ISOM 673

Empirical Assignment 3

Diffusion of Political Parties Founded After Droughts or Floods in India, 1951-1999

Student:

Conner (Haotian) Jin

Under supervision of

Dr. Demetrius Lewis

Emory University

Goizueta Business School

Haotian Jin

Diffusion of Political Parties Founded After Droughts or Floods in India, 1951-1999

ABSTRACT

This exercise tests a theory of quotidian disruptions and political entrepreneurship. I argue that disruptions—either moderate droughts or moderate floods in a district—will influence the foundation of political parties. The theory also incorporates social influence between neighboring districts, that is, when a neighboring district experiences disruption, a focal district will result in a moderate increase in political parties' foundation. The theory is tested through a quasi-experiment design, analyzing electoral data from each Lok Sabha—the lower house of Parliament in India—elections from 1951 to 1999.

INTRODUCTION

The Lok Sabha or the House of the People is the lower house of India's bicameral Parliament. Members of Lok Sabha are elected by adult universal suffrage and a first-past-the-post system to represent their respective regions ("THE CONSTITUTION OF INDIA."). Each state is divided into territorial constituencies in a manner that the ratio between the population of each constituency and the number of seats allotted to it remain the same throughout the state. Political parties are founded before or during the election period, in preparation for competing in elections or advocating certain bills.

Empirical studies show that the formation of political parties is related to disruptive events, either in the territorial constituency or its neighboring constituencies, before the election period. To test this theory, we used electoral data from Lok Sabha elections from 1951 until 1999, paired with meteorological data that details each district's monthly level of rainfall. The data cleaning is performed by creating time windows as an id number for each election period, grouping by election period and district, and summarizing millimeter rainfall with total sum and Standardized Precipitation Index, or SPI with the average number.

After summarizing data, multiple regression model, including panel linear model, panel most likelihood estimation, and panel autoregressive model, is performed to test the following hypothesis:

- For a specific district, more political parties will be founded when droughts or floods occur during the election period.
- For a specific district, more political parties will be founded if its neighboring district experienced a high level of droughts or floods during the previous election period.
- Experiencing droughts or floods might relate differently to the entry and diffusion of political parties depending on their scope.
- For a specific district, Experiencing Droughts or Floods Decreases Political Concentration.
- Existed political parties diffuse across borders, while non-existed political parties influence by rainfall.

For the coefficient of each predictor variable in each regression, the level of statistical significance will be used to identify whether a relationship exists between that specific predictor variable and the outcome variable. This type of analysis will be the major technique throughout the study.

EXPLORATORY DATA ANALYSIS

(Question 1A) Visual Relationship between Level of Rainfall and Political Parties Foundation

The level of rainfall in a district is quantified by two different measures. The first measure is the raw rainfall measure represented in the units of millimeters; the other measure is the Standard Precipitation Index, which is a transformation used by meteorological scientists that normalizes the rainfall according to a historical average. The sum of raw rainfall and yearly average SPI during each election period is calculated (code in Appendix I). The following visualization briefly shows the relationship between the level of rainfall and political parties foundation within each district at a certain election period.

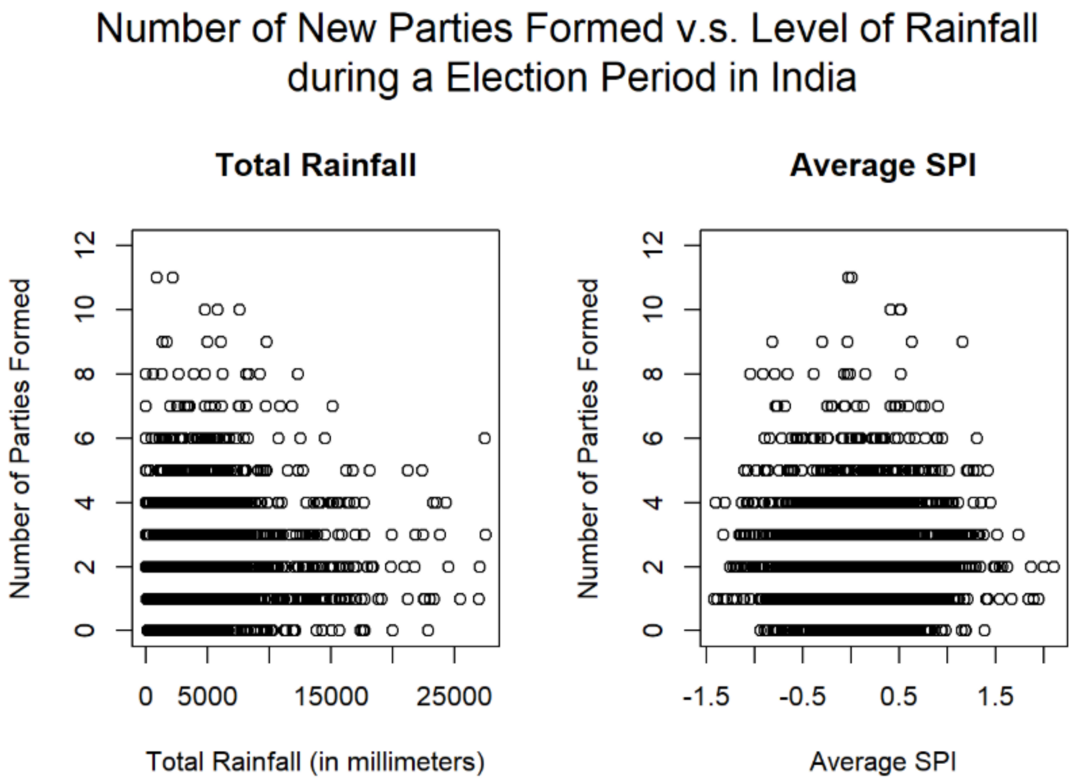

FIG. 1. Scatter plots of the level of rainfall and the number of new parties formed. On the left, the level of rainfall is represented by total raw rainfall (in millimeters) during an election period; On the right, the level of rainfall is represented by average SPI measure during an election period.

As we can see from FIG. 1., if we use total rainfall in millimeters as an indicator for levels of rainfall, we could see that when those districts who have the most number of political parties formation are those who have relatively lower level of rainfall; whereas when the rainfall is extremely high, there is a slight tendency that the number of parties formed would increase as well. However, the influence is not quite clearly purely from the visualization. This is because each data point in the scatter plot represents a district and each district might have different sizes and populations. A similar finding could be shown from the right graph, using SPI as an indicator of the rainfall level.

Therefore, to address this problem, panel linear model or other panel regression techniques would be a better choice for the study. To satisfy the assumption of linear regression, a statistically independent measure for droughts and floods is needed. This modification will also allow us to isolate the effect of economic strain on political parties from other underlying socioeconomic features of a region that might influence its political structure.

(Question 1B) Independence Test of Level of Rainfall from one Period to the Next within a District, as well as from Neighboring Districts from one Election Period to the Next

Hence, we perform a panel regression of a district's current level of the rainfall on its lagged-by-one-period value as well as the lagged value of its neighbor's rainfall. For measuring neighbor's level of rainfall, a border relationship edge list dataset is used to identify neighbors. Consecutively, an average measure is taken among all neighboring districts for each focal district, labeled as the level of rainfall from neighboring districts. In addition, since every election period window contains different numbers of years, a control for the number of years in a period is included. The panel linear model technique itself also specifies a control for the time-invariant features of a district as well as a control for each period.

The regression result is listed below: (Code for summarizing neighboring rainfall level and regression in Appendix II)

TABLE 1. Panel linear regression of level of rainfall in a district on its lagged value and its neighboring districts' average lagged value. All the coefficient results are significant for the four similar simple linear models.

Results

	Dependent variable:			
	total_rain		avg_spi	
	(1)	(2)	(3)	(4)
total_rain_lag1	0.323***			
	(0.016)			
mean_neighbor_rain_lag1		0.218***		
		(0.018)		
avg_spi_lag1			0.126***	
			(0.023)	
mean_neighbor_spi_lag1				0.074***
				(0.027)
year_diff	1,402.281***	1,294.241***	-0.025**	-0.030**
	(31.403)	(34.228)	(0.011)	(0.012)
Observations	2,275	2,017	2,275	2,012
R^2	0.502	0.442	0.016	0.006
Adjusted R^2	0.446	0.378	-0.095	-0.107
F Statistic	1,029.939*** (df = 2; 2044)	716.283*** (df = 2; 1809)	16.572*** (df = 2; 2044)	5.878*** (df = 2; 1805)

*Note: *p<0.1, **p<0.05, ***p<0.01*

From the panel linear model results, we could clearly see that all the linear models have a positive coefficient with strong statistical significance. This suggests the level of rainfall in a focal district **is highly likely not independent** from either its level of rainfall in the previous election period or neighboring districts' level of rainfall in the past, no matter using raw millimeter rainfall or SPI as a measure.

Therefore, due to a lack of independence, a modification is needed to generate a statistically independent measure for droughts and floods.

(Question 1C) Modification of Rainfall Measures

Meteorological scientists consider moderate drought to occur if the Standardized Precipitation Index falls below -1, and moderate floods to occur if it rises above 1. Therefore, a transformation is performed on the rainfall data set to derive the number of moderate extreme weather years during an election period. After that, the same procedure is repeated and the same regression test of the current count of extreme weather in a district on its lagged value and its neighbors' average lagged value is performed as in the previous session. This tests whether extreme weather counts yield an independent measure of rainfall level.

The regression result is listed below: (Code for summarizing neighboring moderate extreme weather counts and regression in Appendix III)

TABLE 2. Panel linear regression of count of extreme weather in a district on its lagged value and its neighboring districts' average lagged value. All the coefficient results are significant for the four similar simple linear models.

Results

	Dependent variable:	
	current	current
	(1)	(2)
extreme_count_lag1	-0.019	
	(0.024)	
mean_neighbor_extreme_weather_lag1		-0.003
		(0.031)
year_diff	0.348***	0.352***
	(0.017)	(0.018)

*Note: *p<0.1, **p<0.05, ***p<0.01*

From the panel linear model results, we could clearly see the biggest difference between the estimates is that the statistical significance of either moderate extreme weather count or the average moderate extreme weather count among neighbors of a focal district has lost. This is a good sign since the isolation of the effect of economic strain on political parties from underlying features of a district is needed.

The fact that the coefficient term for extreme weather count is not statistically significant confirms that: this is a measure that **is independent** from one election period to the next within a district, as well as from neighboring districts from one election period to the next. I could further take this measure to be a quantification of rainfall level because the independence makes the measure robust to use as a predictive variable in regression analysis.

REGRESSION ANALYSIS

(Question 2) More Political Parties will be Formed when Droughts or Floods Occur

After I found that the number of droughts or floods is a statistically independent measure, I performed a regression of the number of new parties founded on the number of droughts or floods. However, it is also likely that the rate of entry of political parties in any particular district in a particular period is “auto-related”, over time, to the rate of entry in the prior periods of this district’s history. Therefore, I used panelAR to take into account district-specific autocorrelation through generalized least square estimator. In detail, I specified “phet” as panel correlation method to consider district fixed effects and specified “psar1” as a correlation feature to control panel-specific autocorrelation. A linear control for each year and the number of years in each period is also included.

The results of the regression are shown below: (Code for the linear model of the number of new parties founded on the number of droughts or floods in Appendix IV)

TABLE 3.1 Panel Regression with Auto Correlation Paris-Winsten Corretion. The outcome variable is number of political parties founded and the predictor variables are counts of drought and flood, including control for number of years in each period and a linear control for each election year.

```
##
## Panel Regression with AR(1) Prais-Winsten correction and panel heteroskedasticity-robust standard errors
##
## Unbalanced Panel Design:
##  Total obs.:       2502 Avg obs. per panel 11.0708
##  Number of panels: 226  Max obs. per panel 14
##  Number of times:  14   Min obs. per panel 1
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   27.393711   5.122711   5.348 9.73e-08 ***
## extreme_count  0.105192   0.031466   3.343 0.000841 ***
## year_diff     -0.202111   0.027018  -7.481 1.02e-13 ***
## year          -0.012409   0.002551  -4.864 1.22e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-squared: 0.2635
## Wald statistic: 56.3381, Pr(>Chisq(3)): 0
```

From the regression results, we could see that all the coefficients, including the intercept, is statistically significant. Specifically, the coefficients for extreme weather counts are positive. **This suggests that a high level of foundation is more likely when SPI index is higher than 1 or below than -1.** One more drought or flood, on average, will increase 0.105 new political party foundation.

In addition, further testing the effect of extreme weather on different kinds of political parties, I performed a regression of the number of specific types of political parties founded on the count of extreme weather years, for each kind of political party. The regression results are shown below (Code for regressions are also in Appendix IV):

TABLE 3.2 Panel Regression with Auto Correlation Paris-Winsten Correction. The outcome variable is the number of political parties founded, for different political parties, and the predictor variables are counts of drought and flood, including control for the number of years in each period and a linear control for each election year.

Results

	Dependent variable:									
	caste	socialist	communist	secular	economic	nationalist	liberal	religious	ethnic	farleft
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
extreme_count	-0.003	0.041***	-0.016*	0.009	0.000	0.007	-0.006	0.005	-0.008	0.031*
	(0.010)	(0.013)	(0.009)	(0.007)	(0.000)	(0.007)	(0.020)	(0.004)	(0.007)	(0.016)
year_diff	-0.087***	-0.059***	-0.011	0.003	0.000	-0.063***	-0.258***	-0.001	0.023***	-0.061***
	(0.008)	(0.009)	(0.007)	(0.005)	(0.000)	(0.006)	(0.016)	(0.003)	(0.005)	(0.012)
year	0.002**	-0.005***	0.00004	0.006***	0.000	-0.015***	-0.022***	-0.001	0.009***	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Constant	-3.546**	11.237***	0.185	-11.150***	0.000	29.763***	45.318***	1.541	-17.697***	8.694***
	(1.689)	(2.221)	(1.461)	(1.281)	(0.000)	(1.190)	(3.224)	(1.324)	(1.006)	(2.785)

*Note: *p<0.1, **p<0.05, ***p<0.01*

From the regression results, we could see that socialist and farleft political parties have significant and positive coefficients; whereas communist political parties have a significant and negative coefficient. The rest types of political parties do not have significant results. This result suggests that socialist, communist, and farleft political parties seem to be more likely than other kinds to be formed when a district experiences extreme weather.

(Question 3) More Political Parties Will be Formed when Droughts or Floods Occur in Neighboring Districts during the Previous Period

My second hypothesis yields that the formation of political parties in a focal district is not only influenced by the number of moderate droughts or moderate floods it experienced, but also influenced by its neighboring districts' experiences. To test this hypothesis, I performed a regression the number of new parties on number of extreme weather counts in the focal district as well as the average number of extreme weather counts in its neighboring districts in the previous election period. Similar to the previous question, I also included a control for district fixed effects, a control for panel-specific autocorrelation, a linear control for each election year, and a control for the length of election period in the regression.

The results of the regression are shown below: (Code for performing the regression in Appendix V)

TABLE 4. Panel Regression with Auto Correlation Paris-Winsten Correction. The outcome variable is the number of political parties founded and the predictor variables are counts of drought and flood and the average counts of drought and flood in the neighboring districts during the previous election period, including control for the number of years in each period and a linear control for each election year.

```
##
## Panel Regression with AR(1) Prais-Winsten correction and panel heteroskedasticity-robust standard errors
##
## Unbalanced Panel Design:
##  Total obs.:       2012 Avg obs. per panel 10.4792
##  Number of panels: 192  Max obs. per panel 12
##  Number of times:  13   Min obs. per panel 1
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      -26.006814   6.264206  -4.152 3.44e-05
## extreme_count                      0.059751   0.033547   1.781  0.07504
## year_diff                         -0.081030   0.029885  -2.711  0.00676
## year                               0.014207   0.003117   4.558 5.49e-06
## mean_neighbor_extreme_weather_lag1   0.141507   0.037709   3.753  0.00018
##
## (Intercept)                      ***
## extreme_count                    .
## year_diff                        **
## year                             ***
## mean_neighbor_extreme_weather_lag1 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-squared: 0.2271
## Wald statistic: 99.0975, Pr(>Chisq(4)): 0
```

From the regression results, we could see that the extreme weather count in the focal district has a positive coefficient with a slightly decreased statistical significance, while the average extreme weather count of its neighboring districts during the previous election period becomes very significant. The positive and significant coefficient on those two predictive variables suggests that even taking into account a district's own droughts and floods, the level of entry of new political parties in a district **will also depend on the number of years its neighboring districts experiencing droughts or floods during the previous election period.**

(Question 4A) Formation of National Scope Parties are More Likely to be Influenced by Droughts or Floods

Experiencing droughts or floods might relate differently to the entry and diffusion of political parties depending on their scope. To test this hypothesis, I performed three separate regression, one each predicting the entry of new national, state, and regional scope political parties by a focal district's own experience of droughts or floods combined with its neighboring districts' average. Similar controls are included as in previous questions.

The results of the regression are shown below: (Code for performing the regression in Appendix VI)

TABLE 5. Three Panel Regressions with Auto Correlation Paris-Winsten Correction for political parties of different scope. The outcome variable is the number of political parties founded for each scope and the predictor variables are counts of drought and flood and average counts of drought and flood in the neighboring districts during the previous election period, including control for the number of years in each period and a linear control for each election year.

Results

	Dependent variable:		
	national	state	regional
	(1)	(2)	(3)
extreme_count	0.052**	-0.002	-0.010
	(0.025)	(0.006)	(0.018)
year_diff	-0.242***	0.022***	0.117***
	(0.021)	(0.007)	(0.014)
year	-0.025***	0.010***	0.033***
	(0.002)	(0.001)	(0.002)
mean_neighbor_extreme_weather_lag1	0.118***	-0.007	0.079***
	(0.027)	(0.006)	(0.017)
Constant	51.947***	-20.534***	-65.493***
	(4.549)	(1.480)	(3.109)

*Note: *p<0.1, **p<0.05, ***p<0.01*

From the regression results, we could see that the coefficient for extreme counts is positive and significant for national scope political parties only; while the coefficient for neighbors' average counts is positive and significant for both national scope and regional scope political parties. This suggests that **national scope political party is most likely to be influenced by number of droughts and floods that a district experienced; state scope political party is not very related to moderate extreme weathers; whereas regional scope political party is not related to its own experience of extreme weathers, but more influenced by the neighboring districts' experience of droughts and floods**. This result aligns well with reality. National scope political parties in general focus equally on both the region of its foundation and neighboring districts; regional scope political parties in general focus on its neighboring districts more; and state scope political parties focus on problems other than economics and weather.

(Question 4B) Experiencing Droughts or Floods Decreases Political Concentration

One would also expect a relation between political concentration and experience of droughts and floods since moderate extreme weather influences the level of entry of political parties. The political concentration is normally measured by the Herfindahl index, which measures the degree to which a few parties command the majority of vote share. The equation could be represented as below:

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

To test whether experience of droughts or floods is related to political concentration in a specific election period, a regression predicting the Herfindahl index is performed as a function of the number of years of droughts or floods that occur in a district in the current election period, as well as the average number of years of droughts or floods that occur in its neighboring districts during the previous election period. Similar control is contained in the regression as previous problems.

The results of the regression are shown below: (Code for performing the regression in Appendix VII)

TABLE 6. Panel Regression with Auto Correlation Paris-Winsten Correction. The outcome variable is the Herfindahl Index and the predictor variables are counts of drought and flood and average counts of drought and flood in the neighboring districts during the previous election period, including control for the number of years in each period and a linear control for each election year.

```
##
## Panel Regression with AR(1) Prais-Winsten correction and panel heteroskedasticity-robust standard errors
##
## Unbalanced Panel Design:
##  Total obs.:       2004 Avg obs. per panel 10.4375
##  Number of panels: 192  Max obs. per panel 12
##  Number of times:  13   Min obs. per panel 1
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       88478.322   4250.111  20.818  < 2e-16
## extreme_count                       -44.996     20.930  -2.150   0.0317
## year_diff                            26.864     14.400   1.866   0.0622
## year                                -43.779      2.127 -20.578  < 2e-16
## mean_neighbor_extreme_weather_lag1  -99.532     19.900  -5.002 6.18e-07
##
## (Intercept)                       ***
## extreme_count                     *
## year_diff                         .
## year                              ***
## mean_neighbor_extreme_weather_lag1 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-squared: 0.6719
## Wald statistic: 685.9796, Pr(>Chisq(4)): 0
```

From the regression results, we could see that the coefficients for both extreme counts and its neighboring districts' average are negative and significant. **This suggests that an increase in the number of years that a district experienced droughts or floods, as well as an increase in its neighboring districts' average will decrease the political concentration in the focal district**. More specifically, since moderate extreme weather leads to the formation of new political parties, the vote will be more diversified in those districts that have experienced a high level of droughts or floods.

(Question 5) Diffusion Effect of New Political Parties Formation

Regions influence each other as political activity diffuses across regions over time. Our hypothesis yields that if a political party has not yet formed in a focal district but existed in its neighboring districts, that political party will be more likely to be founded when the focal district experiences droughts or floods. To analyze this, a new party information data set is included. Through combining this data set with the original data set, two data fields are calculated (Code for data cleaning in Appendix VIII). The existed column counts the number of new political parties being founded in a district, that have contested an election in a neighboring district in any previous election period; the non-existed column, in opposite, counts the number of new political parties being founded in a district, that has not contested an election in its neighboring districts. After combining the data sets, a similar regression is performed on the existed political parties and non-existed political parties as a function of the number of droughts or floods occur in a district in the current election period and the number of droughts or floods that occur in its neighboring district in the previous election period. Similar control is contained in the regression as previous problems.

The results of the regression are shown below: (Code for performing the regression in Appendix VIII)

TABLE 5. Tow Panel Regressions with Auto Correlation Paris-Winsten Correction for political parties of different scope. The outcome variable is number of political parties founded that have or have not existed in its neighboring districts and the predictor variables are counts of drought and flood and average counts of drought and flood in the neighboring districts during the previous election period, including control for number of years in each period and a linear control for each election year.

Results

	Dependent variable:	
	existed	non-existed
	(1)	(2)
extreme_count	-0.016	0.084***
	(0.018)	(0.032)
year_diff	-0.045***	-0.070**
	(0.013)	(0.028)
year	-0.013***	0.024***
	(0.002)	(0.003)
mean_neighbor_extreme_weather_lag1	-0.038**	0.107***
	(0.019)	(0.032)
Constant	26.110***	-44.833***
	(3.338)	(5.418)

*Note: *p<0.1, **p<0.05, ***p<0.01*

From the regression results, we could see that the coefficient of extreme weather count and average neighboring extreme weather count in the previous period is only positive and significant for non-existed political parties. **The results suggest that only the formation of those parties that have not existed in its neighboring parties is related to moderate extreme weather**. Furthermore, the intercept for existed parties is positive and significant, whereas the intercept for non-existed parties is negative and significant. This reveals important information of the diffusion process of political parties. **If a political party has existed in a district's neighboring districts, it will naturally diffuse over time, no matter whether extreme weather exists or not. However, if a political party is newly formed and has never existed in a district's neighboring districts before, it is highly likely to be positively influenced by moderate extreme weather.**

CONCLUSION

In conclusion, the moderate extreme level of rainfall will influence the formation of political parties in India. In particular, for a focal district, the level of entry of new parties will increase as its own experience of moderate droughts or floods increases and as its neighboring districts' average experience of moderate droughts or floods increases. In addition, national scope political parties are more likely to be influenced by extreme weather; whereas regional scope political parties will only be influenced by the focal district's neighboring district's experience of droughts or floods. As a result of new entries of new political parties, the political concentration, the Herfindahl Index, of the focal district would decrease. Moreover, the formation of political parties also diffuses across borders naturally and is not likely influenced by the number of years that a district experiences droughts or floods. However, counts of those extreme weather are crucial in the formation of newly started political parties. More experience of extreme weather increases the likelihood of the formation of political parties that have not existed in the focal district's neighboring districts.

REFERENCES

"THE CONSTITUTION OF INDIA." *National Portal of India*, www.india.gov.in/.

Appendix

Appendix I: Code for Question 1A

```
# creating window
eyear <- unique(districtdt$year)[order( unique(districtdt$year))]
syear <- nafill(shift(eyear,type='lag',n=1),type = "const",fill=1945)
districtdt[,window := as.factor(unlist(lapply(year, function(x) which(x == eyear)))),]
# calculate for each district, each election period, total rain and average spi
raindt <- raindt[year<=1999][, window := as.factor(unlist(lapply(year,function(x) which(x > syear & x <= eyear))))]
rain_amount <- raindt[,.(total_rain = sum(rain,na.rm=TRUE),avg_spi = mean(spi,na.rm=TRUE)),by=.(district,window)]
dt <- as.data.table(left_join(districtdt,rain_amount,by=c("district","window")))
par(mfrow=c(1,2),oma = c(0, 0, 3, 0))
plot(dt$total_rain,dt$new_parties,ylim=c(0,12),main =  'Total Rainfall',xlab = 'Total Rainfall (in millimeters)',ylab = 'Num
ber of Parties Formed')
plot(dt$avg_spi,dt$new_parties,ylim=c(0,12),main =  'Average SPI',xlab = 'Average SPI',ylab = 'Number of Parties Formed')
mtext("Number of New Parties Formed v.s. Level of Rainfall\n during a Election Period in India", outer = TRUE, cex = 1.5)
```

Appendix II: Code for Question 1B

```
# get lag by district
dt[,total_rain_lag1 := shift(total_rain,type='lag',n=1),by=.(state,district)]
dt[, avg_spi_lag1 := shift(avg_spi,type='lag',n=1),by=.(state,district)]

# get neighbors means
border_with_value <- as.data.table(left_join(borderdt,dt[,list(district = district,total_rain_lag1 = total_rain_lag1,avg_spi
_lag1 = avg_spi_lag1,window=window),],by='district'))

neighbor_mean <- border_with_value[!is.na(window)][,.(mean_neighbor_rain_lag1 = mean(total_rain_lag1,na.rm=TRUE),mean_neighb
or_spi_lag1 = mean(avg_spi_lag1,na.rm = TRUE)),by=.(focal_district,window)]

# merge table
dt <- as.data.table(left_join(dt,neighbor_mean,by=c("district" = "focal_district","window"="window")))

# get year difference
syear2 <- syear
syear2[syear2 == 1945] <- 1946
yeardif <- eyear - syear2
dt[,year_diff := as.numeric(unlist(lapply(window,function(x) yeardif[x]))),]

# linear models
rainfall_vs_lag1rainfall <- plm(total_rain ~ total_rain_lag1+year_diff,data=dt,effect = 'twoways',model='within',index='dist
rict')
rainfall_vs_lag1rainfall_neighbor <- plm(total_rain ~ mean_neighbor_rain_lag1+year_diff,data=dt,effect = 'twoways',model='wi
thin',index='district')
spi_vs_lag1spi <- plm(avg_spi ~ avg_spi_lag1+year_diff,data=dt,effect = 'twoways',model='within',index='district')
spi_vs_lag1spi_neighbor <- plm(avg_spi ~ mean_neighbor_spi_lag1+year_diff,data=dt,effect = 'twoways',model='within',index='d
istrict')
```

Appendix III: Code for Question 1C

```
# get counts
extreme_climate <- raindt[!is.na(spi),.(extreme_count = sum((spi < -1 | spi > 1),na.rm=TRUE)),by = .(district,window)]
dt <- as.data.table(left_join(dt,extreme_climate,by=c("district" = "district","window"="window")))


# get neighbor average
dt[,extreme_count_lag1 := shift(extreme_count,type='lag',n=1),by=.(state,district)]
border_with_count <- as.data.table(left_join(borderdt,dt[,list(district = district,extreme_count_lag1 = extreme_count_lag1,window=window),],by='district'))
neighbor_mean_count <- border_with_count[!is.na(window)][,.(mean_neighbor_extreme_weather_lag1 = mean(extreme_count_lag1,na.rm=TRUE)),by=.(focal_district,window)]

# merge table
dt <- as.data.table(left_join(dt,neighbor_mean_count,by=c("district" = "focal_district","window"="window")))

# linear model
extreme_vs_lag1 <- pglm(extreme_count~ extreme_count_lag1+year_diff,data=dt,effect = 'twoways',model='within',index='district',family='poisson')
```

```
extreme_vs_lag1_neighbor <- pglm(extreme_count~ mean_neighbor_extreme_weather_lag1+year_diff,data=dt,effect = 'twoways',model='within',index='district',family='poisson')
```

Appendix IV: Code for Question 2

```
# linear model

dt2 <- dt[!is.na(extreme_count) &!is.nan(extreme_count)]
df2 <- as.data.frame(dt2)


new_party_vs_count <- panelAR(new_parties ~ extreme_count + year_diff + year, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.rm = TRUE)
```

```
caste <- panelAR(new_parties_caste ~ extreme_count + year_diff + year, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.rm = TRUE)
socialist <- panelAR(new_parties_socialist ~ extreme_count + year_diff + year, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.rm = TRUE)
communist <- panelAR(new_parties_communist ~ extreme_count + year_diff + year, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.rm = TRUE)
secular <- panelAR(new_parties_secular ~ extreme_count + year_diff + year, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.rm = TRUE)
nationalist <- panelAR(new_parties_nationalist ~ extreme_count + year_diff + year, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.rm = TRUE)
economic <- panelAR(new_parties_economic ~ extreme_count + year_diff + year, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.rm = TRUE)
liberal <- panelAR(new_parties_liberal ~ extreme_count + year_diff + year, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.rm = TRUE)
religious <- panelAR(new_parties_religious ~ extreme_count + year_diff + year, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.rm = TRUE)
ethnic <- panelAR(new_parties_ethnic ~ extreme_count + year_diff + year, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.rm = TRUE)
farleft <- panelAR(new_parties_farleft ~ extreme_count + year_diff + year, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.rm = TRUE)
```

Appendix V: Code for Question 3

```r
dt2 <- dt[!is.na(mean_neighbor_extreme_weather_lag1) & !is.nan(mean_neighbor_extreme_weather_lag1)]
df2 <- as.data.frame(dt2)

new_party_vs_count_and_neighbor <- panelAR(new_parties ~ extreme_count + year_diff + year  + mean_neighbor_extrem
e_weather_lag1, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.n
a.rm = TRUE)
```

Appendix VI: Code for Question 4A

```r
national_scope <-  panelAR(new_parties_national_scope ~ extreme_count + year_diff + year  + mean_neighbor_extreme
_weather_lag1, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.
rm = TRUE)
state_scope <- panelAR(new_parties_state_scope ~ extreme_count + year_diff + year  + mean_neighbor_extreme_weathe
r_lag1, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.rm = TR
UE)
regional_scope <- panelAR(new_parties_regional_scope ~ extreme_count + year_diff + year  + mean_neighbor_extreme_
weather_lag1, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.r
m = TRUE)
```

Appendix VII: Code for Question 4B

```r
political <- panelAR(political_concentration ~ extreme_count + year_diff + year  + mean_neighbor_extreme_weather_
lag1, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.rm = TRUE
)
summary(political)
```

Appendix VIII: Code for Question 5

```r
## find likelihood of existed new party in neighbor
party <- newParty[,list(district = district,year=year,party_name = party_name)]
party[,window := as.factor(unlist(lapply(year, function(x) which(x == eyear)))),]
border_with_new_party <- as.data.table(left_join(borderdt,party,by='district'))
border_with_new_party_plus_focal <- as.data.table(left_join(border_with_new_party,party[,list(district=district,f
ocal_window = as.numeric(as.character(window)),focal_party = party_name)],by=c('focal_district'='district')))[,wi
ndow:=as.numeric(as.character(window))][!is.na(window) & !is.na(focal_window)]
border_with_new_party_plus_focal <- na.omit(border_with_new_party_plus_focal)

focal_likelihood <- border_with_new_party_plus_focal[,existence:= focal_window > window & focal_party == party_na
me]
focal_likelihood <- unique(focal_likelihood[,list(focal_district=focal_district,focal_window=focal_window,existen
ce = existence,party = focal_party)])[!(party == '' | party == ' ')][,.(existed_likelihood = sum(existence,na.rm
 = TRUE)),by = .(focal_district,focal_window)]

focal_likelihood[,focal_district:=as.character(focal_district)]
focal_likelihood[,focal_window:=as.character(focal_window)]


## join back to original dataset
dt <- as.data.table(left_join(dt,focal_likelihood,by=c("district"="focal_district","window"="focal_window")))
```

```
dt <- dt[!duplicated(dt[,list(d=district,w=window)])]
dt[,non_existed_likelihood := (new_parties - existed_likelihood)]
dt[,existed_likelihood := (existed_likelihood)]
dt <- dt[!is.na(existed_likelihood) & !is.na(non_existed_likelihood)&!is.na(mean_neighbor_extreme_weather_lag1)&!
is.na(extreme_count) &!is.nan(extreme_count)]
df2 <- as.data.frame(dt)
```

Build linear models:

```
existed <- panelAR(existed_likelihood ~ extreme_count + year_diff + year  + mean_neighbor_extreme_weather_lag1, d
ata = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.rm = TRUE)
non_existed <- panelAR(non_existed_likelihood ~ extreme_count + year_diff + year  + mean_neighbor_extreme_weather
_lag1, data = df2, panelVar='district',timeVar='year',autoCorr = 'psar1',panelCorrMethod = 'phet',rho.na.rm = TRU
E)
```

Comparing between models:

```
stargazer(coeftest(existed),coeftest(non_existed),column.labels=c('existed','non-existed'),title = 'Results',alig
n = TRUE,header=FALSE, type='html')
```