Pset3

Gabriel Wang

November 5th, 2025

1 Part 1

1.1 Research Goals

The study aims at causal inference: try to find the what causes(s) the recent prevalence of violent civil conflict around the world.

The author firstly used a Wh-question to lead the following Yes–No questions which detailed specified all other the variables that the author in fact tried to argue against. Then the author used two Wh-questions again to grab readers' attentions and trigger curiosity.

The main strengths are its comprehensiveness, clarity of organization, and well-structured design.

The weaknesses lie in the lack of clarity in its research purpose, redundancy, and the potential confusion regarding which variables the author intends to test.

1.2 Estimands

Table 1: Theoretical and Empirical Estimands

	Original Description	Explanation / Improved Definition
Theoretical Estimand	The theoretical estimand is the causal effect of economic or systemic conditions, etc., on the onset of civil war in a country. The author has not sufficiently defined it.	The theoretical estimand refers to the causal effect of having versus not having certain economic or systemic conditions on a country's probability of civil war onset. In theory, it represents the difference in the expected rate of conflict between the same country under treated and untreated conditions. However, this counterfactual comparison is unobservable in reality.
Empirical Estimand	The empirical estimand is the causal effect of country income and ethnic homogeneity, etc., on probabilities of civil war onset. The author has not sufficiently defined it.	The empirical estimand is operationalized by comparing countries that experienced civil war with those that did not, using a Probit regression framework. It measures the difference in the expected probabilities of civil war onset between treated and untreated cases as observed in the data, serving as an empirical approximation to the theoretical estimand.

1.3 Identification Strategy

- Elaborate series of primary criteria to define violent civil conflicts address the ambiguity in conceptualization and ope-rationalization of civil war

- Analyze the data both with and without the anticolonial wars to ensure robustness.

- Include Prior war as a control variable to eliminate the effects from previous years when discussing the reasons of onset of civil wars

- Testing the impacts of income while controlling ethnicity, and vice versa, to examine the causal effects. Examination of other variables such democracy and freedom are also regressed and addressed in a similar way.

1.4 Assessment of Findings

Provide an overall assessment of the paper and its conclusions. Does the identification strategy support the authors' claims? For example, could the regression coefficients be credibly interpreted as causal effects if causal inference is the goal?

- I think the support is limited. Although adding variables of economy, ethnicity, polity, etc will help exploration of the correlations that the author is interested in, there exists multicollinearity. For example, the geographical continuity is seen to be highly relevant to economic development. But in the regression of this article, the author seems not to mitigate the possible multicollinearity.

- The regression model does not interact Ethnic fractionalization and Per capita income to examine the H_2: The effect of ethnic diversity on the probability of civil war should increase at higher levels of per capita income, which is supposed to explore under certain economic development level, what will the relation between ethnic diversity and probability of civil war looks like. The sample shortages fall into other interactions of variables that are talked in the series of hypothesis.

Does the model adequately represent the real-world data-generating process?

- With Huntington's idea taken, relation between the economic growth and civil war may not be linear. And many cases show that domestic instability and chaos happened in a time period that the country's economy grows. So I would expect another quadratic regression that uses squared term of economic variable.

- I did not see no country-specific intercept. The model does not include country fixed effects; therefore, it does not account for unobserved, time-invariant characteristics across countries.(Such as the impacts of countries' own colonization history towards institutions and therefore civil wars)

Does the data credibly measure the phenomena being studied?

- Per capita income does not necessarily represent the modernization of a country. For example, a leader could use oil export to make extra revenue to buy winning coalitions and suppress the disobedient, causing a phenomena of declining of civil wars. But the money does not flow into the common people. So other measurement about economy can be used in the regression as a placebo test.

- Classification of democracy, Anocracy, and autocracy is probably to arbitrary to measure the institutions of countries around the world. A continual measurement of polity such as Winning Coalition from Selectorate Theory could be used to see whether the result will be different.

- Using oil export to measure Primary Commodity Exports is somehow insufficient because there are so many genres of Primary Commodity, which requires different labor participation and political state capacity. For example, mining typically requires a larger labor force and stronger organizational capacity than oil extraction.

1.5 Broader Contribution

- Its main contribution is to challenge the ever-dominant and mainstream opinion that end of Cold War, ethnicity, religious fractionalization cause the civil war. Even if we could not conclude that they do not have effects on civil war, we should realize there are not absolute causal relations between these variables and onset of civil wars; economic variables and inclusive economic development still matter in domestic political orders and peace.

- This article highlights the distinctions between the duration and onset of civil wars, anti-colonial wars, and "pure" civil wars, suggesting that researchers should be attentive to these differences when selecting and measuring variables.

2 Part2

Question 1

```r
##Use the birth_year variable to create a new age variable
data$age <- 2017-data$birth_year
```

Question 2

About the central tendency

```r
colnames(data)
mean(data$ft_dem,na.rm = TRUE)

median(data$ft_dem,na.rm = TRUE)
```

The median feeling for the Democratic Party is 51. The mean feeling for the Democratic Party is 49.37996
I chose "party_id" from the categorical demographic. variables.

About the central tendency: for democratic

```r
mean(data$ft_dem[data$party_id == "Democrat"],na.rm = TRUE)

median(data$ft_dem[data$party_id == "Democrat"],na.rm = TRUE)
```

The median feeling for the Democratic Party among people who are Democratic is 84. The mean feeling for the Democratic Party among people who are Democratic is 79.85251

About the central tendency: for republican

```r
mean(data$ft_dem[data$party_id == "Republican"],na.rm = TRUE)

median(data$ft_dem[data$party_id == "Republican"],na.rm = TRUE)
```

The mean of feeling for demoractic party of people who are Republican is 21.49618; the median of feeling for demoractic party of people who are Republican is 12.

About spread

```r
var(data$ft_dem,na.rm = TRUE)
sd(data$ft_dem,na.rm = TRUE)
```

The variance of feeling for demoractic party is 1201.352, the sdandard deviation is 34.66053.

About spread among democractic

```r
var(data$ft_dem[data$party_id == "Democrat"],na.rm = TRUE)

sd(data$ft_dem[data$party_id == "Democrat"],na.rm = TRUE)
```

The variance of feeling of democratic peoples for demoractic party is 389.3891, the sdandard deviation is 19.73294.

About spread among republican

```r
var(data$ft_dem[data$party_id == "Republican"],na.rm = TRUE)

sd(data$ft_dem[data$party_id == "Republican"],na.rm = TRUE)
'''The variance of feeling of Republican people
for demoractic party is 519.5677, the sdandard deviation is 22.79403'''
```

About the density plots

Histogram of ft_dem among Democrat
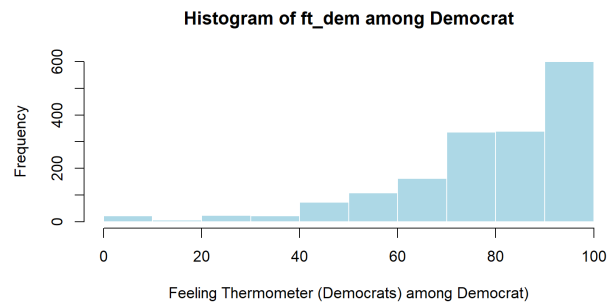
H

Figure 2: Histogram of ft_dem among Democrat

```
hist(data$ft_dem,
     main = "Histogram of ft_dem",
     xlab = "Feeling Thermometer (Democrats)",
     col = "lightblue",
     border = "white")
```
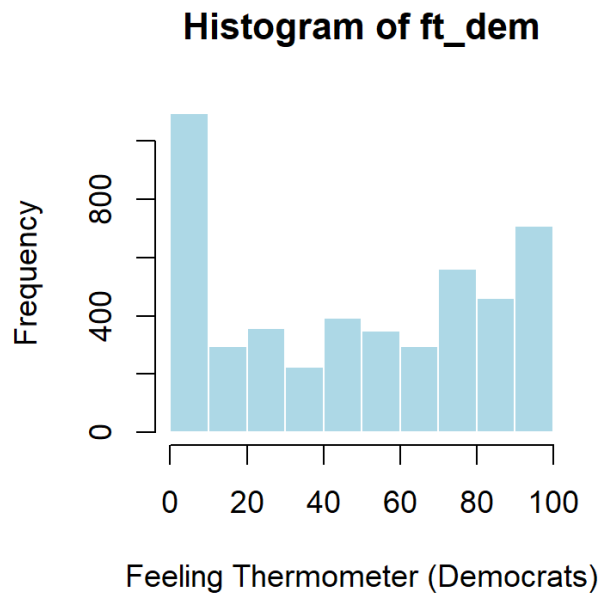

Figure 1: Histogram of ft_dem

```
hist(data$ft_dem[data$party_id == "Democrat"],
     main = "Histogram of ft_dem among Democrat",
     xlab = "Feeling Thermometer (Democrats) among Democrat)",
     col = "lightblue",
     border = "white")
```

```
hist(data$ft_dem[data$party_id == "Republican"],
     main = "Histogram of ft_dem among Republican",
     xlab = "Feeling Thermometer (Democrats) among Republican)",
     col = "lightblue",
     border = "white")
```
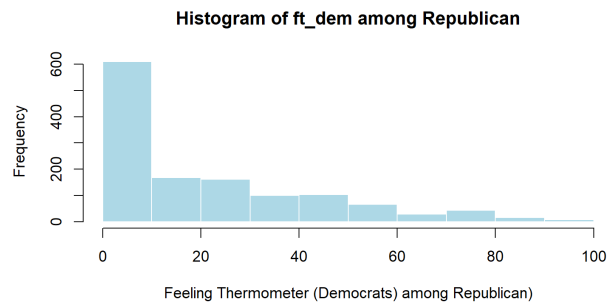
Figure 3: Histogram of ft_dem among Republican

Question 3

```
library(dplyr)
Model3 <- lm(ft_dem ~ party_id, data=data)
summary(Model3)
install.packages("stargazer")
library(stargazer)
stargazer(Model3,
          type = "latex",title="Model Question3",
          covariate.labels = c("Independent", "Not sure", "Other", "Republican"))
```

Table 2: Model Question3

	Dependent variable:
	ft_dem
Independent	−38.647***
	(0.875)
Not sure	−36.700***
	(3.736)
Other	−47.948***
	(2.409)
Republican	−58.356***
	(0.920)
Constant	79.853***
	(0.607)
Observations	4,740
R^2	0.480
Adjusted R^2	0.480
Residual Std. Error	25.003 (df = 4735)
F Statistic	1,092.990*** (df = 4; 4735)
Note:	*p<0.1; **p<0.05; ***p<0.01

Question 4

```
new_data <- data %>%
```

```
  filter(party_id %in% c("Democrat", "Republican")) %>%
  droplevels() %>%
  mutate(binary_party_id = if_else(party_id == "Republican", 1, 0))
```

Question 5

```
Model5 <- glm(binary_party_id ~ ft_immig + age + age:ft_immig , data=new_data,family
    = binomial(link = "logit"))
summary(Model5)
stargazer(Model5,
          type = "latex",title="Model Question5")
```

Table 3: Model Question5

	Dependent variable:
	binary_party_id
ft_immig	−0.022***
	(0.008)
age	0.021**
	(0.009)
ft_immig:age	−0.0002
	(0.0001)
Constant	0.465
	(0.522)
Observations	3,023
Log Likelihood	−1,833.775
Akaike Inf. Crit.	3,675.550
Note:	*p<0.1; **p<0.05; ***p<0.01

I estimate a logistic regression to predict the probability of being Republican using respondents' age, their feeling thermometer toward immigrants, and the interaction between the two. Older individuals and those less favorable toward immigrants are generally more likely to identify as Republican. The interaction term allows me to test whether age moderates this relationship. The logit specification ensures that predicted probabilities stay between 0 and 1 and allows interpretation in terms of odds ratios.

2.1 Question 6

The coefficient of age $(0.021**)$is statistically significant at the 5% level, suggesting that a growing age has a positive and significant effect on growing likelihood of being Republican.
The coefficient of ft_immig $(−0.022***)$is statistically significant at the 1% level, suggesting that a unit of increase in feeling towards immigrants(more positive feelings towards immigrants) will decrease $−0.022$ likelihood of being Republican, carrying a negative and significant effect.
The interaction term between ft_immig and age does not show significance, suggesting that the effect of immigration attitudes on Republican identification does not vary systematically by age.

2.2 Question 7

```
new_data$pred <- predict(Model5, new_data, type = "response")
ggplot(new_data, aes(x = ft_immig, y = pred)) +
  geom_point(alpha = 0.4) +
  geom_smooth(method = "glm",
```

```
                method.args = list(family = "binomial"),
                color = "green", se = TRUE) +
    theme_minimal() +
    labs(title = "Question 7: Relation between ft_immig and predicted binary_party_id",
         x = "Immigrant Feeling Thermometer",
         y = "Predicted Probability")
```
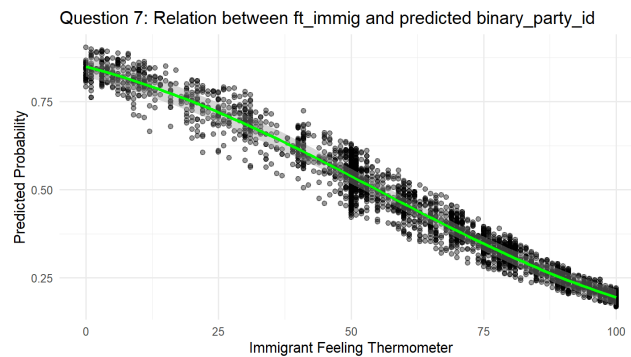

Figure 4: Question 7: Relation between ft_immig and predicted binary_party_id

The plot shows that feeling thermometer towards immigrants is negatively correlated with the likelihood of being a republican. More possitive feeling towards immigrants a person has, the less likely he is a republican. But we cannot say there is a causal effect. The first reason is that being a republican could also shape his or her feeling towards some issues with the example of immigrant issues. There are lots of literature suggesting that. The second reason is that: in this regression, we have just included few variables in the regression of explaning the likelihood of being a republican or democrat, with many confounding factors remained.