Top of Form

|  |  |  |
| --- | --- | --- |
| **Client:** | Hanna Sistek | **File:** 23-056 |
| **Dept:** | Political Science | **Faculty:**  **Student:** |
| **Date:** | 09-10-2024 | **Initial Meeting:**  **Follow-up:** |
| **Consultant and Attendees:** Hanna Sistek, Sumeeth Guda, Dr. Cherie Maestras, Dr. Bruce Craig | | |
| **Statement of Problem:**  To test a reputational cost theory of disinformation dissemination by political elites and to measure the impact of media fractionalization on the disinformation. | | |
| **Goal of this Project:** Journal Publication, PhD Dissertation | | |
| **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 2016. The time span of her current data she is focusing on the years between 2018-2023. Her data comes from the survey organization V-Dem which rated over 100 countries based on polarization and disinformation criterion for democratic governments. The expert raters rate the country’s polarization on a scale from [-5, +5] . From this data, the client developed a theory with respect to the reputational cost to answer the following research questions:   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?   Originally the client was continuing in the summer of 2024, and they needed help fitting their data using plm(). Specifically, she was having predictors being dropped from the model without explanation. Because plm() is a specialized mixed-effect program, we suggested using the more flexible LMM software for analysis. The methodologies and analysis techniques would be discussed during the fall semester.  The ultimate goal for the client is to create a model to test their theory about the tradeoff between reputational cost for politicians and increasing disinformation and political polarization to emerge as key actors in a majoritarian system. | | |
| **Progress of project at this time:** Analysis (All data have been collected). | | |
| **Relevant information presented at the meeting:**  The client informed everyone that out of the 100 or so countries that were rated, she is focusing on 70 of the countries that were scored for misinformation. Each of her countries are democracies, however most of them aren’t majoritarian-based systems (Winner takes all). The method by which they were scored was by having 5 expert raters in each country (Subject matter experts / Political Scientists) took a survey posed by the V-Dem organization and used Bayesian methods to rate the country’s polarization scale within [-5, +5]. An important disclosure the client brought up was that she is dealing with 2 constraints with the data. The first is a time constraint. For the purposes of her project, she is focusing on the time window from 2016 to 2023, because most of her countries don’t have data points before 2016. The second part is that her data are still being updated, with the most recent update happening in 03/2024. The client is staying in this time window, because any other time window is dropping the predictor variables (Mostly random effects). Her variables are the following:   * 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.   The client mentioned that she wanted to show that the reputational costs are being affected by political polarization and media fractionalization (To what degree news is covered in different ways). She would do this through running regression models through the 70 countries and analyzing the trends. She originally chose a plm() approach in the Summer of 2024 at the recommendation of one of her colleagues, however this was not cleared by the SCS and ultimately the project progress stalled until the Fall 2024 semester.  Broadly speaking the fixed-effects model is based on: *disinfo~polarization+media\_frac+polarization\*media\_frac*  One of the issues which she had was that when she introduced the control variables into her model with most of them being random effects, ultimately, they got dropped due to multicollinearity issues. This was an issue, because according to Dr. Maestas the control variables are necessary for Hanna’s model. A small concern Dr. Craig brought up was that some of Hanna’s predictors could change over time. As Dr. Maestas mentioned, the goal is to whittle down the time frame, control any alternative explanations reviewers will care about, and the model is presentable. In Hanna's case, she has a dependent variable in the predictor and wants to explore the relationships. However, there can be different relations within each of the countries. Dr. Craig couldn't tell which variables are changing by year and which ones are changing by country and region.    As said by Dr. Maestas regarding the background of their study: Elites are back casting to what they believed in 2016. There are a lot of differences in change between countries over time. She feels the change within the country over time is interesting to analyze for the project, however the data is based on cross-sectional differences compared to temporal differences. Since the main goal for the time being is to publish in a journal, Dr. Maestas’s concern is that a journal reviewer will discount the results of the model. They want to create a robust model which produces significant results for all time periods tracked with the survey.  As Dr. Craig said to Hanna with respect to the stratification, he told her to pick a specific year and to do the analysis on a year-by-year basis or across years.. If the variables don't change much over the years, he worries that the dataset is being considered bigger than it already is. Hence the results will not be valid and therefore we’d need to build in appropriate correlations to reduce the effective sample size.  Hanna wants to test her model across all of the countries in the study. She wants to make sure her theory works with the 70 countries under both majoritarian and non-majoritarian systems. The new model would be:  *lmer(Disinfo ~ polarization + media\_frac + media\_frac\*polarization), disinfo ~ [-5, +5]*  The response is smooth enough to consider as numeric.  As mentioned by Dr. Maestas, the raters are aggregating the score across the elite responses are based on the aggregate input of the scores. Hanna is assuming that the data are prepared in the best of the abilities of the elite raters, since she can't change the methods of collecting the data. Because of this, she needs to consider the error with the data. Which contributes to the noise of the model.  What Hanna and Dr. Maestas want to do for this semester is to finalize the model for analysis, as well as conduct robust diagnostic checks to validate the results. It is determined that they will use the mixed effect model for this project at the suggestion of the SCS, however R has a variety of mixed effect model libraries, however we need to find an appropriate mixed model. For the time being it was decided to use lmer() since there are built in diagnostics under this model.  When Hanna ran the model over the summer using only the fixed effects. The model assumed constant slope and constant variance across all the countries, with only the intercept differing between the countries. Dr. Craig brought up the idea that the mean structure of the model might not change with the political system, hence the constant slope and varying intercepts. By bringing country in as a random effect, this allows for the variance structure to change and ultimately the model can have different slopes between the countries. Additionally, by exploring the different political systems (Majoritarian, Non-Majoritarian) separately, ultimately the different variability can be explored. This was important because under the combined model with both of the systems, it assumed a constant variance. | | |
| **Recommendations for Analysis:**  The first task recommended to Hanna was to plot each of her 70 countries she is focusing on against each of the disinformation and polarization. This was recommended in part because she needs to explore the data: To answer what variables vary by country and what variables vary by year within country to see what model is appropriate for her data. The expected pattern is that the countries with flat slope lines can be labeled as non-majoritarian and those with increasing slopes can be labeled as majoritarian with respect to the government.    As recommended by Dr. Craig, because the mixed effect model will assume the same variation across the model for the fixed effects under a majoritarian system if the whole. It is important to compare the variance and covariance structures of 3 different mixed effect models (Majoritarian, Non-Majoritarian, Original Data Set). This can be done through splitting the data set by the “Majoritarian” variable and running the model through:  >>> lmer(disinfo~polarization + media\_fractionalization + polarization\*media\_fractionalization + polarization\*country + media\_fractionalization\*country)  The interaction terms are needed to add random effects to the model. They are important to add, because if the model only has fixed effects, then ultimately the slopes for all the countries will be assumed to be the same with different intercepts. Adding interactions changes the variability and theoretically will allow for different slopes for the different countries.  It was noted that in the meeting that lmer() is one of many R mixed effect model packages. lmer() was recommended by Dr. Craig as this is the popular approach for statisticians, as well as plm() is built on top of lmer() hence it is better to tap into the full lmer() rather than a niche application model. It is one of the consultant’s responsibilities to determine which mixed effect model approach is best for the client’s needs.  The client and consultant need to look at the fits of the model. The selection criteria (BIC, AICc) can be used to compare the models. In addition to plotting out the residuals and seeing if there are any patterns. | | |
| **Who will carry out these actions?**  Client:   * Split the data set into 3 separate entities (Majoritarian, Non-Majoritarian, Original) * Fit the variables of interest (disinfo, polarization, media fractionalization) which are the fixed effects, with the random effects (country, country\*polarization, country\*media fractionalization) into the mixed effect model. * Share the results with the consultant.   Consultant:   * Work with the client to determine which mixed effect model approach is best for their data. If there is continued trouble with R mixed effect packages, then potentially approach the problem with a SAS implementation where the mixed effect packages (PROC MIXED and PROC GLMMIX) are the only 2 packages. * Work with the client to help understand and compare the models against each other. | | |
| **Status:** Follow up meeting needed | | |

Bottom of Form