Confounding of Selection and Influence

Cecilia Cavero-Sanchez

2024-11-17

- There are important challenges in conducting causal inference on contagion effects in observational data.
- (Fyfe and Desmarais 2024) show how we can use the "split-haves" test, robust to confounding, and apply it to studies of contagion effects.
- In this tutorial I will go over the method and several replication examples.

The split halves test

- Observational data are subject to confounding when identifying contagion/influence effects because of the co-existence of homophily and influence.
- The SH test isolates the impact of contagion by assuming the pre-existence of a network in the data without conditioning on it.

1. Test data and adjust it for non-stationarity.
2. Randomly split observational time-series cross-section data into two halves based on node (country in country-year data).
3. Calculate mean values for each half for every time period.
4. Run regression setting time t means as the dependent variable and $t-1$ means of each half as independent variables.
5. Perform steps $1-3$ N times to recover a mean and $p-$value that indicate whether contagion is present or not.
6. Contagion signal is the average value of the estimated relationship between the mean value of the first half at time t with the mean value of the second half at time $t-1$, conditional on the mean value of the first half at $t-1$. In a way, it is the relationship between both halves at different time pints.
7. The $p-$value is calculated as the minimum of two proportions, the proportion of times the contagion signal is > 0 and the proportion of times when it is > 0. We obtain the $p-$value by multiplying the minimum proportion by 2 for a two-tailed test of whether there is contagion in the data.
8. The estimate of general contagion tells us the average effect of a one-unit increase in the outcome value of any other node in the following year.

Applying the split-halves test

- I will use three of the replication examples in (Fyfe and Desmarais 2024) to illustrate the use of the split-halves test and its impact on results of previous studies.

Confirmation of Contagion: Conflict Onset

- The first replication is of (Buhaug and Gleditsch 2008), who find that there is a neighborhood effect of armed conflict.
- DV is binary indicator of conflict onset and a three-level ordinal variable indicating the type of conflict.
- First, here is the replication of the original study.

```
## Buhaug and Gleditsch
rm(list = ls())
# Libraries
library(ggplot2)
library(dplyr)
library(broom)
library(nnet)
library(dplyr)
library(haven)

# Original data
rep <- read_dta("hb_ksg_replication.dta")

# Model 1
model1 <- glm(allons3 ~ neighall + neighpol + I(neighpol^2) +
                neighlgdp + peaceall,
              data = rep, family = binomial(link = "logit"))

# Model 2
model2 <- glm(allons3 ~ ncivwar + neighpol + I(neighpol^2) +
                neighlgdp  + peaceall,
              data = rep, family = binomial(link = "logit"))

# Model 3
model3 <- glm(allons3 ~ neighall + neighpol + I(neighpol^2) +
                neighlgdp + polity2l + I(polity2l^2) +
              lgdp96l + lnpop + peaceall,
              data = rep, family = binomial(link = "logit"))

# Model 4
model4 <- glm(allons3 ~ ncivwar + neighpol + I(neighpol^2) +
                neighlgdp + polity2l + I(polity2l^2) +
              lgdp96l + lnpop + peaceall,
              data = rep, family = binomial(link = "logit"))

# Table 2, conflict neighbors only
# Model 5
model5 <- glm(allons3 ~ lnblength + lndist + ethlink2 +
                lneighbref + pop_nc + nterr + lbd_cum +
                polity2l + I(polity2l^2) + lgdp96l + lnpop +
                peaceall,
              data = subset(rep, ncivwar == 1),
              family = binomial(link = "logit"))

# Model 6
model6 <- glm(allons3 ~ lnblength + confbord + ethlink2 +
```

```r
                lneighbref + pop_nc + nterr + lbd_cum +
              polity2l + I(polity2l^2) + lgdp96l + lnpop +
                peaceall,
              data = subset(rep, ncivwar == 1),
              family = binomial(link = "logit"))

# Model 7, multinomial logit of terr and gov conflicts
model7 <- multinom(mons3 ~ ethlink2 + lneighbref + nterr +
                    polity2l + I(polity2l^2) +
                     lgdp96l + lnpop + peaceall,
                  data = subset(rep, ncivwar == 1))
```

```
## # weights:  30 (18 variable)
## initial  value 2906.928116
## iter  10 value 747.699419
## iter  20 value 582.810436
## iter  30 value 550.835875
## final  value 550.816856
## converged
```

```r
# Summary
library(texreg)
screenreg(list(model1, model2, model3, model4))
```

```
##
## ==================================================================
##                  Model 1      Model 2      Model 3      Model 4
## ------------------------------------------------------------------
## (Intercept)        -0.09        -0.52        -3.11 **     -3.31 **
##                    (0.88)       (0.87)       (1.07)       (1.04)
## neighall            0.59 *                    0.32
##                    (0.28)                    (0.30)
## neighpol            0.01         0.01        -0.00         0.00
##                    (0.02)       (0.02)       (0.02)       (0.02)
## neighpol^2         -0.00        -0.00         0.00         0.00
##                    (0.00)       (0.00)       (0.00)       (0.00)
## neighlgdp          -0.39 ***    -0.37 ***    -0.03        -0.02
##                    (0.11)       (0.11)       (0.15)       (0.15)
## peaceall           -0.02 **     -0.02 **     -0.01 *      -0.01 *
##                    (0.01)       (0.01)       (0.01)       (0.01)
## ncivwar                          0.67 ***                  0.48 **
##                                 (0.15)                    (0.15)
## polity2l                                      0.00         0.01
##                                              (0.01)       (0.01)
## polity2l^2                                   -0.01 ***    -0.01 ***
##                                              (0.00)       (0.00)
## lgdp96l                                      -0.30 *      -0.27 *
##                                              (0.12)       (0.12)
## lnpop                                         0.30 ***     0.28 ***
##                                              (0.05)       (0.05)
## ------------------------------------------------------------------
## AIC              1849.17      1832.01      1795.70      1786.60
## BIC              1889.93      1872.77      1863.64      1854.54
```

```
## Log Likelihood  -918.59      -910.01      -887.85      -883.30
## Deviance        1837.17      1820.01      1775.70      1766.60
## Num. obs.       6591         6591         6591         6591
## ================================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

```r
screenreg(list(model5, model6, model7))
```

```
##
## =====================================================
##                  Model 1      Model 2      Model 3
## -----------------------------------------------------
## (Intercept)        -3.83 **     -3.85 **
##                    (1.37)       (1.38)
## lnblength          -0.04        -0.04
##                    (0.10)       (0.10)
## lndist             -0.01
##                    (0.04)
## ethlink2            0.61 *       0.62 *
##                    (0.26)       (0.26)
## lneighbref          0.04         0.04
##                    (0.02)       (0.02)
## pop_nc             -0.03        -0.03
##                    (0.08)       (0.08)
## nterr               0.61 **      0.61 **
##                    (0.20)       (0.20)
## lbd_cum            -0.05        -0.05
##                    (0.05)       (0.05)
## polity2l            0.02         0.02
##                    (0.02)       (0.02)
## polity2l^2         -0.01        -0.01
##                    (0.00)       (0.00)
## lgdp96l            -0.23        -0.23
##                    (0.12)       (0.12)
## lnpop               0.32 ***     0.32 ***
##                    (0.06)       (0.06)
## peaceall           -0.01        -0.01
##                    (0.01)       (0.01)
## confbord                         0.01
##                                 (0.26)
## 1: (Intercept)                               -6.13 ***
##                                              (1.80)
## 1: ethlink2                                   0.63
##                                              (0.40)
## 1: lneighbref                                -0.02
##                                              (0.04)
## 1: nterr                                      1.18 ***
##                                              (0.31)
## 1: polity2l                                   0.04
##                                              (0.02)
## 1: polity2l^2                                -0.00
##                                              (0.01)
## 1: lgdp96l                                   -0.49 **
##                                              (0.19)
```

```
## 1: lnpop                                            0.54 ***
##                                                    (0.09)
## 1: peaceall                                        -0.03 *
##                                                    (0.02)
## 2: (Intercept)                                     -4.84 ***
##                                                    (1.45)
## 2: ethlink2                                         0.68 *
##                                                    (0.32)
## 2: lneighbref                                       0.06 *
##                                                    (0.02)
## 2: nterr                                            0.14
##                                                    (0.25)
## 2: polity2l                                        -0.01
##                                                    (0.02)
## 2: polity2l^2                                      -0.01 *
##                                                    (0.00)
## 2: lgdp96l                                          0.05
##                                                    (0.15)
## 2: lnpop                                            0.06
##                                                    (0.08)
## 2: peaceall                                        -0.00
##                                                    (0.01)
## ---------------------------------------------------------
## AIC               991.09       991.12      1137.63
## BIC              1067.47      1067.50      1243.49
## Log Likelihood   -482.54      -482.56      -550.82
## Deviance          965.09       965.12      1101.63
## Num. obs.        2632         2632         2646
## K                                             3
## =========================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

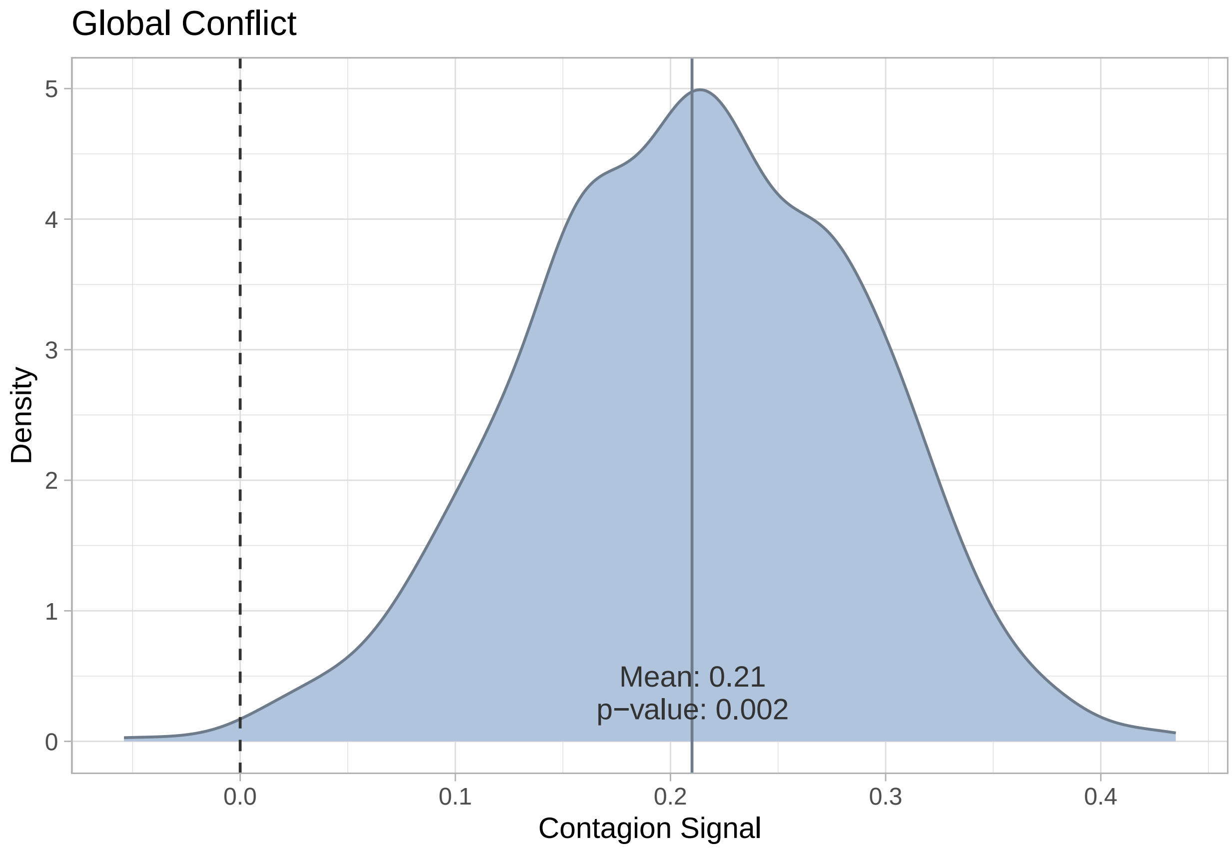
- The results from models $1-4$ (full sample analysis) show that the presence of neighboring conflicts (neighall) and the number of neighboring civil wars (ncivwar) significantly increases the likelihood of conflict onset.
- The results from models $5-7$ show that ethnic links to neighboring conflicts (ethlink2) and the number of territorial conflicts in neighboring countries (nterr) significantly increases the risk of conflict spillover.
- Below, by applying the split halves test, we see that this is an example where the SH test confirms the author's main results of contagion.

```r
## Replication
# Libraries
library(maditr)
library(ggplot2)
library(tidyr)
library(ggplot2)
library(haven)
library(ContagionTest) # can download from GitHub
#devtools::install_github("rebekahfyfe/ContagionTest")

# Data
d <- read.table("conflict.tab", header = T)
```

```r
# Removing duplicate rows
d <- d[-4439,]
d <- d[-5014,]

# Selecting necessary columns (country, date, DV)
d1 <- d %>%
  select("abbrev", "year", "allons3") %>%
  pivot_wider(names_from = year, values_from = allons3)
d1 <- as.data.frame(d1)

# Changing null values to NAs (treating as missing data)
d1[d1 == 'NULL'] <- NA

# Formatting for split-halves test
d1 <- STFormat(d1)

# Running split-halves contagion test
simmodels <- lag_pc_test(d1, 1000, 3, T, 0.1,
                         lagWin = 1, missingData = T)
```

```
## [1] "Did not take 1st difference"
```

```r
# Summary of models
simmodels <- as.data.frame(simmodels)
names(simmodels) <- c("intercept","t-1coef","counterpart")

# Calculate contagion signal
xmean <- mean(simmodels$counterpart) ## input this in the plot
xmean <- round(xmean, digits = 4)

# P-value of the signal (proportion of results < 0)
pval <- sum(simmodels$counterpart < 0) / 1000  ## pvalue
pval <- round(pval, digits = 3)

# Density graph of results
density_graph(simmodels, 1000, xmean, 0.5, xmean, 0.25,
              title = "Global Conflict")
```



- Using $1,000$ random splits, the SH returns a positive contagion signal and a $p-$value < 0.01.
- The expected prevalence of civil conflict onset in one country increases by approximately 0.02 for every 0.1 increase in lagged civil war prevalence among the other countries.
- The results support the findings of Buhaug and Gleditsch.

Challenging Non-Contagion: Pro-Democracy Protests

- The authors find that, contrary to some strands of literature, pro-democracy protests do not diffuse to other countries (Brancati and Lucardi 2018).
- DV is protest onset.
- This is an example where the SH test challenges the author's results of no contagion.

```
## Brancati and Lucardi
# Violence in the Netherlands data, from Braun 2011
d <- as.data.frame(read.delim(unzip("violneth.zip")))

# Selecting only the necessary columns
d <- d[, c(3, 5, 21:55)]

# Creating a singe variable for dates
T1 <- c(rep(1, 30), rep(0, 1065))
d$TT1 <- rep(T1, 474)
```

```r
date <- seq(as.Date("2001-01-01"),
            as.Date("2003-12-31"), by = "days")
d$date <- rep(date, 474)

# Changing to wide format
d <- d %>% select(number, date, countinc) %>%
  pivot_wider(names_from = date, values_from = countinc)

# Formatting to be used with congation test
d <- STFormat(d)

# Running parallel contagion test
simmodNVio <- lag_pc_test(d, 1000, 1, T, 0.05)
```

```
## [1] "Took 1st difference"
```

```r
summary(simmodNVio)
```

```
##   (Intercept)        c(j1mean.tm1, j2mean.tm1) c(j2mean.tm1, j1mean.tm1)
##  Min.   :0.0003944   Min.   :0.0979            Min.   :0.05258
##  1st Qu.:0.0004100   1st Qu.:0.1482            1st Qu.:0.11515
##  Median :0.0004112   Median :0.1638            Median :0.13170
##  Mean   :0.0004115   Mean   :0.1646            Mean   :0.13075
##  3rd Qu.:0.0004129   3rd Qu.:0.1802            3rd Qu.:0.14722
##  Max.   :0.0004291   Max.   :0.2437            Max.   :0.19829
```

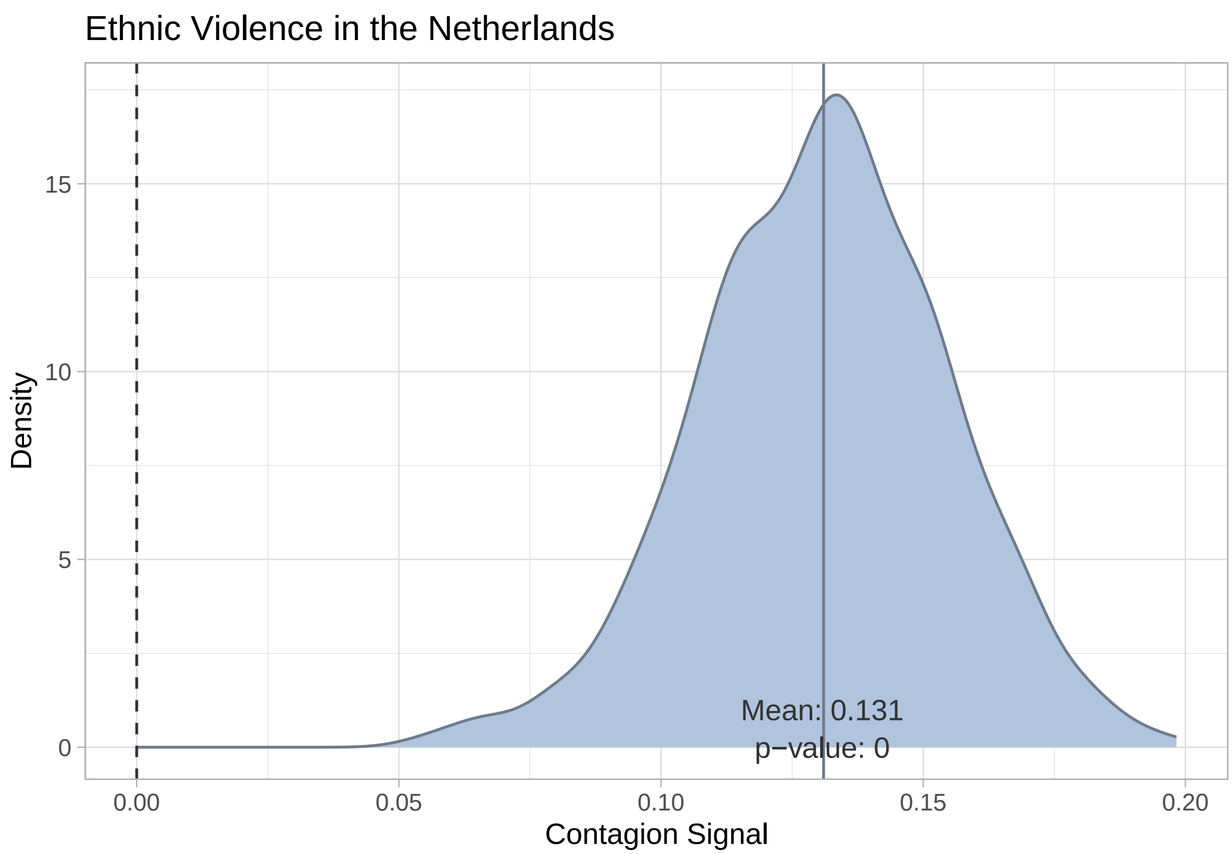
```r
# Creating a dataframe with results
simmodNVio <- as.data.frame(simmodNVio)
names(simmodNVio) <- c("intercept","t-1coef","counterpart")

# Calculating mean (contagion signal)
mean <- mean(simmodNVio$counterpart) ## input this in the plot below
(mean <- round(mean, digits = 10))
```

```
## [1] 0.1307509
```

```r
# Significance of the signal, proportion of means less than 0
pval <- sum(simmodNVio$counterpart < 0) / 1000  ## pvalue
(pval <- round(pval, digits = 3))
```

```
## [1] 0
```

```r
# Density graph of results
density_graph(simmodNVio, 1000, mean, 1, mean, 0,
              title = "Ethnic Violence in the Netherlands")
```



- While Brancati and Lucardi find no contagion effect, the SH test shows that there is indeed statistical evidence of contagion.

Challenging Contagion: Civilian Targeting

- The authors find that there is a spillover effect that results in the spreading of violence against civilians by armed actors (Lis, Spagat, and Lee 2021).
- DV is the Civilian Targeting Index, which ranges from 0 to 1000.
- This is an example where the SH test challenges the author's results of contagion.

```r
## Replication data
d <- read_dta("civiliantargeting.dta")

# Selecting necessary columns (country, date, DV)
d <- d %>%
  select(actor_id, year, cti)

# Changing to wide format
d <- dcast(d, actor_id ~ year, value.var = "cti")

# Formatting to be used with split-halves contagion test
d <- STFormat(d)
```

```r
# running contagion test
lslres <- lag_pc_test(d, 1000, 1, T, 0.05, 1, F)
```

```
## [1] "Took 1st difference"
```

```r
summary(lslres)
```

```
##   (Intercept)      c(j1mean.tm1, j2mean.tm1) c(j2mean.tm1, j1mean.tm1)
##  Min.   :-1.7334   Min.   :-0.58013          Min.   :-0.28110
##  1st Qu.:-1.2921   1st Qu.:-0.40000          1st Qu.:-0.02571
##  Median :-1.2359   Median :-0.34034          Median : 0.04104
##  Mean   :-1.2402   Mean   :-0.33310          Mean   : 0.03548
##  3rd Qu.:-1.1891   3rd Qu.:-0.26924          3rd Qu.: 0.10138
##  Max.   :-0.6576   Max.   :-0.02375          Max.   : 0.27829
```

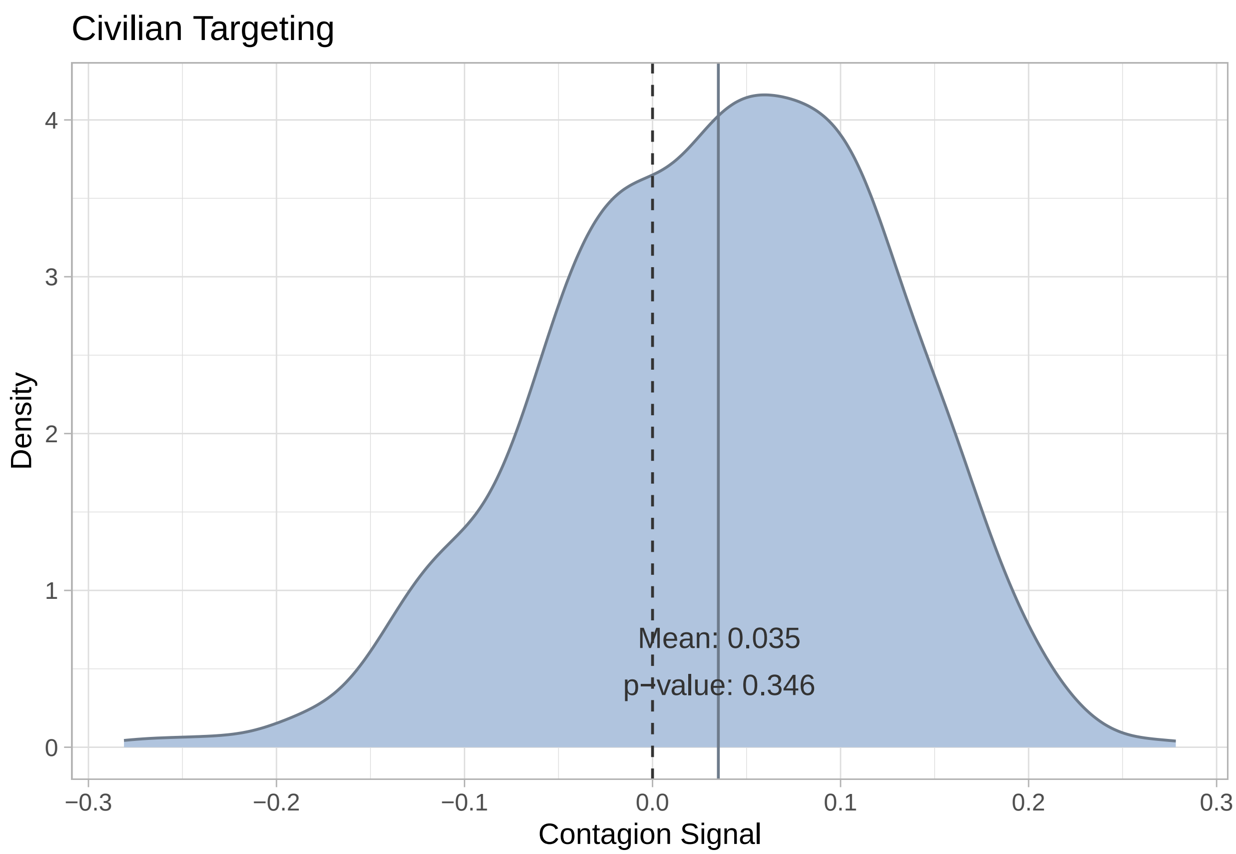
```r
# Creating a data frame with results
lslres <- as.data.frame(lslres)
names(lslres) <- c("intercept","t-1coef","counterpart")

# Calculating mean (contagion signal)
lslresmean <- mean(lslres$counterpart) ## input this in the plot below
(lslresmean <- round(lslresmean, digits = 10))
```

```
## [1] 0.03548029
```

```r
# Significance of the signal, proportion of means less than 0
lslrespval <- sum(lslres$counterpart < 0) / 1000  ## pvalue
(lslrespval <- round(lslrespval, digits = 3))
```

```
## [1] 0.346
```

```r
# Density graph of results
density_graph(lslres, 1000, lslresmean, 0.7, lslresmean, 0.4,
              title = "Civilian Targeting")
```



- While Lis, Spagat and Lee find a contagion effect, the SH test shows that there is no statistical evidence of contagion.

References

Brancati, Dawn, and Adrián Lucardi. 2018. "Why Democracy Protests Do Not Diffuse." *Journal of Conflict Resolution* 63 (10): 2354–89. https://doi.org/10.1177/0022002718815957.

Buhaug, Halvard, and Kristian Skrede Gleditsch. 2008. "Contagion or Confusion? Why Conflicts Cluster in Space." *International Studies Quarterly* 52 (2): 215–33. https://doi.org/10.1111/j.1468-2478.2008.00499.x.

Fyfe, Rebekah, and Bruce Desmarais. 2024. "Causal Evidence for Theories of Contagious Civil Unrest." *International Studies Quarterly* 68 (4). https://doi.org/10.1093/isq/sqae124.

Lis, Piotr, Michael Spagat, and Uih Ran Lee. 2021. "Civilian Targeting in African Conflicts: A Poor Actor's Game That Spreads Through Space." *Journal of Peace Research* 58 (5): 900–914. https://doi.org/10.1177/0022343320961150.