October 30, 2020

Social Network Analytics, Empirical Exercise #3
Due on Wednesday, Novenber 11, 2020 at 8:00am

Diffusion of political parties founded after extreme weather events in India

Setting up diffusion data: This exercise analyzes electoral data from each Lok Sabha—the lower house of Parliament in India—election from 1951 until 1999. These data are compiled at the district level. The electoral data is paired with meteorological data that details each district's monthly level of rainfall. We can use the occurrence of extreme weather events in a district as a proxy for economic disruption. We will use the data to answer the question of whether abnormal levels of rainfall during the time leading up to an election cause more political parties to enter into a district during that election. We will also use the data to find out whether there is a diffusion process of political parties entering into districts. A striking feature of this data is that the initial catalyst of political parties entering a region, rainfall, is random (or independent) within region across time, so the analysis is able to isolate and model the diffusion of political activity that is essentially stimulated by chance.

- The file "election_results.csv" contains a table with details about election results in each parliamentary district, compiled over each of the election periods. Not all districts exist for all of there election periods (some districts are created later on, and some districts dissolve). The data contain columns with information about the candidates, their districts, their political affiliations, and the districts:
  - "state" represents the geographic state in which the election takes place
  - "district" represents the parliamentary district in which the election takes place
  - "year" represents the election year
  - "name"represents the candidate's name
  - "gender" represents the candidate's gender
  - "age" represents the candidate's age in years
  - "party_name" represents the political party the candidate is running under
  - "party_issue" represents the type of issues and political identity of the candidate's political party
  - "party_scope" represents how broad or narrow is the geographic scope of of the political party
  - "vote_count" represents the number of votes each candidate earns in the election
  - "language" represents the primary language spoken in the district
  - "unrest" represents the number of events of political unrest, such as demonstrations, riots, and official protests that have occurred in a district in the interval leading up to the election
- The file "monthly_rainfall.csv" contains a table with monthly rainfall totals, measured in milliliters, for electoral districts in India, collected from meteorological stations around the country. The districts are coded to uniquely match the districts in "district_information.csv" without needing to refer to states.
- The file "border_information.csv" contains an edge list illustrating which districts share borders. Each row represents a border-sharing relationship between two districts, and this relationship is undirected.

1. First, we will set up the relationship between rainfall and political party diffusion, and then modify the rainfall measure to generate a statistically independent measure for droughts. This modification will allow us to isolate the effect of economic strain on political parties from other underlying features of a region that might influence its political structure.

   To work with rainfall in a way that is comparable across different regions across time, climate researchers use a measure called a Standardized Precipitation Index (SPI in the equation

below), which normalizes raw rainfall measures according to a the yearly historical average for the region. This transformation uses a Pearson type III distribution, which can be achieved in R through the `PearsonDS` library using the `qpearsonIII()` function.

For this transformation, the distribution's scale and shape will be equal to

$$scale = \frac{s_2}{\overline{x} - c}$$

$$shape = \frac{(\overline{x} - c)^2}{s_2}$$

where $\overline{x}$ will be the yearly mean across districts and s_2 will be the yearly variance across districts. You can assume that since the skewness of rainfall is positive, the location parameter c is equal to 0.

To get the Standardized Precipitation Index, normalize each district-year observation for each district i in year t:

$$SPI_{it} = \frac{rain_{Pearson\ III} - \overline{rain}_{Pearson\ III}}{s_{rain_{Pearson\ III}}}$$

where the mean and standard deviation, $s_{rain_{Pearson\ III}}$, are both calculated at the district level.

(A) Create a figure, for example, a scatter plot, showing the visual relationship between the average SPI in a district across each year from the prior election—or from the beginning of the data for the first election—leading up to the current election, and the number of political parties that are founded in a district. Consider the party to be "founded" if it has not competed in an election in this district prior to this election.

(B) Using the election-period level rainfall measures created above, show that the SPI is not independent from one election period to the next within a district, as well as from neighboring districts from one election period to the next. It is possible to show this relationship by regressing a district's current level of the rainfall variable on (1) its lagged value and (2) the lagged value of its neighbors' rainfall variable. For computing the neighbors' value, you can use an average of each of the surrounding districts' values. Include a control in the regression for the number of years in the election period, and use a fixed effects specification to control for the time-invariant features of a district as well as a control for each election period. This can be accomplished using the plm package, using a model specified in the form of `plm(outcome variable ~ predictor variables, data, effect = "twoways", model = "within", index = "district")`, where "twoways" "within" provide both sets of fixed effects.

(C) Climate scientists consider moderate droughts to occur if the Standardized Precipitation Index falls below -1, and moderate floods to occur if it rises above 1. Create a measure that sums the number of years a district experiences either moderate droughts or floods during the interval starting from the year following the previous election up until the year of the current election. Perform the same test as in (B), using this new transformed measure. This measure will form the basis for the predictors used in the remainder of the regressions in Questions 2-5. Since this is a count outcome that is reported as a discrete number of years, use a regression adopted for data of this form—this can be accomplished with the pglm package, using a model specified in the form of `pglm(outcome variable ~ predictor variables, data, effect = "twoways", model = "within", index = "district", family = "poisson")`. What differences do you see between the estimates?

2. Next, let's analyze whether there are more new political parties when droughts or floods occur in a district or in its neighboring districts.

(A) Run a regression predicting the number of new political parties that are formed as a function of the number of years a district experiences droughts or flooding in the interval starting from the year following the previous election up until the year of the current election. The number of new political parties that enter a district is a discrete count

outcome, so we should use a regression format adopted for counts, as in (1C). Include a control in the regression for the number of years in the election period, and a control for the time-invariant features of a district, as in Question 1. Are certain kinds of political parties, based on the issues they cater to, more likely to be formed when a district experiences extreme weather?

(B) Now that we have established the baseline effect, we can look at how political activity stimulated by droughts or floods in one district might affect political activity in another district.

Use a similar regression to (A) to show that, even when taking into account a district's own droughts and floods, that district's degree of political foundings will also depend on the number of years its neighboring districts experience years of droughts or flooding in the interval starting from the year following two elections ago, up until the year of the previous election—the election lead-up interval before the current one. Include a control in the regression for the number of years in the current election period, and a control for the time-invariant features of a district.

3. Extreme weather events like droughts or floods can erode the stability of political systems and wear away at the entrenched power bases of large, national-scale parties that have difficulty responding to the needs of affected regions.

   Perform a regression similar to Question 2B to determine whether experiencing droughts or floods relate to political concentration. The Herfindahl Index, or HHI, is a measure of political concentration that measures the degree to which a few parties command the majority of the shares of votes in each election:

$$Herfindahl = \sum_{i}^{n} (vote\ share_i)^2$$

   where the vote share is the count of votes divided by the total number of votes received by all candidates in the election in this district in this election period. The HHI can be computed with the `HHI` package using the `hhi()` function.

   What does this regression illustrate in terms of the HHI's concentration or fragmentation of political power in districts affected by extreme weather?

4. To understand the diffusion process more, we want to analyze whether the candidates that contest elections in a district are the same individuals that have contested elections in nearby regions in the past, or if brand new candidates are responsible for driving the diffusion process.

   To analyze this, run two separate regressions similar to Question 2B predicting the likelihood of (1) a new candidate contesting an election in a district, that has contested an election within a distance of two neighboring district in any previous election period, and (2) new political parties being founded in a district that have not contested an election in a neighboring district in any previous election period.

   What do these regressions illustrate about the diffusion of political organizing?

5. Last, we want to understand how political diffusion processes relate to the overall political trends in the country.

   One theory is that new parties are more likely to diffuse into areas that are more closely contested in terms of political competition, creating space for more issues to be discussed. To test this theory, run a regression predicting the likelihood of a new political party contesting an election in a district, that has contested an election in a neighboring district in any election period, as a function of the district's prior political concentration given by the HHI of the vote share of the issues covered by its political parties in "party_issue".

   Another theory is that new parties are more likely to diffuse into areas where a national party has suffered a recent loss, creating a power vacuum. To test this theory, run a regression predicting the likelihood of a new political party contesting an election in a district, that has contested an election in a neighboring district in any election period, as a function of the number of losses a party of national scope suffers in the district and its neighboring districts in the prior election.

   Does either one of the theories seem more plausible?