Hamed Nikookalam

PS3

Table of Contents

Part1 1

Part2 4

Part1

1. The primary goal of Fearon and Laitin's article is causal explanation. They seek to explain why civil wars have been so prevalent since 1945, why some states experience civil war while others do not, and whether this is due to post–Cold War changes, ethnic diversity, or structural conditions favoring insurgency (low income, weak state capacity, rough terrain, large populations). Forecasting and description are not their central aim.
   The authors clearly state these goals in the abstract and introduction and articulate them further through a set of hypotheses contrasting ethnic-diversity and grievance explanations with an insurgency-based theory. This makes it easy for readers to see what is at stake and what the paper is trying to show. A weakness is the sheer breadth of the project. The authors ask a very general question: what explains the global pattern of civil wars from 1945 to 1999? And test a large set of hypotheses in a single cross-national framework. This breadth is a strength in terms of ambition and scope, but it forces them to work at a very aggregate country–year level and to treat heterogeneous conflicts and mechanisms as if they were driven by the same structural variables. That limits how deeply they can probe the micro-level processes or differences across types of civil wars.

2. On the theoretical estimand, the article does not have just a single theoretical estimand. Instead, the authors lay out a set of 11 hypotheses (H1–H11), and in Lundberg's terms each of these hypotheses implicitly corresponds to its own theoretical estimand. All 11 share the same basic structure: the unit of analysis is a country–year; the outcome is the annual probability that a civil war begins in that country–year (according to their coding rules); and the target population is all independent states above their population threshold between 1945 and 1999. For each hypothesis, the theoretical estimand can be understood as the average difference in this yearly probability of civil war onset between country–years that would have one value of a key condition and country–years that would have another value of that condition, holding other relevant characteristics fixed in this population.
   In more concrete terms, the hypotheses about ethnic and religious diversity define estimands such as the average difference in annual onset risk between otherwise similar

country–years that would be more versus less ethnically (or religiously) fractionalized. The hypotheses about democracy, civil liberties, discrimination, and inequality define estimands like the average difference in yearly civil war onset probability that would arise if a country were more democratic, had fewer discriminatory language or religion policies, or had lower income inequality. The hypotheses about insurgency-favoring structural conditions—such as low income, rough or mountainous terrain, large population, political instability, new statehood, and oil exporter status—define estimands such as: the average change in a country–year's probability of civil war onset associated with being poorer rather than richer, more rather than less mountainous, larger rather than smaller, newly independent rather than long-established, politically unstable rather than stable, or oil-exporting rather than not, holding other factors fixed. Alongside these causal-type quantities, the paper also implicitly uses some descriptive theoretical estimands, such as the fraction of states at war in each year and the rate of civil war onset over time.

On empirical estimand, the estimands are defined in terms of the observed data and the logit models they estimate. Here it is useful to say that the empirical estimand is the way they measure those theoretical target quantities with the data they actually have. Concretely, they construct a binary outcome for each country–year indicating whether a civil war begins in that year and compile 6,610 country–years of data from 1945 to 1999. They then estimate logit models of civil war onset with lagged covariates such as GDP per capita, population, proportion mountainous, new statehood, political instability, oil exporter status, ethnic and religious fractionalization, discrimination, regime type, inequality, and region and time controls. For each predictor X, the empirical estimand is the change in the log-odds of civil war onset—and, by transformation, the implied change in the predicted probability of onset—associated with a one-unit change in X, conditional on the other regressors in this sample. They also derive empirical estimands in the form of predicted probabilities over multiple years for different combinations of income and ethnic composition.

All in all, the theoretical estimands live at the level of "what we want to know about the world" (how different conditions change the probability of civil war onset in the population of states), whereas the empirical estimands are "how we measure those quantities using the specific country–year data and the logit model" (the coefficients and marginal effects in the sample).

3. Their Fearon and Laitin's identification strategy is a selection-on-observables design using a cross-national country–year panel and logit regression with lagged covariates. Their empirical claim is that structural conditions that make insurgency feasible (low income,

weak states, rough terrain, large populations, political instability, new states, oil) are what really drive civil war onset, and that ethnic diversity and grievance variables do not. To support this claim, they assemble data on 6,610 country–years (independent states, 1945–1999) and model a binary outcome indicating whether a civil war onset occurs in that year. Their modeling strategy is to estimate logit regressions that include insurgency-feasibility variables and ethnic/grievance variables in the same specification. If, conditional on the full set of controls, the insurgency variables are strongly associated with higher onset risk while ethnic fractionalization, democracy, civil liberties, discrimination, and inequality are not, they interpret this pattern as evidence in favor of their theory. They reinforce this with robustness checks: alternative civil war codings, different death thresholds, exclusion of anticolonial wars, and models with region and time dummies or subsamples.

Unlike designs based on instrumental variables, regression discontinuity, or difference-in-differences, they do not exploit a quasi-experimental shock; instead, they rely on observational panel data and rich controls to argue that their empirical estimand is a reasonable approximation to the theoretical estimands.

4. Fearon and Laitin's findings are very convincing for the broad question they ask, even though the design is not "clean causal identification". Their country–year logit models consistently show that low income, large populations, rough terrain, political instability, new statehood, and oil dependence are strongly associated with civil war onset, while ethnic and religious fractionalization, democracy, civil liberties, discrimination, and inequality generally are not once income is controlled. This pattern is robust across many alternative codings, samples, and specifications, so the evidence strongly supports their comparative claim that insurgency-feasibility conditions matter more than ethnic diversity per se. However, because the strategy is observational "regression with controls," the coefficients cannot be taken as fully credible causal effects: they rely on a selection-on-observables assumption, and unobserved factors (state-building histories, regional security environments, external interventions, etc.) may still confound the relationships. The country–year logit model is a reasonable but coarse approximation to the real data-generating process, which operates through subnational, dynamic, and strategic mechanisms. The data themselves are standard and widely used, but key concepts—civil war onset, state capacity, insurgency conditions, and ethnic diversity—are measured with noise and indirect proxies. So, the estimates are best interpreted as strong, theory-consistent conditional associations rather than fully identified causal effects.

5. Yes. Even with identification limits, the paper powerfully reshapes how we think about civil war. It shows that structural conditions linked to state weakness and insurgency feasibility (low income, rough terrain, instability, new states, oil) are far more predictive of civil war onset than ethnic diversity or broad "grievance" measures. That pushes theory

and policy away from “ancient hatreds” stories and toward questions of state capacity, governance, and the conditions under which rebels can operate. So even if we treat the coefficients as strong correlations rather than clean causal effects, the study still offers a very influential and useful framework for understanding where civil wars are most likely to occur and how they might be prevented.

Part2

1. Code:

```
# 1. Read the data and create age

therm <- read.csv("thermometers.csv")

therm$age <- 2017 - therm$birth_year
```

2. Code:

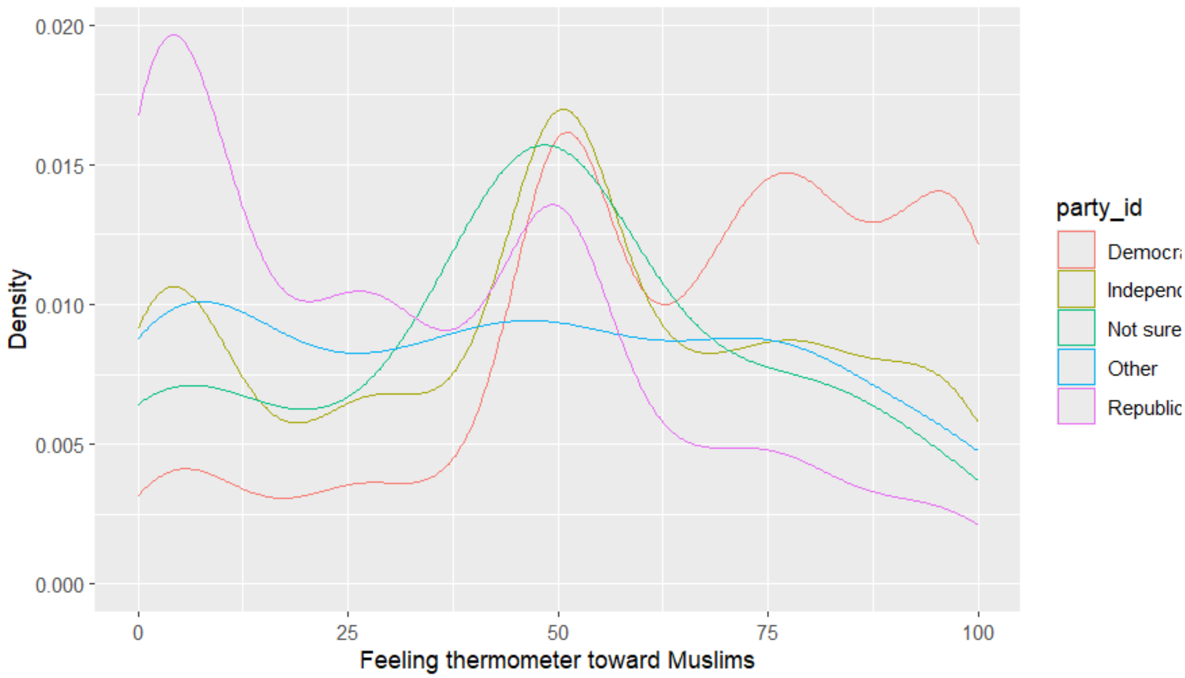
```
# Load packages once

library(dplyr)

library(ggplot2)


# 2. Describe ft_muslim by party_id

summary(therm$ft_muslim)

sd(therm$ft_muslim, na.rm = TRUE)


therm %>%

  group_by(party_id) %>%

  summarize(

    mean   = mean(ft_muslim, na.rm = TRUE),

    median = median(ft_muslim, na.rm = TRUE),

    sd     = sd(ft_muslim, na.rm = TRUE),

    n      = n()

  )


ggplot(therm, aes(ft_muslim)) +

  geom_histogram(na.rm = TRUE) +

  labs(x = "Feeling thermometer toward Muslims",
```

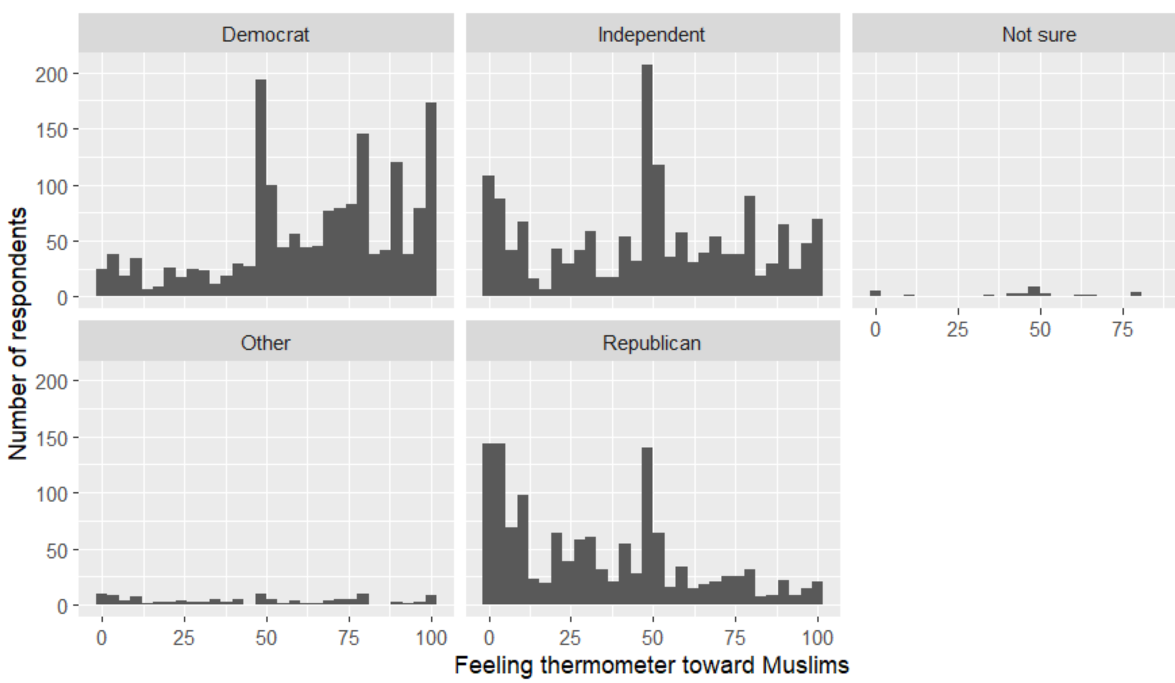
```
      y = "Number of respondents")


ggplot(therm, aes(ft_muslim)) +
  geom_histogram(na.rm = TRUE) +
  facet_wrap(~ party_id) +
  labs(x = "Feeling thermometer toward Muslims",
      y = "Number of respondents")


ggplot(therm, aes(ft_muslim, color = party_id)) +
  geom_density(na.rm = TRUE) +
  labs(x = "Feeling thermometer toward Muslims",
   y="Density")
```

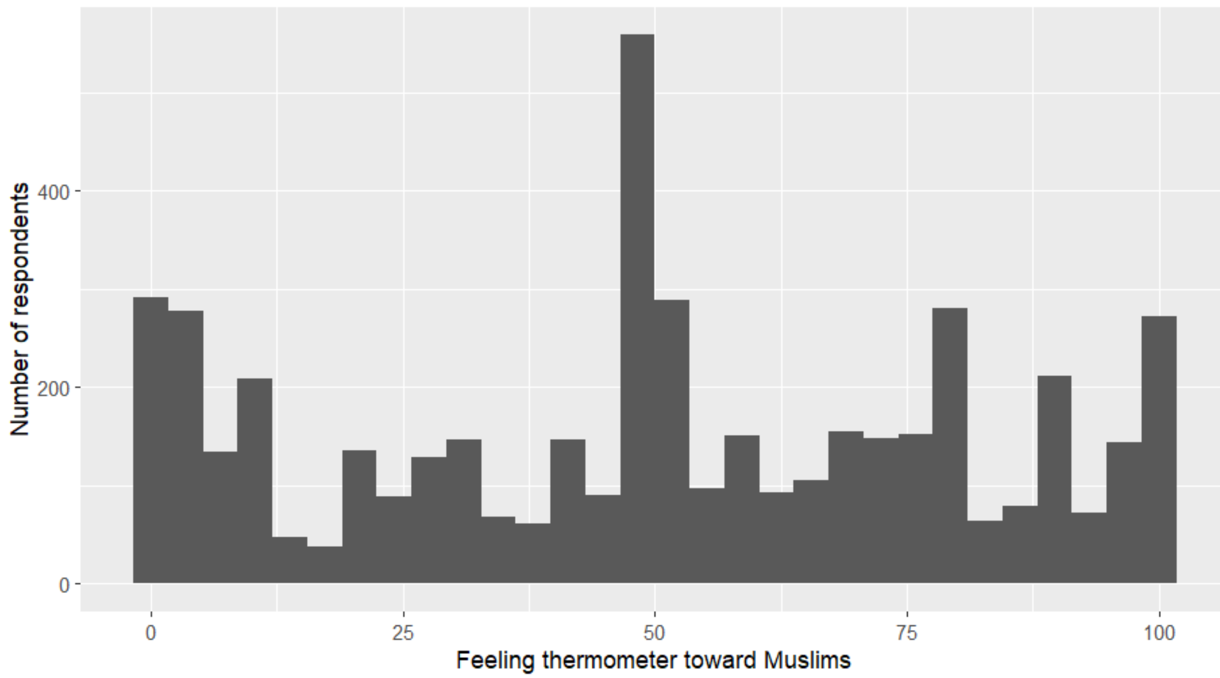
Output:

```
   Min. 1st Qu.  Median    Mean 3rd Qu.
   0.00   25.00   50.00   50.01   76.00
   Max.    NA's
 100.00     257
> sd(therm$ft_muslim, na.rm = TRUE)
[1] 30.85642
```

```
# A tibble: 5 × 5
  party_id     mean median    sd     n
  <chr>       <dbl>  <dbl> <dbl> <int>
1 Democrat     64.6   70    26.7  1734
2 Independent  49.0   50    30.0  1658
3 Not sure     47.1   50    28.4    55
4 Other        45.1   48.5  32.2   130
5 Republican   33.3   30    27.6  1412
```


Density
Feeling thermometer toward Muslims
party_id
Democrat
Independent
Not sure
Other
Republican
0.000
0.005
0.010
0.015
0.020
0
25
50
75
100


Democrat
Independent
Not sure
Other
Republican
Number of respondents
Feeling thermometer toward Muslims
0
50
100
150
200
0
25
50
75
100



For all respondents, the histogram shows that ratings toward Muslims cover the full 0–100 scale, with a big spike around 50 and many people also choosing very cold (near 0) or very warm (near 100) scores.

The density and faceted histograms by party_id show clear partisan differences: Democrats' distribution is shifted toward warmer ratings, Republicans toward colder ratings, and Independents/Other/Not sure are clustered more around the middle of the scale.

3. Code:

```
# 3. Regression: conditional mean of ft_muslim by party_id

therm$party_id <- factor(therm$party_id)

model_ft <- lm(ft_muslim ~ party_id, data = therm)

summary(model_ft)
```

Output:

```
Call:

lm(formula = ft_muslim ~ party_id, data = therm)


Residuals:

   Min       1Q     Median     3Q       Max

-64.605  -22.002   0.998    20.998    66.721
```

Coefficients:

```
                     Estimate Std. Error
(Intercept)           64.6053     0.6922
party_idIndependent  -15.6034     0.9920
party_idNot sure     -17.5428     4.1377
party_idOther        -19.5036     2.6923
party_idRepublican   -31.3266     1.0420
                    t value Pr(>|t|)
(Intercept)          93.330  < 2e-16 ***
party_idIndependent -15.729  < 2e-16 ***
party_idNot sure     -4.240 2.28e-05 ***
party_idOther        -7.244 5.05e-13 ***
party_idRepublican  -30.065  < 2e-16 ***
---
Signif. codes:
  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
  0.1 ‘ ’ 1
Residual standard error: 28.26 on 4727 degrees of freedom
  (257 observations deleted due to missingness)
Multiple R-squared:  0.1618,	Adjusted R-squared:  0.1611
F-statistic: 228.1 on 4 and 4727 DF,  p-value: < 2.2e-16
```

This regression estimates the mean Muslim thermometer score for each party. The intercept (about 64.6) is the mean for Democrats, and the other coefficients show that Independents, “Not sure,” “Other,” and especially Republicans have colder feelings toward Muslims (e.g., Republicans are about 31 points lower on average than Democrats).

4. Code:

```
# 4. Keep only Democrats and Republicans, create binary party
therm_dr <- subset(therm, party_id %in% c("Democrat", "Republican"))
therm_dr$party_bin <- ifelse(therm_dr$party_id == "Republican", 1, 0)
table(therm_dr$party_id, therm_dr$party_bin)
therm_dr$sex <- factor(therm_dr$sex)
```

Output:

```
               0    1
  Democrat  1734    0
```

Independent	0	0
Not sure	0	0
Other	0	0
Republican	0	1412

5. Code:

```
# 5. Linear probability model

model_party <- lm(party_bin ~ ft_muslim * sex + age, data = therm_dr)

summary(model_party)
```

Output:

```
Call:
lm(formula = party_bin ~ ft_muslim * sex + age, data = therm_dr)

Residuals:
    Min       1Q   Median       3Q 
-0.90774 -0.34463 -0.07362  0.33909 
    Max 
 1.00094 

Coefficients:
                    Estimate  Std. Error
(Intercept)        0.7008742  0.0427146
ft_muslim         -0.0075289  0.0003425
sexMale            0.0823752  0.0300320
age                0.0015961  0.0006024
ft_muslim:sexMale -0.0003974  0.0005110
              t value Pr(>|t|)

(Intercept)         16.408  < 2e-16 ***
ft_muslim          -21.984  < 2e-16 ***
sexMale              2.743  0.00613 ** 
age                  2.649  0.00810 ** 
ft_muslim:sexMale  -0.778  0.43679    
---
Signif. codes:  
  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
  0.1 ‘ ’ 1

Residual standard error: 0.4295 on 2979 degrees of freedom
  (162 observations deleted due to missingness)
Multiple R-squared:  0.2531,	Adjusted R-squared:  0.2521 
F-statistic: 252.3 on 4 and 2979 DF,  p-value: < 2.2e-16
```

I model party_bin (1 = Republican, 0 = Democrat) as a function of the Muslim thermometer, sex, age, and an interaction between the thermometer and sex. I include ft_muslim because attitudes toward Muslims are plausibly related to party identification, age to capture generational differences, and sex because men and women often differ in their partisan patterns. The interaction ft_muslim * sex allows the association between the thermometer and the probability of being Republican to vary by gender.

6. They represent the change in the predicted probability of being Republican (party_bin = 1):

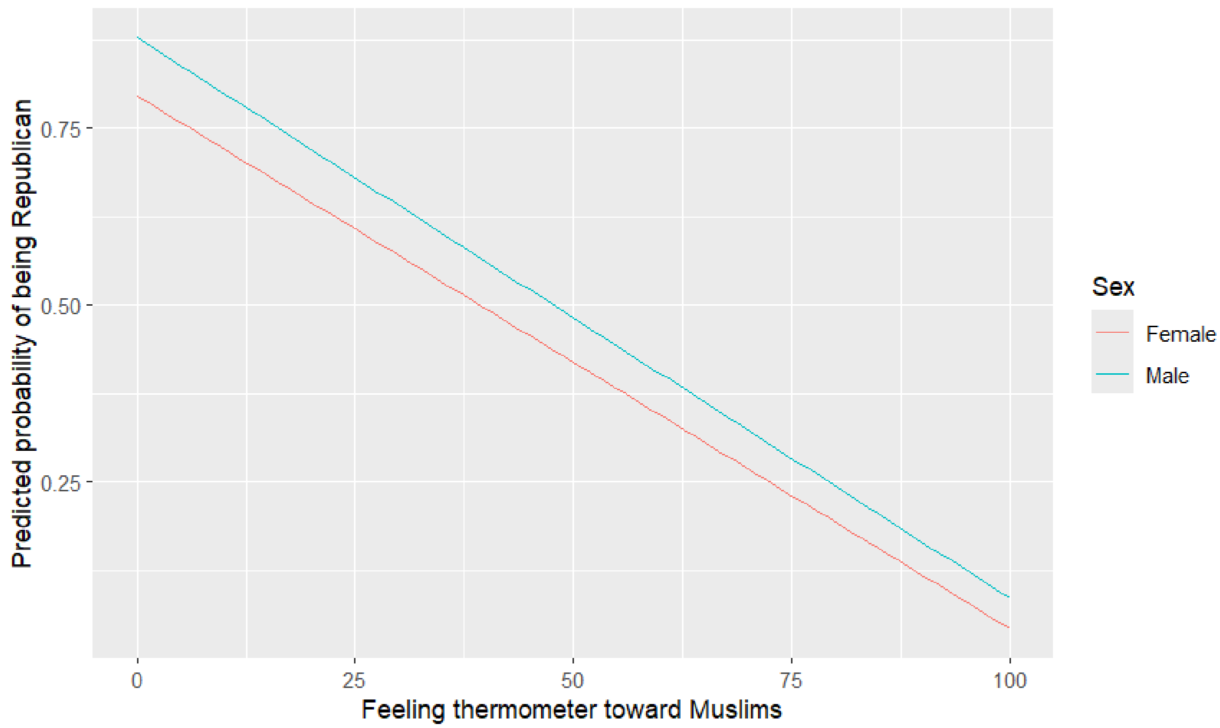
- For a numeric variable (ft_muslim, age): change in probability for a 1-unit increase.
- For a categorical variable (sexMale): the difference in probability between that category and the reference group.

7.

Code:

```
# 7. Predicted values as ft_muslim changes
mean_age <- mean(therm_dr$age, na.rm = TRUE)
newdata <- expand.grid(
  ft_muslim = seq(
    from = min(therm_dr$ft_muslim, na.rm = TRUE),
    to   = max(therm_dr$ft_muslim, na.rm = TRUE),
    length.out = 100
  ),
  sex = c("Female", "Male"),
  age = mean_age
)
newdata$pred <- predict(model_party, newdata = newdata)
ggplot(newdata, aes(ft_muslim, pred, color = sex)) +
  geom_line() +
  labs(
    x = "Feeling thermometer toward Muslims",
    y = "Predicted probability of being Republican",
    color = "Sex")
```

Output:



The plot shows that as the Muslim thermometer increases (people feel warmer toward Muslims), the predicted probability of being Republican declines sharply for both men and women: those with very cold ratings have a high predicted probability of being Republican, while those with very warm ratings have a much lower predicted probability. Men are slightly more likely to be Republican than women at the same thermometer value, which matches the positive coefficient on sexMale. However, this should not be interpreted as a causal effect of feelings toward Muslims on party identification, because the data are observational and party ID and unobserved factors (e.g., ideology, media consumption) likely also shape thermometer scores.