huai_project

2024-04-02

A normal comparison of the air pollution in northern cities compared to southern cities would not adequately capture the causal effect of the Huai River policy due to the range of values being too large to indicate a significant difference in means. Ebenstein et all overcomes this barrier by a regression discontinuity design, directly comparing effects of air pollution on those immediately north and south of the Huai River. This comparison provides greater similarity between differing groups, and though we still incur a fundamental problem of economics by being unable to observe different outcomes on a single individual, regression discontinuity attempts to solve this problem by observing two individuals whose differences are negligible except being given treatment. In the study, the outcome variable is particulate matter, and the assignment variable is degrees north or south of the Huai River.

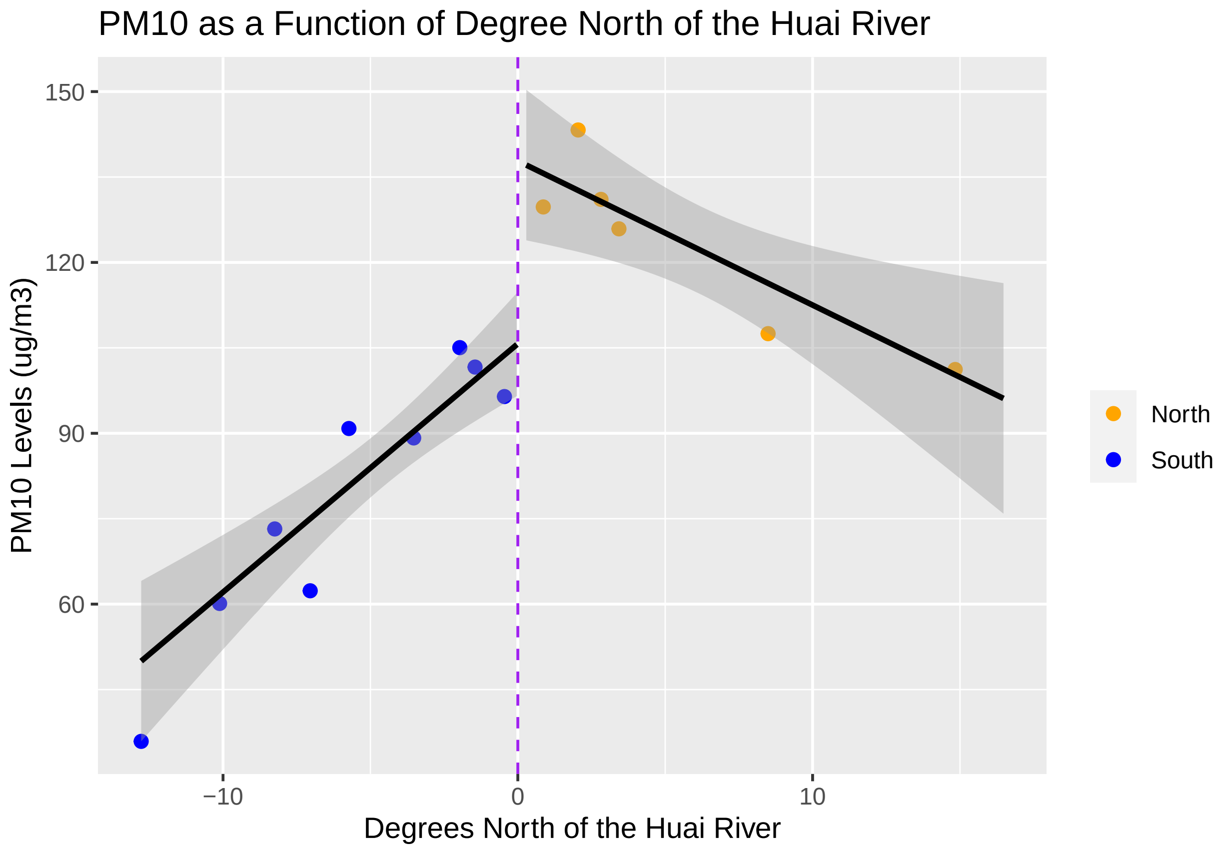
The identification assumption for regression discontinuity design is that change in Di is the only reason for discrete jumps in Yi around c. The first graphs are consistent with that assumption given that other observed factors did not jump discreetly around the latitude of the Huai River.

Accordingly, our regressions for each other observed characteristic reflect this characteristic by not being statistically significant around the Huai River. This is also indicated by the confidence interval for each regression including 0, which in effect, means our results are not statistically significant from 0.

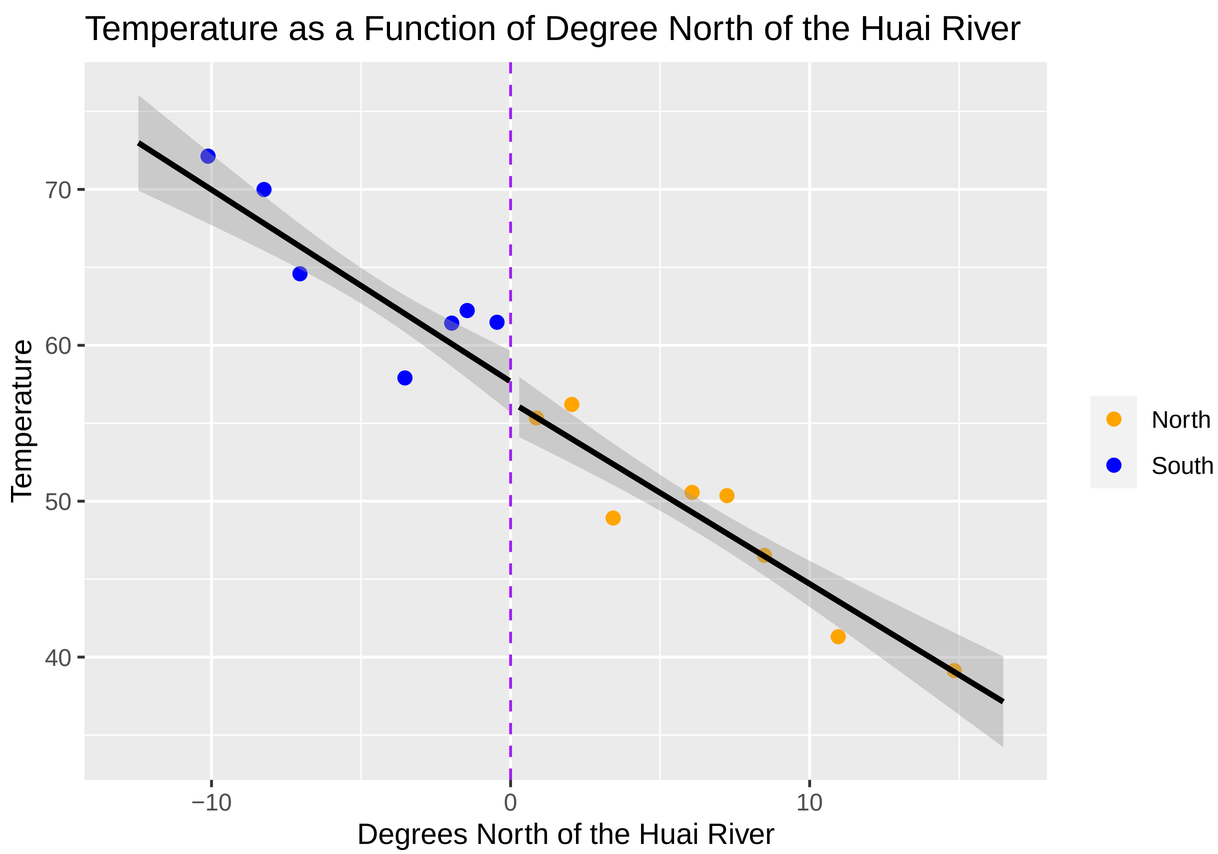
```
## Warning: Removed 2 rows containing non-finite values (‘stat_smooth()‘).
```

```
## Warning: Removed 5 rows containing non-finite values (‘stat_smooth()‘).
```

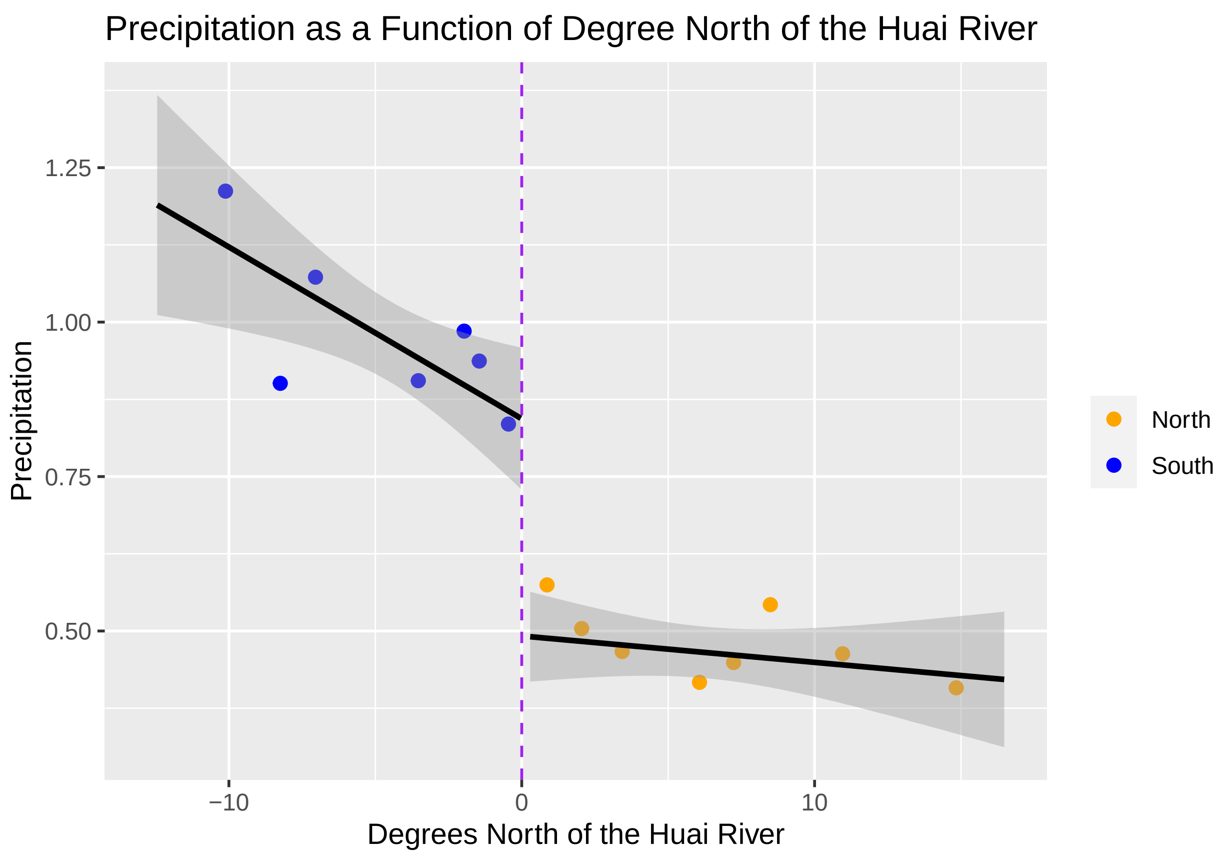
```
## Warning: Removed 6 rows containing missing values (‘geom_point()‘).
```



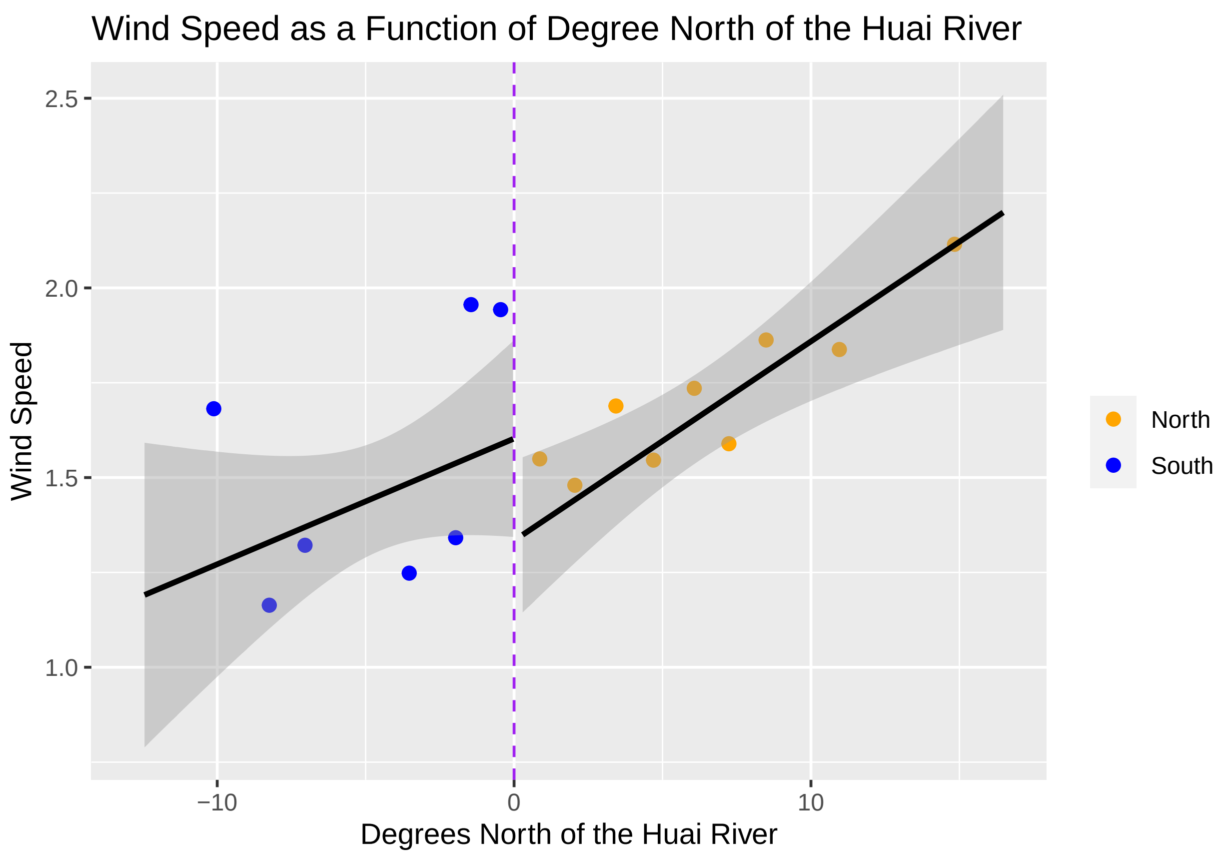

```
## Warning: Removed 6 rows containing non-finite values (`stat_smooth()`).

## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).

## Warning: Removed 6 rows containing missing values (`geom_point()`).
```




```
## Warning: Removed 6 rows containing non-finite values (‘stat_smooth()‘).

## Warning: Removed 2 rows containing non-finite values (‘stat_smooth()‘).

## Warning: Removed 6 rows containing missing values (‘geom_point()‘).
```




```
## Warning: Removed 4 rows containing non-finite values (`stat_smooth()`).

## Warning: Removed 1 rows containing non-finite values (`stat_smooth()`).

## Warning: Removed 5 rows containing missing values (`geom_point()`).
```



These binned scatterplot organizes data into subsets within a given range of values, and assign value to the bins based on the number of observations that fall within that specific range.

```
##
## Call:
## lm(formula = y ~ ., data = dat_step1, weights = weights)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -67.00 -14.01   0.24  12.91 174.10
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 105.7261     5.6187  18.817  < 2e-16 ***
## D            32.1029     8.1151   3.956 0.000117 ***
## x             4.3624     0.9961   4.379 2.22e-05 ***
## x_right      -6.8941     1.2655  -5.448 2.04e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 28.49 on 150 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.3841, Adjusted R-squared:  0.3718
## F-statistic: 31.18 on 3 and 150 DF,  p-value: 1.007e-15
```

```
## Sharp RD estimates using local polynomial regression.
```

```
##
## Number of Obs.                  154
## BW type                       mserd
## Kernel                      Uniform
## VCE method                      HC1
##
## Number of Obs.                   80           74
## Eff. Number of Obs.              44           33
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                   4.525        4.525
## BW bias (b)                   8.423        8.423
## rho (h/b)                     0.537        0.537
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    48.817    17.950     2.720     0.007    [13.636 , 83.999]
##         Robust         -         -     2.637     0.008    [13.888 , 94.329]
## =============================================================================
```

The result between the rdrobust and the hand regression lies within the coefficient and the standard errors. The hand regression has a smaller estimate and smaller standard errors (32.10, 8.11), where the rdrobust has a larger coefficient and standard errors (48.817, 17.950). Both P-Values indicate statistical significance (being smaller than 0.01).

```
## Covariate-adjusted Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                  154
## BW type                       mserd
## Kernel                      Uniform
## VCE method                       NN
##
## Number of Obs.                   80           74
## Eff. Number of Obs.              45           33
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                   4.851        4.851
## BW bias (b)                   8.655        8.655
## rho (h/b)                     0.561        0.561
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    48.748    18.437     2.644     0.008    [12.613 , 84.884]
##         Robust         -         -     2.520     0.012    [11.621 , 92.916]
## =============================================================================
```

```
## Covariate-adjusted Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                  154
## BW type                      Manual
## Kernel                   Triangular
```

```
## VCE method                         NN
##
## Number of Obs.                     80            74
## Eff. Number of Obs.                80            74
## Order est. (p)                      1             1
## Order bias  (q)                     2             2
## BW est. (h)                    30.000        30.000
## BW bias (b)                    30.000        30.000
## rho (h/b)                       1.000         1.000
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    33.452     9.335     3.583     0.000    [15.155 , 51.749]
##         Robust         -         -     3.146     0.002    [16.848 , 72.512]
## =============================================================================

## Covariate-adjusted Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                    154
## BW type                        Manual
## Kernel                        Uniform
## VCE method                         NN
##
## Number of Obs.                     80            74
## Eff. Number of Obs.                80            74
## Order est. (p)                      2             2
## Order bias  (q)                     3             3
## BW est. (h)                    30.000        30.000
## BW bias (b)                    30.000        30.000
## rho (h/b)                       1.000         1.000
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    43.635    14.036     3.109     0.002    [16.125 , 71.146]
##         Robust         -         -     3.006     0.003    [18.343 , 87.089]
## =============================================================================

## Covariate-adjusted Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                    154
## BW type                        Manual
## Kernel                        Uniform
## VCE method                         NN
##
## Number of Obs.                     80            74
## Eff. Number of Obs.                 9             3
## Order est. (p)                      0             0
## Order bias  (q)                     1             1
## BW est. (h)                     1.000         1.000
## BW bias (b)                     1.000         1.000
## rho (h/b)                       1.000         1.000
##
```

```
## ================================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## ================================================================================
##   Conventional    37.334     4.481     8.332     0.000    [28.551 , 46.116]
##         Robust         -         -     4.838     0.000    [25.052 , 59.171]
## ================================================================================
```

In the optimal bandwidth approach, we analyze symmetrical bandwidth. Typically a smaller bandwidth provides a more accurate estimate around the threshold, but it may lead to noise in the data, while a larger bandwidth reduces noise but might smooth out the treatment effect. If two different bandwidths are used for these regressions, then regressions will be estimated based on different samples, which will complicate the computation of standard errors for the estimates.

Typically, the data kernel is rectangular, meaning that all data is weighted the same in the regression. However, in the alternative kernel approach, a triangular kernel weights the data in descending order from the regression discontinuity in an upside-down cone shape. In theory, this will increase accuracy of the estimates due to the heavier weight placed on observations closer to the regression discontinuity, rather than weighting all observations equally. On a personal note, I think this kernel is the most accurate and most statistically valid purely based on the regression discontinuity design.

When comparing functional forms, the standard regression will be linear in parameters. We can alter this and create a quadratic or greater polynomial to better understand the underlying effect of the Huai River policy. In doing this, we estimate a greater discontinuity between North and South, but also larger standard errors, with a larger confidence interval. However, our results still remain statistically significant, implying validity to our regression design, albeit with a larger potential range of values.

The Smallest-Difference-In-Group-Means Estimator can be chosen by limiting the bandwidth to only one observation from each side of the regression discontinuity. Our P-Values are statistically significant, with a small range confidence interval. Our standard errors are the smallest out of any regression, which can be directly attributed to only 2 values being used by the regression design.

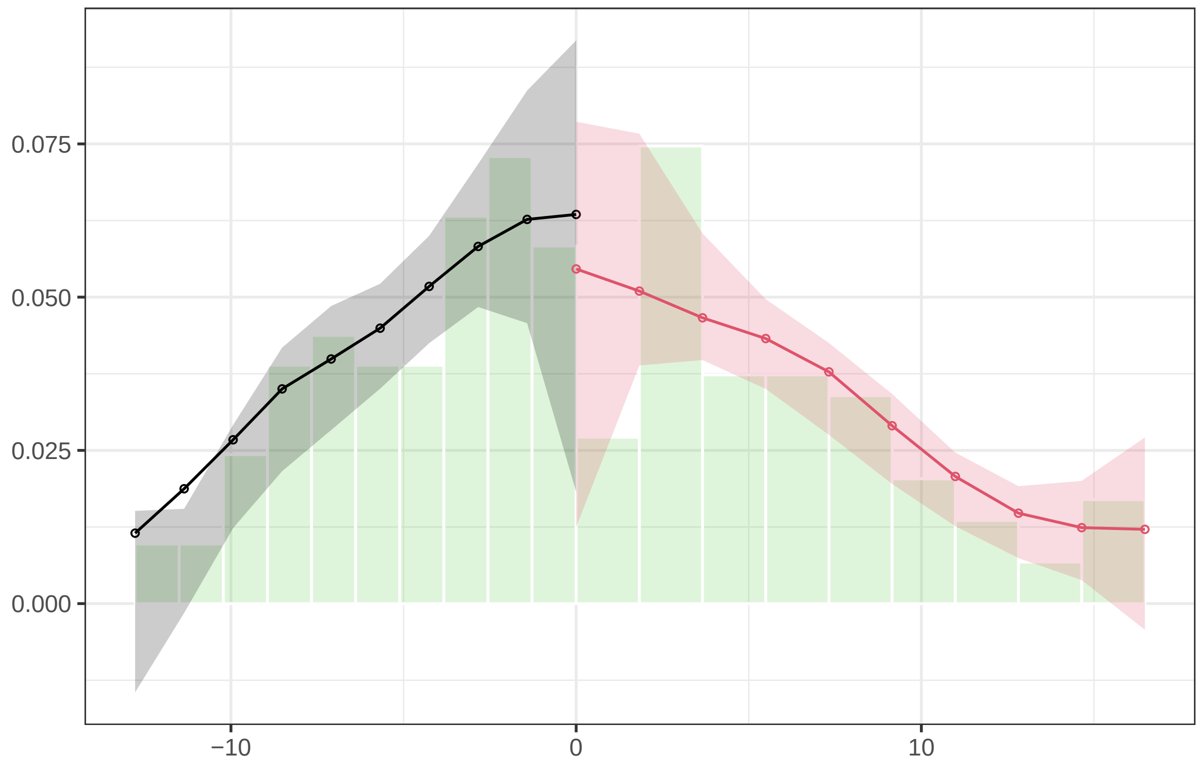
```
##      Regression Discontinuity Regression Discontinuity Symmetric
## [1,]                48.817327                          48.748334
## [2,]                20.520926                          20.738752
## [3,]                13.888024                          11.621374
## [4,]                94.328576                          92.915786
## [5,]                 4.525134                           4.851457
## [6,]                44.000000                          45.000000
##      Regression Discontinuity Kernel Regression Discontinuity Linear From
## [1,]                        33.45211                             43.63529
## [2,]                        14.20025                             17.53750
## [3,]                        16.84809                             18.34290
## [4,]                        72.51205                             87.08861
## [5,]                        30.00000                             30.00000
## [6,]                        80.00000                             80.00000
##      Regression Discontinuity Small Group Means
## [1,]                                  37.333641
## [2,]                                   8.704074
## [3,]                                  25.051998
## [4,]                                  59.171342
## [5,]                                   1.000000
## [6,]                                   9.000000
```

This table indicates summary information for each regression.

```
## Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                  153
## BW type                       mserd
## Kernel                      Uniform
## VCE method                       NN
##
## Number of Obs.                   76           77
## Eff. Number of Obs.              38           29
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                   3.703        3.703
## BW bias (b)                   6.874        6.874
## rho (h/b)                     0.539        0.539
## Unique Obs.                      76           77
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -5.410     4.551    -1.189     0.235   [-14.331 , 3.510]
##         Robust         -         -    -0.991     0.322   [-17.099 , 5.615]
## =============================================================================

## Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                  153
## BW type                       mserd
## Kernel                      Uniform
## VCE method                       NN
##
## Number of Obs.                   76           77
## Eff. Number of Obs.              27           18
## Order est. (p)                    1            1
## Order bias  (q)                   2            2
## BW est. (h)                   2.894        2.894
## BW bias (b)                   5.387        5.387
## rho (h/b)                     0.537        0.537
## Unique Obs.                      76           77
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -0.203     0.077    -2.620     0.009    [-0.355 , -0.051]
##         Robust         -         -    -2.170     0.030    [-0.362 , -0.018]
## =============================================================================

## Sharp RD estimates using local polynomial regression.
##
## Number of Obs.                  156
## BW type                       mserd
## Kernel                      Uniform
## VCE method                       NN
##
## Number of Obs.                   78           78
```

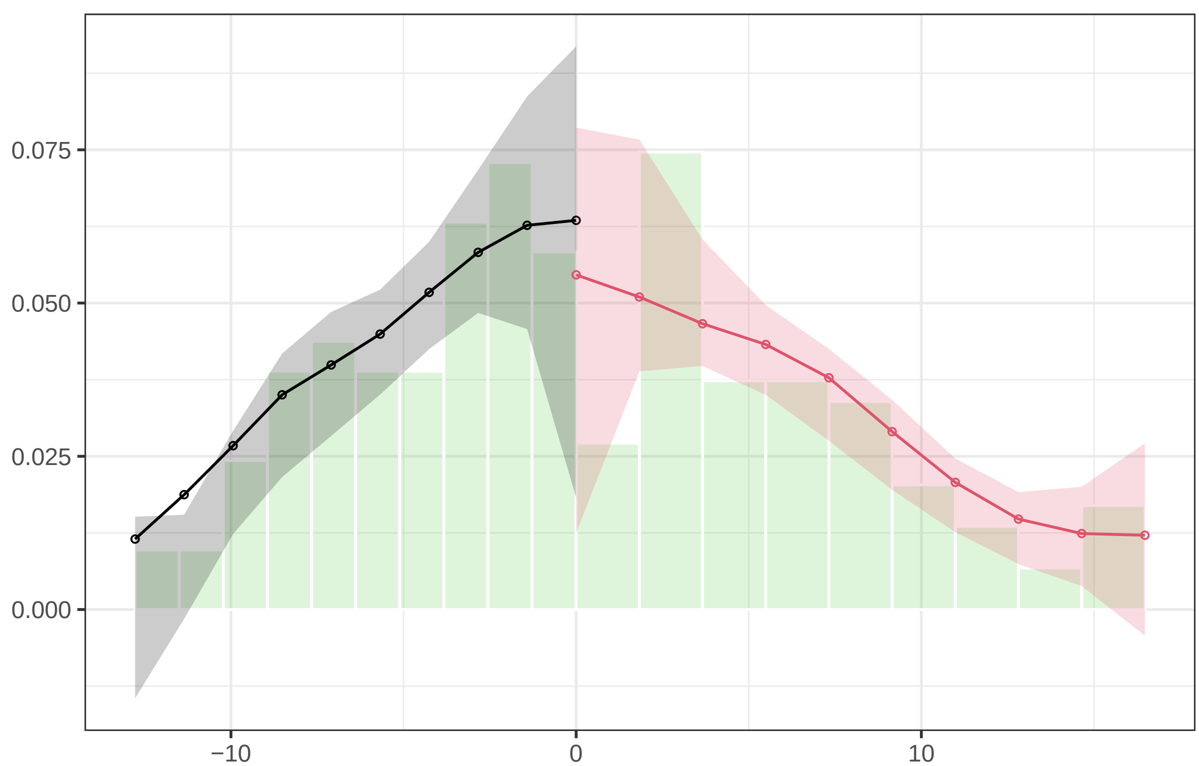
```
## Eff. Number of Obs.              32          22
## Order est. (p)                    1           1
## Order bias  (q)                   2           2
## BW est. (h)                   3.199       3.199
## BW bias (b)                   5.987       5.987
## rho (h/b)                     0.534       0.534
## Unique Obs.                      78          78
##
## =============================================================================
##         Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
## =============================================================================
##   Conventional    -0.621     0.332    -1.874     0.061    [-1.271 , 0.028]
##         Robust         -         -    -1.626     0.104    [-1.406 , 0.131]
## =============================================================================
```

The covariate smoothness tests indicate no discernable difference North and South of the Huai River. This reinforces our assumptions that "treatment" stems exclusively from the Huai River Policy.

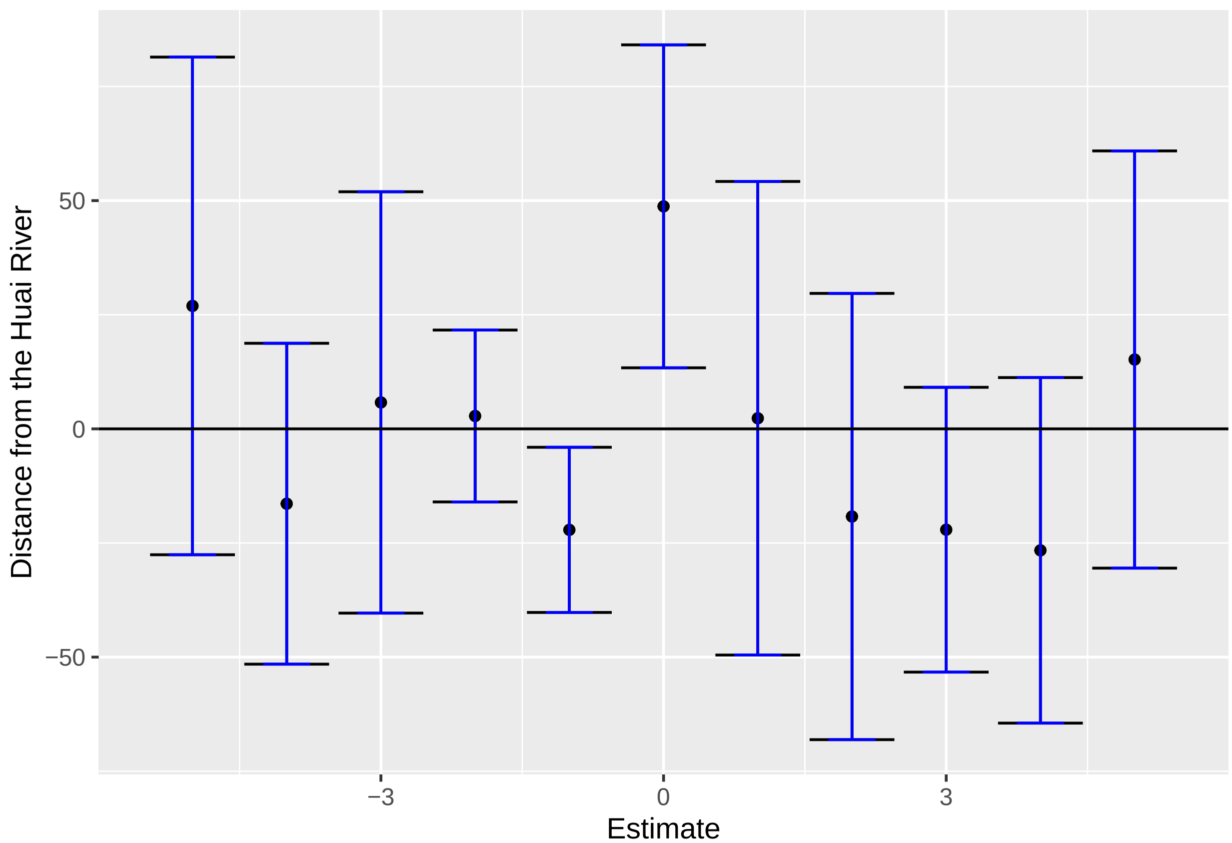



```
## $Estl
## Call: lpdensity
##
## Sample size                                      82
## Polynomial order for point estimation    (p=)    1
## Order of derivative estimated            (v=)    1
## Polynomial order for confidence interval (q=)    2
## Kernel function                                  triangular
```

```
## Scaling factor                                    0.50625
## Bandwidth method                                  user provided
##
## Use summary(...) to show estimates.
##
## $Estr
## Call: lpdensity
##
## Sample size                                       79
## Polynomial order for point estimation    (p=)     1
## Order of derivative estimated            (v=)     1
## Polynomial order for confidence interval (q=)     2
## Kernel function                                   triangular
## Scaling factor                                    0.4875
## Bandwidth method                                  user provided
##
## Use summary(...) to show estimates.
##
## $Estplot
```



In the context of this study, manipulation could be a potential concern for those living immediately North and South of the Huai River. Given the similar environments, regional culture, local language, and short distance needed to travel, there is at least a non-zero chance that some form of manipulation may have occurred. Thus, a check for manipulation will be useful to ensure "compliance" in the regression discontinuity.



The placebo test helps indicate the results from the Huai River policy are impactful. By comparing other areas north and south of the Huai River, the economists determine that PM10 and Life Expectancies express confidence intervals that include 0, which means the results are not statistically significant.

The interpretation for the "placebo effect" states that the emphasis of the Huai River comes from the Huai River being used as the line to demarcate the subsidized heating policy. Here we can see that other locations have no impact on our estimates given a change in distance from the Huai River because the confidence intervals include 0, making them statistically irrelevant.

In essence, if the Huai River policy was moved 1 degree south, we would see the same effect in that location, rather than at the Huai River.

In the context of Ebenstein et al, the estimate here could specifically be identified as the discontinuity. Additionally, the RDs for this project are all sharp tests of the policy. This is established by 100% "compliance". For this example, everyone North of the Huai River received and "complied" with the treatment, and everyone South of Huai did not receive treatment and "complied" with the control. Translated, it indicates that all participants' treatment status changes at the cutoff, with 0 observations not "complying" with the assigned outcome.

The Local Average Treatment Effect estimates the causal effect of the treatment for the subpopulation of compliers, i.e., those who would change their behavior based on the treatment assignment. In turn we can say that the effect of PM10 on health is a local treatment effect based on the nature of compliance. However, we should proceed with caution based on the regression formats being different than the observed regression discontinuity.