**Client 23-056 Hanna Sistek**

Meeting time: 09/10, Tuesday, 12:30 AM

Current stage: Analysis (All data has been collected)

Goal: Journal Publication / PhD dissertation

Supported by a grant or contract: No

**Background:**

* The client is a PhD student in Political Science who needs help analyzing their data and understanding their results pertaining to how disinformation and bias within media impacts the political polarization amidst various countries. Specifically analyzing the polarization gap from years before and after the Trump presidency. For the current data she is focusing on the years between 2018-2023.
* The specific dataset she is analyzing is from an external organization which rated the political polarization of 100 democratic majoritarian countries. They measured the media fractionalization as well through this study.
* The client wants to measure the effects of disinformation given the media fractionalization and polarization for each country in this survey.
* Originally, she came to the SCS as a returning client in Summer 2024 as she had technical issues with running her model across a panel regression model in R (plm()) as she mentioned that some of her colleagues and other researchers in political science used this method for similar projects, under the impression that her project was a longitudinal data analysis project. The issue in summer 2024 she was having was that her plm() model was dropping all the random effects in her model and only keeping the fixed effects. Using plm() was not discussed with the SCS before, hence for the majority of the summer 2024 semester, the consultant the SCS professors worked with the client to make sense of her project, understanding why she chose plm() for her analysis method, and realizing that the problems she is facing may by a more complex conceptual statistics problem than an analysis problem.
* A second follow up meeting was decided towards the end of the summer 2024 semester to make sense of the whole problem. But not without steering the client to look into the mixed effect model approach (lmer()), and group countries based off their respective slopes to analyze.

**Variables:**

Unit of Analysis: country-year

DV: Party dissemination of disinformation domestically (disinfo)

IVs: polarization, media fractionalization, electoral system(categorical variable), compulsory voting, rule of law, max level of populism speech on party level, aggregated to country level, country id.

Controls: level of democracy, inequality, internet use/social media use, net migration.

**Research Questions from Client:**

1. Politicians in democracies should care about reputation for re-election chances. So why risk it by disseminating disinformation?
2. What explains the global variation in disinformation dissemination domestically by political elites?

**Statistical Issues:**

1. The client isn’t that familiar with mixed / fixed effect regression and needs help understanding the findings.
2. From summer 2024, when she tried running her model across the countries, it assumed a single slope with different intercepts across the countries. She did not include interaction terms in her model, nor did she group the countries together based on slopes.

**Discussion points.:**

1. Need her to explain how she decided to originally choose the plm approach, understand what question she wants to solve, and get a more secure understanding of her project itself.
2. What variables in her project are fixed and random effects?
3. After the summer 2024 semester, did she reformulate her model to have interaction terms + grouping the similar slope countries.
   1. How did she group the countries together?
4. If she reran the lmer() model, what type of output did she get?
5. I understand that she chooses the 2018–2023-time span as this provides a more complete dataset, but since one of her aims was to understand disinformation before and after the Trump presidency, why didn’t she include the data before 2016 in her analysis?
6. Is her ultimate research question to show that the disinformation has increased since 2016?

**Professors Comments:**

Dr. Craig:

Sumeeth:

                plm() is simply a package that fits a set of linear mixed models to data where N experimental units are observed at the same K time points.  As you point out, it sounds like K=5 and N=100.

                The LMMs fit by plm assumed a constant slope over time and either fixed or random country-specific intercepts.  I provided code to you that also allowed country-specific slopes.  I suggested moving to another LMM package because of the limitations within plm()

                So we’ll see where she is at.  There are covariates she had at the country level and other covariates that are available at the country-year level.

Prof Craig

Dr. Gu:

There are ways to fit longitudinal data and various packages could be used to fit models in various situations, but I remain skeptical about the "doability" of  formal modeling in this project: different countries have different political systems, election cycles, and societal cultures, so I am not sure how to justify merging too many countries together plus aligning various calendars ...

CG