootstrap and
# regression coefficient from model1
set.seed(123)
B = 1000
n = nrow(darfur$peacefactor)
tau_boot = rep(NA, B)
for (i in 1:B) {
    idx_boot = sample(1:n, size = n, replace = T)
    data_boot = darfur[idx_boot, ]
    ols_boot = lm(peacefactor ~ directlyharmed + female + village +
        age + farmer_dar + herder_dar + pastvoted + hhsize_darfur,
        data = data_boot)
    tau_boot[i] = coef(ols_boot)["directlyharmed"] - delta *
        gamma
}

# 95% Confidence interval for tau using bootstrap
se.boot = sd(tau_boot)
z_crit = qnorm(0.95)
low = tau - z_crit * se.boot
up = tau + z_crit * se.boot
# 95% CI using bootstrap method
c(low, up)
```

```
## directlyharmed directlyharmed
##     0.01522134     0.09941030
```

```r
# 95% Confidence interval for tau using bootstrap and plug
# in estimate for tau1
set.seed(123)
B = 1000
n = nrow(darfur$peacefactor)
tau_boot1 = rep(NA, B)
for (i in 1:B) {
    idx_boot = sample(1:n, size = n, replace = T)
    data_boot = darfur[idx_boot, ]
    ols_boot = lm(peacefactor ~ directlyharmed + female + village +
        age + farmer_dar + herder_dar + pastvoted + hhsize_darfur,
        data = data_boot)
    newdata1 = data.frame(directlyharmed = 1, female = data_boot$female,
        village = data_boot$village, age = data_boot$age, farmer_dar = data_boot$farmer_dar,
```

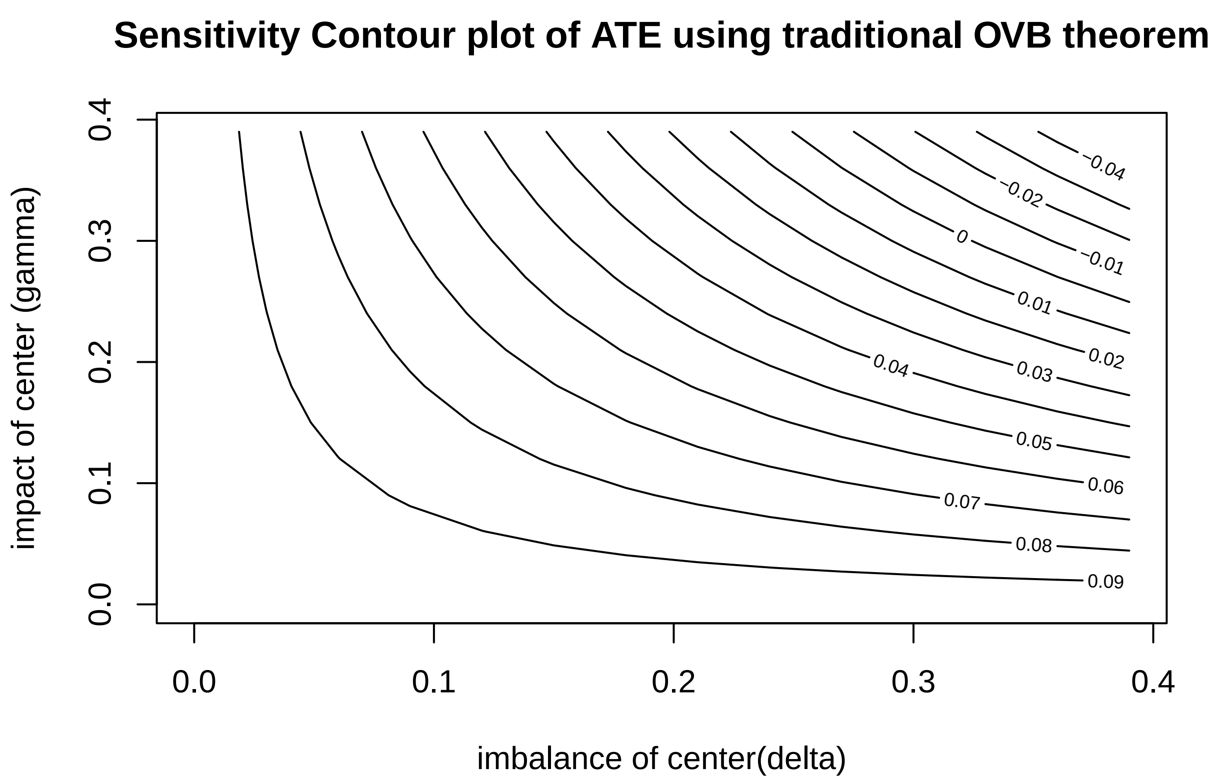
```
        herder_dar = data_boot$herder_dar, pastvoted = data_boot$pastvoted,
        hhsize_darfur = data_boot$hhsize_darfur)
    Ey1.boot = mean(predict(ols_boot, newdata = newdata1))
    newdata2 = data.frame(directlyharmed = 0, female = data_boot$female,
        village = data_boot$village, age = data_boot$age, farmer_dar = data_boot$farmer_dar,
        herder_dar = data_boot$herder_dar, pastvoted = data_boot$pastvoted,
        hhsize_darfur = data_boot$hhsize_darfur)
    Ey0.boot = mean(predict(ols_boot, newdata = newdata2))
    ate_boot = Ey1.boot - Ey0.boot
    tau_boot1[i] = ate_boot - delta * gamma
}

# 95% Confidence interval for tau using bootstrap
se.boot = sd(tau_boot1)
z_crit = qnorm(0.95)
low = tau - z_crit * se.boot
up = tau + z_crit * se.boot
# 95% CI using bootstrap method
c(low, up)
```

```
## directlyharmed directlyharmed
##     0.01522134     0.09941030
```

```
# ATE function
ate = function(delta, gamma) {
    tau1 = 0.0973
    tau = tau1 - delta * gamma
}
```

```
# sensitivity analysis
delta = seq(0, 0.4, 0.03)
gamma = seq(0, 0.4, 0.03)
eta = outer(delta, gamma, ate)
contour(delta, gamma, eta, levels = c(0.09, 0.08, 0.07, 0.06,
    0.05, 0.04, 0.03, 0.02, 0.01, 0, -0.01, -0.02, -0.03, -0.04),
    xlab = "imbalance of center(delta)", ylab = "impact of center (gamma)",
    main = "Sensitivity Contour plot of ATE using traditional OVB theorem")
```




```
tilde.yols = resid(ols)
sd_ratio = sd(tilde.yols)/sd(tilde.di)
sd_ratio
```

Section 6.4: Code for Section 5.2

```
## [1] 0.6507676
```

```
# residual variation of outcome Y_i explained by unobserved
# confounder Z_i = R^2_y~z|D,W = 12%
r2.yz1dw = 0.12
# residual variation of treatment D_i explained by
# unobserved confounder Z_i = R^2_D~z|W = 1%
r2.dz1w = 0.01
# estimated bias
est.bias = sqrt(r2.yz1dw * r2.dz1w/(1 - r2.dz1w)) * sd_ratio

# estimated adjusted tau using plug in principle using OVB
# theorem
tau.est1 = est_tau1 - est.bias
tau.est1
```

```
## [1] 0.074659
```

```
tau.est2 = adjusted_estimate(ols, treatment = "directlyharmed",
    r2dz.x = r2.dz1w, r2yz.dx = r2.yz1dw)
tau.est2
```

```
## directlyharmed
##       0.074659
```

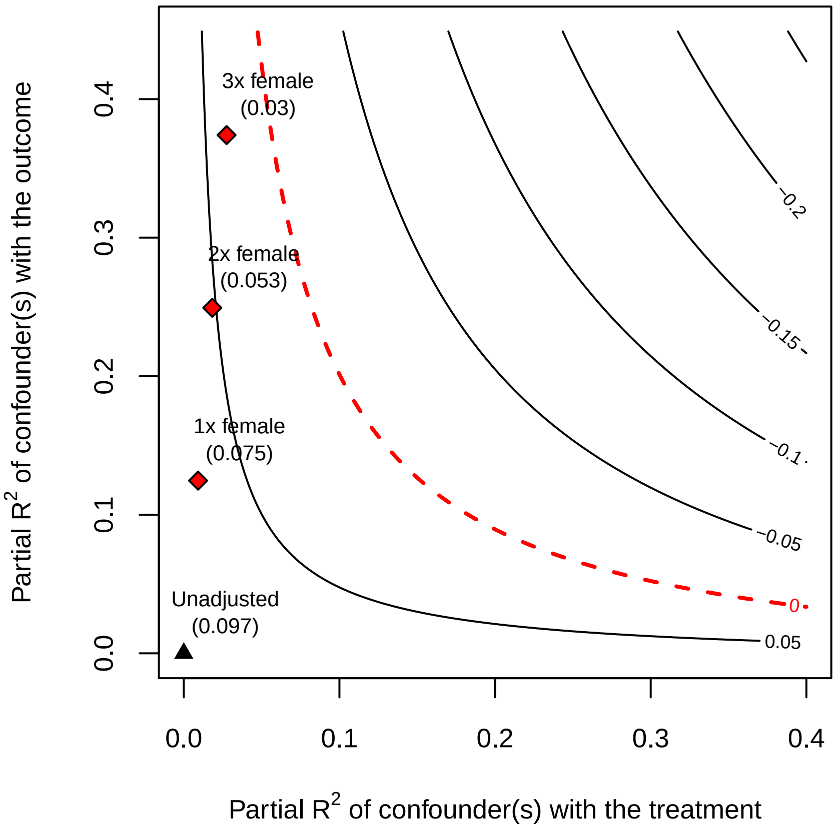
```r
# 95% confidence interval for tau using bootstap
set.seed(123)
B = 1000
n = nrow(darfur$peacefactor)
tau_boot = rep(NA, B)
for (i in 1:B) {
    idx_boot = sample(1:n, size = n, replace = T)
    data_boot = darfur[idx_boot, ]
    y.dvfx = lm(peacefactor ~ directlyharmed + female + village +
        age + farmer_dar + herder_dar + pastvoted + hhsize_darfur,
        data = data_boot)
    resid_y.dvfx = resid(y.dvfx)
    d.fvx = lm(directlyharmed ~ female + village + age + farmer_dar +
        herder_dar + pastvoted + hhsize_darfur, data = data_boot)
    resid_d.fvx = resid(d.fvx)
    sd_ratio = sd(resid_y.dvfx)/sd(resid_d.fvx)
    bias = abs(sqrt(r2.yz1dw * r2.dz1w/(1 - r2.dz1w)) * sd_ratio)
    tau_boot[i] = coef(y.dvfx)["directlyharmed"] - bias
}

se.boot = sd(tau_boot)
z_crit = qnorm(0.95)
low = tau.est2 - z_crit * se.boot
up = tau.est2 + z_crit * se.boot
# 95% CI using bootstrap method
c(low, up)
```

```
## directlyharmed directlyharmed
##     0.03251059     0.11680740
```

```r
# sensitivity analysis
darfur.sensitivity <- sensemakr(model = ols, treatment = "directlyharmed",
    benchmark_covariates = "female", kd = 1:3, ky = 1:3, alpha = 0.05)

plot(darfur.sensitivity)
```




```
# ovb_minimal_reporting(darfur.sensitivity, format =
# 'html')
```

References:

1. Carlos Cinelli and Chad Hazlett. Making sense of sensitivity: Extending omitted variable bias. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82(1):39–67, 2020. 2. Chad Hazlett. Angry or weary? how violence impacts attitudes toward peace among darfurian refugees. Journal of Conflict Resolution, 64(5):844–870, 2020. 3. Carlos Cinelli, Jeremy Ferwerda, and Chad Hazlett. sensemakr: Sensitivity analysis tools for ols in r and stata. Available at SSRN 3588978, 2020.