"Problem Set 3 (PS3)"

Joseph Kim

2025-12-12

Part 1: Paper Analysis 1. Research Goals

The primary goal of Fearon and Laitin (2003) is to explain the causes of civil war onset across countries and over time. In particular, the authors seek to identify the structural and political conditions under which civil wars are more likely to begin.

The study is best understood as aiming at causal explanation rather than mere description or prediction. The authors explicitly challenge prevailing explanations that attribute civil war primarily to ethnic diversity, ethnic grievances, or the end of the Cold War. Instead, they argue that conditions favoring insurgency—such as low per capita income, weak state capacity, rough terrain, and political instability—are the central causes of civil war onset.

While the article's theoretical ambition is causal, the empirical strategy relies on observational cross-national data and multivariate regression rather than modern causal inference designs such as difference-in-differences or matching. As a result, the authors do not claim to identify causal effects in a strict experimental sense, but instead present evidence consistent with a causal interpretation based on statistical control and theory-driven modeling.

A key strength of the paper is that the authors clearly articulate their research objective and directly engage with conventional wisdom, demonstrating that factors such as ethnic diversity and the end of the Cold War are far less influential than commonly believed once state capacity and insurgency-related conditions are taken into account.

2. Estimands

The theoretical estimands in Fearon and Laitin (2003) are the causal effects of various political, economic, and structural conditions on the likelihood of civil war onset. Each hypothesis articulated in the paper implies a distinct theoretical estimand, defined as the effect of a given factor on the probability that a country experiences the onset of civil war in a given year.

For example, the paper theorizes about the effect of ethnic and religious diversity, political regime type, discriminatory state policies, income inequality, and especially conditions favoring insurgency (such as low per capita income, rough terrain, political instability, and weak state capacity) on civil war onset. In theoretical terms, these estimands are counterfactual in nature, comparing the risk of civil war under different values of these conditions while holding other factors constant.

The empirical estimands correspond to how these theoretical quantities are operationalized in the data. Specifically, the authors estimate logit models in which the dependent variable is a binary indicator of civil war onset in a country-year. The empirical estimands are the changes in the log-odds (and, by implication, the probability) of civil war onset associated with changes in the independent variables included in the regression models.

The regression coefficients reported in the tables are estimators of these empirical estimands. While they provide evidence about the direction and magnitude of associations, the link between the empirical and theoretical estimands relies on strong assumptions about model specification and the absence of omitted variable bias.

3. Identification Strategy

Fearon and Laitin (2003) rely on an observational identification strategy based on statistical control. The authors estimate multivariate logit models in which civil war onset is regressed on a set of theoretically motivated covariates intended to capture both competing explanations (e.g., ethnic diversity, democracy, grievance) and conditions favoring insurgency.

Identification of the empirical estimands rests on the assumption that, conditional on the included control variables, the remaining association between each explanatory variable and civil war onset approximates the corresponding theoretical effect. In other words, the strategy assumes that the model sufficiently accounts for major confounders using observed covariates.

The paper does not employ design-based causal inference methods such as experiments, difference-in-differences, or matching. As a result, causal interpretation depends on strong assumptions about correct model specification and the absence of important omitted variables.

The authors strengthen their argument by examining marginal effects and by testing whether key relationships—especially the effect of per capita income—persist across alternative samples and model specifications. These robustness checks provide supporting evidence but do not establish causal identification on their own.

4. Assessment of Findings

Fearon and Laitin (2003) present strong empirical evidence that factors associated with insurgency—such as low per capita income, political instability, rough terrain, and large population—are robustly correlated with civil war onset. In contrast, commonly cited explanations based on ethnic diversity or grievance perform poorly once these factors are taken into account.

However, the identification strategy only partially supports the strength of the authors' causal claims. While the paper provides an intuitive narrative linking state weakness and insurgency to civil war, the theoretical mechanisms connecting some covariates to civil war onset are not always fully developed.

The paper also advances a large number of hypotheses, which risks diluting theoretical focus. In addition, the empirical models largely exclude interaction terms, potentially overlooking how combinations of factors jointly shape civil war risk.

The most prominent weakness of the article is its reliance on a broad multivariate regression strategy that assumes causal interpretation through statistical control alone. As a result, the regression coefficients are more credibly interpreted as conditional associations rather than definitive causal effects.

5. Broader Contribution

Despite its methodological limitations, Fearon and Laitin (2003) makes an important contribution to the study of civil war. The article demonstrates that several factors—particularly low per capita income, political instability, rough terrain, and large population—are strongly and consistently correlated with civil war onset.

Most notably, the paper challenges the widely held belief that ethnic and religious diversity are primary causes of civil war. By shifting the focus away from cultural explanations and toward state capacity and insurgency-related conditions, the article reshaped the civil war literature and set the agenda for subsequent research.

Part 2, Question 1: Create age variable

I examine the distribution of the feeling thermometer toward Black Americans (ft_black) both overall and by sex. Overall, the distribution is skewed toward higher values, indicating generally warm attitudes. The mean and median are relatively close, suggesting limited skewness.

When broken down by sex, women report slightly warmer feelings toward Black Americans on average than men. The overall spread of the distribution is similar across groups, although men exhibit slightly greater dispersion.

```
# load data + create age
thermometers <- read.csv("data/thermometers.csv")
thermometers$age <- 2017 - thermometers$birth_year
```

Part 2, Question 2: Descriptive statistics

```
summary(thermometers$ft_black)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   51.00   76.00   71.33   91.00  100.00     131
```

```
sd(thermometers$ft_black, na.rm = TRUE)
```

```
## [1] 23.98103
```

```
aggregate(ft_black ~ sex, data = thermometers, mean, na.rm = TRUE)
```

```
##      sex ft_black
## 1 Female 74.34345
## 2   Male 68.03488
```

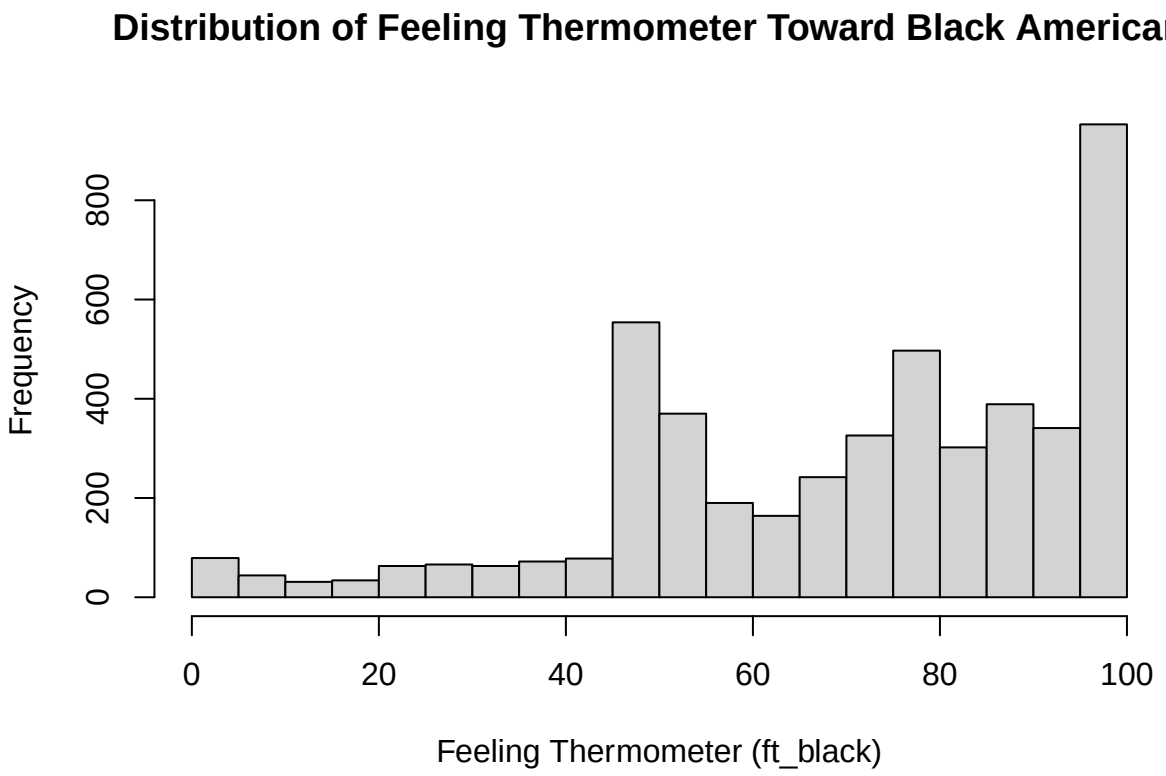
```
aggregate(ft_black ~ sex, data = thermometers, median, na.rm = TRUE)
```

```
##      sex ft_black
## 1 Female       80
## 2   Male       72
```

```
aggregate(ft_black ~ sex, data = thermometers, sd, na.rm = TRUE)
```

```
##      sex ft_black
## 1 Female 23.33186
## 2   Male 24.25076
```

```
hist(thermometers$ft_black,
breaks = 20,
main = "Distribution of Feeling Thermometer Toward Black Americans",
xlab = "Feeling Thermometer (ft_black)",
col = "lightgray")
```



Part 2, Question 3:Conditional Mean Regression

To estimate the conditional mean of ft_black by sex, I fit a linear regression with sex as a categorical predictor. This model estimates differences in average feelings toward Black Americans between men and women.

```
model_q3 <- lm(ft_black ~ sex, data = thermometers)
summary(model_q3)
```

```
##
## Call:
## lm(formula = ft_black ~ sex, data = thermometers)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -74.343 -18.035   4.965  20.965  31.965
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  74.3435     0.4721 157.466   <2e-16 ***
## sexMale      -6.3086     0.6829  -9.238   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 23.78 on 4856 degrees of freedom
##   (131 observations deleted due to missingness)
## Multiple R-squared:  0.01727,    Adjusted R-squared:  0.01707
## F-statistic: 85.34 on 1 and 4856 DF,  p-value: < 2.2e-16
```

Part 2, Question 4:Binary Party Identification

I restrict the sample to respondents who identify as Democrats or Republicans and create a binary party identification variable, where 1 indicates Democrat and 0 indicates Republican.

```
thermo_dr <- subset(thermometers, party_id %in% c("Democrat", "Republican"))
thermo_dr$party_bin <- ifelse(thermo_dr$party_id == "Democrat", 1, 0)
```

Part 2, Question 5:Multiple Linear Regression with Interactions

I estimate a multiple linear regression model predicting party identification using feelings toward Black Americans and interactions with sex, age, and education. This specification allows the association between racial attitudes and party identification to vary across demographic groups.

```
thermo_dr$sex  <- as.factor(thermo_dr$sex)
thermo_dr$educ <- as.factor(thermo_dr$educ)

model_q5 <- lm(
party_bin ~ ft_black +
ft_black:sex +
ft_black:age +
ft_black:educ,
data = thermo_dr
)
summary(model_q5)
```

```
##
## Call:
## lm(formula = party_bin ~ ft_black + ft_black:sex + ft_black:age +
##     ft_black:educ, data = thermo_dr)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8804 -0.4769  0.2197  0.4188  0.8271
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       1.729e-01  2.763e-02   6.258 4.44e-10 ***
## ft_black                          7.967e-03  7.131e-04  11.173  < 2e-16 ***
## ft_black:sexMale                 -1.346e-03  2.314e-04  -5.818 6.58e-09 ***
## ft_black:age                     -3.151e-05  8.909e-06  -3.538  0.00041 ***
## ft_black:educ4-year              -1.817e-05  3.619e-04  -0.050  0.95996
## ft_black:educHigh school graduate -1.148e-03  3.638e-04  -3.155  0.00162 **
## ft_black:educNo HS               -1.016e-03  9.548e-04  -1.065  0.28715
## ft_black:educPost-grad            1.560e-04  3.957e-04   0.394  0.69347
```

```
## ft_black:educSome college         -1.493e-04  4.077e-04  -0.366  0.71416
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4748 on 3058 degrees of freedom
##   (79 observations deleted due to missingness)
## Multiple R-squared:  0.09058,    Adjusted R-squared:  0.0882
## F-statistic: 38.07 on 8 and 3058 DF,  p-value: < 2.2e-16
```

Part 2, Question 6:Interpretation of Coefficients

Because the dependent variable is binary and the model is estimated using OLS, the coefficients represent changes in the predicted probability of identifying as a Democrat associated with a one-unit change in the independent variables, holding other factors constant.

The coefficient on ft_black indicates that warmer feelings toward Black Americans are associated with a higher probability of identifying as a Democrat for respondents in the reference categories. The interaction terms show that this relationship is moderated by gender and age, and to a lesser extent by education.

Part 2, Question 7:Predicted Values and Interpretation

I plot predicted values from the regression model as ft_black varies across its observed range, holding sex at the reference category, age at its sample mean, and education at the reference category.

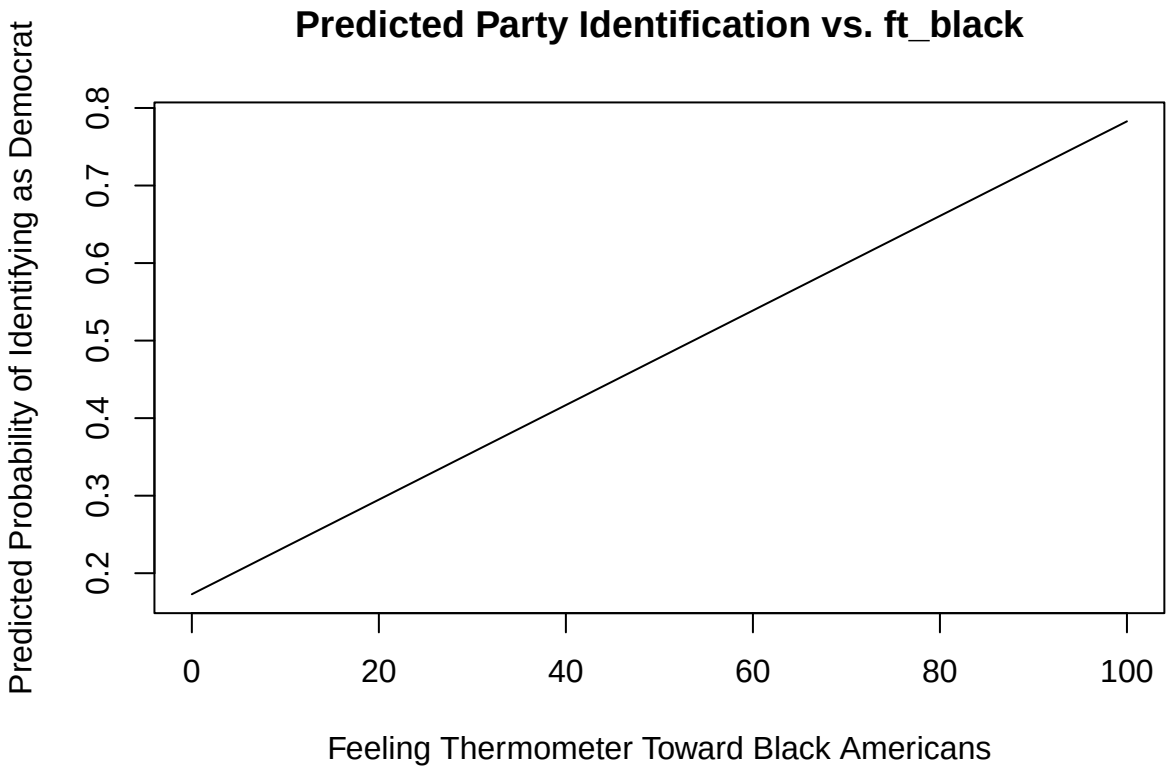
```r
sex_ref  <- if ("Female" %in% levels(thermo_dr$sex)) "Female" else levels(thermo_dr$sex)[1]
educ_ref <- levels(thermo_dr$educ)[1]
age_ref  <- mean(thermo_dr$age, na.rm = TRUE)

ft_seq <- seq(min(thermo_dr$ft_black, na.rm = TRUE),
max(thermo_dr$ft_black, na.rm = TRUE),
by = 1)

pred_data <- data.frame(
ft_black = ft_seq,
sex  = factor(rep(sex_ref, length(ft_seq)), levels = levels(thermo_dr$sex)),
age  = rep(age_ref, length(ft_seq)),
educ = factor(rep(educ_ref, length(ft_seq)), levels = levels(thermo_dr$educ))
)

pred_data$pred_party <- predict(model_q5, newdata = pred_data)

plot(pred_data$ft_black, pred_data$pred_party,
type = "l",
xlab = "Feeling Thermometer Toward Black Americans",
ylab = "Predicted Probability of Identifying as Democrat",
main = "Predicted Party Identification vs. ft_black")
```



The plot shows a strong positive association between feelings toward Black Americans and the predicted probability of identifying as a Democrat. However, this relationship should not be interpreted as causal. The analysis relies on observational survey data, and unobserved confounders and reverse causality may jointly influence racial attitudes and party identification.