Montana Wildfires

Miles D. Williams

2023-02-23

Background

The state of Montana has a problem. Forest fires are a regular seasonal occurrence. To limit the spread and damage caused, the Forest Service offers to inspect the properties of individuals to ensure they are "fire proofed." If more people take advantage of this service, forest fires wouldn't be eliminated, but they would spread less intensely and pose a lower risk of damage and loss of life. Unfortunately, this service is voluntary and estimates suggest that less than 1% of the population take advantage of it. So, in conjunction with the Forest Service, we've conducted a randomized controlled trial (RCT) in the form of behaviorally informed letters intended to highlight the risks individuals pose to their neighbors by not ensuring that their property is fire proofed. We want to know, did the letters increase applications for property inspections?

Design of the Experiment and Expected Data

Households in two counties were randomly assigned to either treatment or control. Theoretically, each household, regardless of characteristics, had an equal chance of getting a letter.

For each household, we have the following data:

- `county`: a county identifier
- `hhincome`: household income
- `acres`: the size of the property in acres
- `urban`: an indicator for whether the property is urban or rural

We additionally will have information about which households received a letter and which submitted an application for an inspection. These variables will be `letter` and `app` respectively.

Analysis

Analysis Plan

Will estimate the effect of letters sent to households using the estimator specified by Lin (2013). I will specify my regression model like so:

```
eq <- app ~ letter * (ccounty + curban + chhincome + cacres)
```

In the above, each of the covariates interacted with letter will be mean centered.

I will then estimate this model using the `lm_robust()` function from the `{estimatr}` R package using HC2 standard errors. Since these errors are randomization justified, they should provide appropriate randomization based inferences about the effect of the letters.

Analysis Implemented

Below I'm reading in the code for the analysis.

```
data_location <- here::here("DPR 201", "Data", "montana_wildfires.csv")
Data <- read_csv(data_location)
```

If we take a glimpse at the data, we can get a sense for its structure:

```
glimpse(Data)
```

```
## Rows: 3,000
## Columns: 13
## $ county_id    <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ county       <chr> "County A", "County A", "County A", "County A", "County A~
## $ county_size  <dbl> 1000, 1000, 1000, 1000, 1000, 1000, 1000, 1000, 1000, 100~
## $ county_urban <dbl> 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.~
## $ hh_id        <chr> "0001", "0002", "0003", "0004", "0005", "0006", "0007", "~
## $ urban        <dbl> 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, ~
## $ hhincome     <dbl> 301785.438, 273852.605, 39062.613, 89131.648, 7496.056, 2~
## $ acres        <dbl> 1.7256400, 3.3747046, 7.2634906, 2.9203985, 15.4241460, 3~
## $ letter       <dbl> 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ app0         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ app1         <dbl> 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, ~
## $ app          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ got_letter   <dbl> 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, ~
```

With the data, I first need to make mean-centered versions of the covariates:

```
cmean <- function(x) x - mean(x, na.rm=T)
Data %>%
  mutate(
    ccounty = cmean(county == "County A"),
    curban = cmean(urban),
    chhincome = cmean(log(hhincome)),
    cacres = cmean(log(acres))
  ) -> Data
```

Now I'll create my formula object:

```
eq <- app ~ letter * (ccounty + curban + chhincome + cacres)
```

Next, I'll estimate the model:

```
fit <- lm_robust(eq, Data, se_type = "HC2")
```

With this, we can now take a look at the results.

Results

The below table summarizes the output from the regression model.

The estimate on `letter` indicates that the effect of the treatment on inspection applications by households was approximately 4.1 percentage points ($p < 0.001$, 95% CI = [1.674, 6.517]). We can reject the null that the letters had no effect on applications.

We can visualize the substantive impact of letters relative to baseline. The below figure summarizes the percentage of household applications among the control group and calculates the estimated counterfactual

	Model 1
(Intercept)	0.010***
	(0.002)
letter	0.041***
	(0.012)
ccounty	−0.001
	(0.005)
curban	0.000
	(0.005)
chhincome	0.005*
	(0.002)
cacres	0.002
	(0.002)
letter:ccounty	0.001
	(0.021)
letter:curban	−0.054*
	(0.026)
letter:chhincome	−0.010
	(0.010)
letter:cacres	0.002
	(0.010)
R^2	0.021
Adj. R^2	0.018
Num. obs.	3000
RMSE	0.119

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 1: Statistical models

application rate due to receiving a letter. The letter would have led to a more than 5 percent application rate.