Online Appendix: Constructing Generalizable Geographic Natural Experiments

June 6, 2022

Contents

1 Appendix A: Covariates

As mentioned in the manuscript, we use data from Keele et al. (2015) to construct the generalizable local geographic natural experiment. We use five pretreatment covariates: age, gender, turnout in 2004 and 2006, and housing prices.

Age corresponds to the subtraction between 2000 and the year of birth. We use the year 2000 as the reference since the template will be constructed using data from the 2000 census. Gender is a binary indicator for male individuals. Turnout takes the form of two binary indicators of having voted in the 2004 and 2006 elections. Keele et al. (2015) provide a categorical variable for a voter's history, using fine instead of mean balance for that covariate to constrain its marginal distribution. In our application we use mean balance to simplify the method. Therefore, we transform the original categorical variable into two binary indicators of turnout in 2004 and 2006. Finally, housing values corresponds to the median value of houses sold within a 500-m radius of a voter's residence between the years 2006 and 2008.

2 Appendix B: Covariate Balance

Table A1 reports the differences in standard deviation units between the treated and control groups before matching. We can see that there is an imbalance for housing prices (i.e., the standardized differences are greater than 0.1). Tables A2, A3, A4, and A5 report the standardized differences between the treated and control group after using regular matching, the city as the template, the state as the template, and the country as the template. We can see that all of the matched samples are balanced with respect to the five pretreatment covariates of interest.

Table A1: Balance before matching

Covariates	Standardized differences
Housing prices	0.13
Turnout 2004	0.02
Turnout 2006	0.04
Male	0.06
Age	0.02

Table A2: Balance regular matching

Covariates	Standardized differences
Housing prices	0.05
Turnout 2004	0.05
Turnout 2006	0.05
Male	0.05
Age	0.04

Table A3: Balance template matching: city

Covariates	Standardized differences
Housing prices	0.05
Turnout 2004	0.00
Turnout 2006	0.00
Male	0.07
Age	0.00

Table A4: Balance template matching: state

Covariates	Standardized differences
Housing prices	0.00
Turnout 2004	0.00
Turnout 2006	0.00
Male	0.02
Age	0.00

Table A5: Balance template matching: country

Covariates	Standardized differences
Housing prices	0.00
Turnout 2004	0.00
Turnout 2006	0.01
Male	0.01
Age	0.00

3 Appendix C: City, State, and Country-level Characteristics

To ensure that our matched covariates are representative at the city, state and country level, we provide the average housing prices, median age, voter turnout rate, and proportion of males at the three levels stated above using data from various sources. Table A6 provides the data with their corresponding hyperlinks.[1] We use census data for age and gender, the Wisconsin Realtors association and US Census Bureau for housing values,[2] and data from the city of Milwaukee, the state of Wisconsin, and the US Census Bureau for turnout.

Table A6: Average estimates

Variables	City	State	Country wide
Age 2000 (years)	30.0	36.0	35.3
Housing Values 2007 and 2008 (dollars)	152,995.5	162,407.0	214,545.5
Proportion of males 2000 (percentage)	47.8	49.4	49.1
Turnout 2004 (percentage)	69.9	72.9	63.8
Turnout 2006 (percentage)	56.9	50.9	47.8

[1] Links to the various sources can be accessed by clicking on the numbers.

[2] We could not find data for housing values at the city and state levels for the year 2006. Therefore, we used an average housing value for the years 2007 and 2008.

4 Appendix D: Using Milwaukee and the United States as templates

Table A7 and A8 report the averages for the five covariates after template matching when using the city of Milwaukee and the country, respectively, as templates. They illustrate how the matched samples resemble the target population of interest.

Table A7: After template matching: city

Covariates	Treated	Control	City
Housing prices	152,798.70	150,599.90	152,996
Turnout 2004	0.72	0.72	0.70
Turnout 2006	0.54	0.54	0.57
Male	0.46	0.49	0.48
Age	30.86	30.87	30

Table A8: After template matching: country

Covariates	Treated	Control	Country
Housing prices	212,170.30	212,162.00	214,546
Turnout 2004	0.66	0.66	0.64
Turnout 2006	0.50	0.50	0.48
Male	0.48	0.48	0.49
Age	36.16	36.15	35.30

5 Appendix E: Using cardinality matching to construct a generalizable geographic natural experiment

In this section, we show how to use cardinality matching with the `designmatch` package for `R` to construct a generalizable geographic natural experiment. We borrow some explanations from (Visconti and Zubizarreta, 2018), where the authors explain how to implement a traditional cardinality matching procedure.

First, we load the packages `designmatch`, `foreign`, and `gurobi`. The package `gurobi` enhances the performance of `designmatch`, and is particularly useful when having large data sets such as in this case.[3]

```
> library(designmatch)
> library(foreign)
> library(gurobi)
```

Then, we load and sort the data by the treatment. The treatment indicator is given by `t_ind`.

```
> d = read.dta("Final-Data.dta")
> d = d[order(d$t_ind, decreasing=TRUE), ]
```

Next, we define the covariates we will use to obtain mean balance in `mom_covs`. Also, we define the moments of the target distribution in `mom_targets`. In this example, the target distribution is the state of Wisconsin, and `mom_targets` are the first moments of this distribution. We also define the moment balance requirements; in this case, to balance the means of the covariates up to 0.05 standardized differences between the matched groups and the template (or 0.1 standardized differences between the matched treated and matched control groups).

```
> t_ind = d$t_ind
```

[3] Details about how to install `gurobi` can be found here: https://cran.r-project.org/web/packages/prioritizr/vignettes/gurobi_installation.html.

```
> mom_covs = cbind(d$hprice_mean,d$turnout04,d$turnout06,d$male,d$age)
> mom_targets = c(162407,0.729,0.509,0.494,36)
> mom_tols = absstddif(mom_covs, t_ind, .05)
> mom = list(covs = mom_covs, tols = mom_tols, targets = mom_targets)
```

Then, we specify the solver parameters.

```
> t_max = 60*30
> name = "gurobi"
> approximate = 0
> solver = list(name = name, t_max = t_max, approximate = approximate,
+ round_cplex = 0, trace = 1)
```

And we find the optimal matches.

```
> out_1 = cardmatch(t_ind, mom = mom, solver = solver)
Building the matching problem...
Gurobi optimizer is open...
Finding the optimal matches...
Optimal matches found
```

We extract the indices of the matched treated and control units, and check the number of units that were matched.

```
> t_id_1 = out_1$t_id
> c_id_1 = out_1$c_id
> length(t_id_1)
[1] 9924
> length(c_id_1)
[1] 9924
```

Finally, we save the matched sample, which is what we will use to compute covariate balance and treatment effects (see replication files).

```
d_match_state = d[c(t_id, c_id), ]
```

6 Appendix F: Examples of geographic natural experiments

In the manuscript we mentioned different types of natural experiments based on geography such as Geographic Natural Experiments (GNE), Geographic Local Natural Experiments (GLNE), Geographic Regression Discontinuity Designs (GRDD), and Geographic Natural Experiments with Matching (GLNE + Matching). In this section, we provide relevant examples of studies published in prominent journal using these approaches.

Table A9: Examples of studies that exploit geographic natural experiments

Study	Substantive focus	Natural Experiment	Research Design
McCauley and Posner (2019)	Political environment and religious saliency	Political border between Cote d'Ivoire and Burkina Faso	GNE
Posner (2004)	Salience of political culture	Political border between Zambia and Malawi	GNE
McNamee (2019)	Colonial Rule and Salience of ethnicity	Colonial border in Namibia	GLNE
Bram and Munger (2022)	Economic interest and voting in secession	Isohyet line	GLNE
Morris and Miller (2021)	Polling place consolidation and turnout	Municipal boundary line	GRDD
Harvey and West (2020)	Discrimination Statutes and Public Accommodations	State boundary line	GRDD
Keele and Titiunik (2018)	All-mail voting and turnout	County boundary in city	GLNE + Matching

References

Bram C and Munger M (2022) Where you stand depends on where you live: county voting on the texas secession referendum. *Constitutional Political Economy* 33(1): 67–79.

Harvey A and West EA (2020) Discrimination in public accommodations. *Political Science Research and Methods* 8(4): 597–613.

Keele L and Titiunik R (2018) Geographic natural experiments with interference: The effect of all-mail voting on turnout in colorado. *CESifo Economic Studies* 64(2): 127–149.

Keele L, Titiunik R and Zubizarreta JR (2015) Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 178(1): 223–239.

McCauley JF and Posner DN (2019) The political sources of religious identification: Evidence from the burkina faso–côte d'ivoire border. *British Journal of Political Science* 49(2): 421–441.

McNamee L (2019) Indirect colonial rule and the salience of ethnicity. *World Development* 122: 142–156.

Morris K and Miller P (2021) Voting in a pandemic: Covid-19 and primary turnout in milwaukee, wisconsin. *Urban Affairs Review* : 10780874211005016.

Posner DN (2004) The political salience of cultural difference: Why chewas and tumbukas are allies in zambia and adversaries in malawi. *American Political Science Review* 98(4): 529–545.

Visconti G and Zubizarreta JR (2018) Handling limited overlap in observational studies with cardinality matching. *Observational Studies* 4: 217–249.