THE EFFECT OF CIRCUIT COURT DECISIONS AND JUDGE CHARACTERISTICS ON SOCIAL AND POLITICAL ATTITUDE OF THE PUBLIC: CAUSAL ANALYSIS

BHASKAR GURRAM, NAMAN KUMAR, PRANEETH YENUGUTALA, ROHIT MUTHYALA (ADVISORS: DANIEL CHEN AND ELLIOT ASH)

PROBLEM DEFINITION

We present an application of machine learning and causal inference to study the effect on social and political attitude of the public, based on court rulings and characteristics of judges involved. Our goal is to determine if the sentiment of the judges in the court rulings could be used to infer the public opinion. We also find the characteristics of the judge(s) that are important for determining this causal effect by using them as instruments, which helps to ensure that the determined causal effect is indeed consistent and unbiased.

We use American National Election Survey data, Court rulings data, and judge characteristics data for our analysis. The result can be used to inspect how public attitude can be changed by changing judge characteristics.

INTRODUCTION

There are 13 appellate courts that sit below the U.S. Supreme Court, and they are called the U.S. Courts of Appeals. The 94 federal judicial districts are organized into 12 regional circuits, each of which has a court of appeals. The appellate court's task is to determine whether the law was applied correctly in the trial court. Appeals courts consist of three randomly selected judges and do not use a jury. Each judge has a different background and characteristics.

The rulings of these courts can be a major factor in shaping the social and political attitude of the public. These rulings are verbose written text of several paragraphs. We calculate paragraph and case level similarities using word2vec to certain target values. We also use NLP to calculate sentiment for each case that could be used to understand the context of these paragraphs. These could then be aggregated at a case level by getting the weighted average of the sentiment, with similarities of the cases being the weight. We find the effect of this weighted sentiment towards the attitudes of the public.

DATASETS

COURT RULINGS/CASE OPINIONS

The dataset is the case level opinion of every case collected yearly from 1880 to 2013. Each case file contains a list of paragraphs of the case proceedings.

BLOOMBERGCASELEVEL_TOUSES

This database consists of case level details of each case and contains 387,898 case records from 1880 to 2013. Each case record has 134 features which corresponds to general information for the case.

BLOOMBERGVOTELEVEL_TOUSES

This dataset consists of case level details including judge's biographical characteristics. It contains 1,163,694 vote records and has 414 features. For each case, there are 3 records; one for each judge, with the common case characteristics and individual judge characteristic as features.

ANES

This dataset consists of the results of surveys conducted between 1948 and 2008. Each record in these surveys corresponds to the answers of each individual participant. The questions that has been asked in three or more Time Series studies (once every two years) are included in this data. There are 951 distinct questions that were asked across the years, 40 of these are chosen by us, known as feeling thermometer variables.

FEATURE ENGINEERING / VARIABLE COMPUTATIONS

ENDOGENOUS VARIABLE

The data was from a case i with panel of judges J, where judges were indexed by j, in circuit c, at year t. There are a set of thermometer targets, indexed by k. Each paragraph d in case i has a similarity to each target, W^k_{id}, and a sentiment, S_{id}. Let W^k_i denote the vector of paragraph similarities in case i. Let S_i denote the vector of sentiments in case i. The case-level sentiment towards target k is defined as: $S^k_i = W^k_i S_i$, which is the dot product. The circuit-year level endogenous regressor is therefore S_{ckt}, the average case-level sentiment toward k for each case in circuit-year ct.

INSTRUMENT VARIABLE

Each case has a similarity to each target, S^k_i. Therefore, we have to weigh the judge assignment instruments in the same way. Let $\overline{W}^k_i$ be the average similarity to k of case i. When aggregating the judge-bio instruments Z, weight the average by $\overline{W}^k_i$, the similarity to k of each case i to which they are assigned.

EXOGENOUS VARIABLE

The outcome variable is defined by circuit c and year t and target k. It is the average thermometer score, Y_{ckt}, for all respondents in the ANES in that circuit-year.

FIRST STAGE REGRESSION

We want to use OLS to estimate

$S_{ckt} = \Upsilon_{ck} + \Upsilon_{kt} + \Upsilon_{ct} + Z'_{ckt}\Upsilon_Z + \eta_{ckt}$

S_{ckt}, the weighted average sentiment toward target k for cases in circuit c during year t

Υ_{ck}, a set of dummy variables (fixed effects) for each circuit-target

Υ_{kt}, a set of dummy variables (fixed effects) for each year-target

Υ_{ct}, a set of dummy variables (fixed effects) for each circuit-year

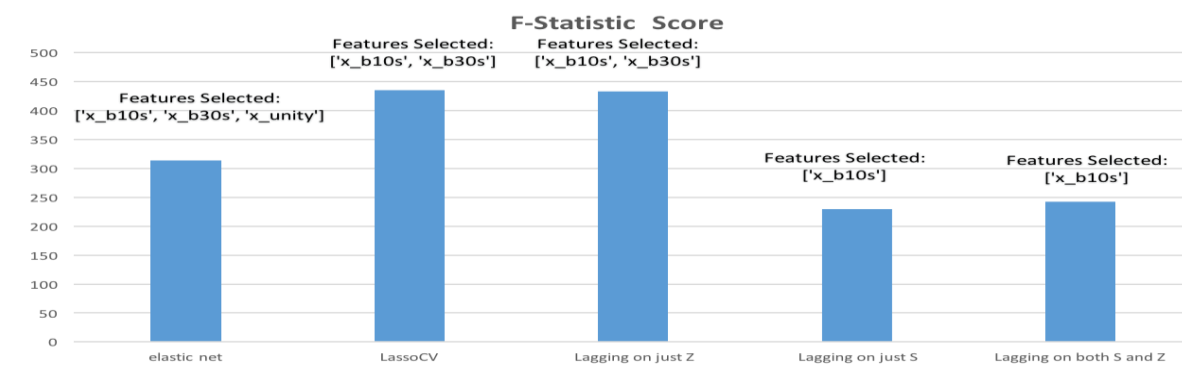
Z_{ckt}, weighted average judge biographical characteristics in circuit c and year t, weighted by the similarity to target k of the cases to which the judges are assigned.

η_{ckt}, error term.

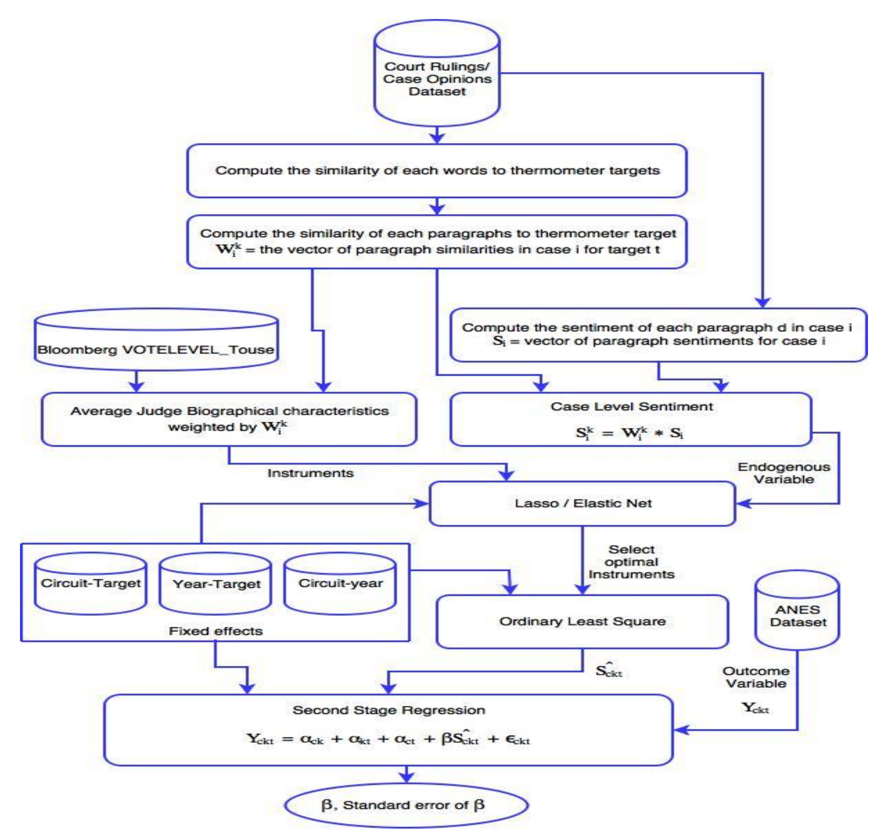
IMPORTANT WORDS AND FIRST STAGE RESULTS

Thermometer Variable	TOP-1	TOP-2	TOP-3	Thermometer Variable	Top-1	Top-2	Top-3
Democrats	Committeemen	Electors	Votes	Political Independents	Political	Partisan	Ideological
Republicans	Voters	Candidates	Canvassers	Political Parties	Parties	Litigants	Tribes
Protestants	Intervenors	Respondents	Petitioners	Poor People	Poor	Gentle	Habits
Catholics	Muslims	Indonesians	Coptic	Republican Party	Party	Nonparty	Creditor
Jews	Christians	Russians	Germans	Women Right Activist	Right	Desire	Opportunities
Blacks	Negroes	Minorities	Nonwhites	Young People	People	Teenagers	Young
Whites	Caucasians	Anglos	Promotions	Asian Americans	Americans	Asian	African
Southerners	Villains	Traitors	Ninja	Congress	Legislature	Legislation	Drafters
Big Business	Business	Big	Thriving	Environmentalists	Longliners	Fishers	fishermen
Labor Unions	Labor	Unions	Union	Anti-Abortionists	Anti	Combats	Prevention
Liberals	Gruffalo	Muggers	Baruch	Federal Government	Government	Federal	State
Conservatives	Greeks	Italians	Partisans	Illegal Aliens	Aliens	Illegal	Immigrants
Military	Civilian	Naval	Uniformed	Christian Fundamentalists	Christian	Pentecostal	Buddhist
Policemen	Patrolmen	Youths	Officers	Radical Students	Students	Pupils	Schools
Black Militants	Black	White	Hispanic	Farmers	Growers	Dairies	Merchants
Civil Rights Leaders	Rights	Civil	Interests	Feminists	Entrepreneurs	Ambassadors	Competitors
Chicanos Hispanics	Hispanics	Latinos	Chicanos	Evangelical Groups	Groups	Churches	Congregations
Democratic Party	Litigant	Candidate	party's	Elderly	Infant	Homeless	Autistic
Middle Class People	Residents	Class	counties	Supreme Court	Supreme	Court	Sjc
People on Welfare	on	upon	lives	Women	Females	Males	youngsters

F-STAT AFTER FIRST STAGE OLS



DESIGN AND METHODOLOGY



2SLS

EQUATION

$Y_{ckt}=\alpha_{ck}+\alpha_{kt}+\alpha_{ct}+\beta\hat{S}_{ckt}+\epsilon_{ckt}$

Y_{ckt} – Outcome variable from ANES

$\hat{S}_{ckt}$ – Received after first stage of IV2SLS

LASSOCV (NO LAG)

coef	std err	t	P>|t|	[0.025	0.975]
-143.0589	44.051	-3.248	0.001	-229.420	-56.698

LAG ON Z

coef	std err	t	P>|t|	[0.025	0.975]
-257.4727	45.193	-5.697	0.000	-346.075	-168.871

LAG ON S

coef	std err	t	P>|t|	[0.025	0.975]
-130.0840	42.898	-3.032	0.002	-214.185	-45.983

LAG ON S AND Z

coef	std err	t	P>|t|	[0.025	0.975]
-257.4710	45.194	-5.697	0.000	-346.074	-168.868

PERFORMANCE EVALUATION AND EXPERIMENTS

FEATURE SELECTION

Lasso, elastic-net and random forest is used.
Features selected via Lasso give the best F-stat value.

F-TEST

The test is performed after first stage regression.
A high value of F-stat (higher than 12) indicates causality.

USING LAG IN VARIOUS SETTINGS (ADVANCED FEATURE SELECTION)

Lag values for Endogenous and Instrument variables is used.
Results are computed using multiple such values for best setting.

ENDOGENEITY BIAS HANDLING (FIXED EFFECT MODELLING)

Dummy variable is plugged in to account for fixed affect.
Missing variable bias/ endogeneity bias is hence accounted and removed.

CONFIDENCE INTERVAL

Confidence interval for second stage regression is determined.
High absolute value in the confidence interval signals more confidence on conclusion.

P-VALUE

Small p value signifies a good prediction.
A p value below 0.05 is considered good, which is achieved.

HIGH VALUE OF COEFFICIENT

The coefficient of second stage regression is generally used to evaluate result.
A high absolute value signifies a good result.
A high absolute value is achieved in our experiments.

CONCLUSION AND FUTURE WORK

CONCLUSION

We observed that, if the judge sentiment is more positive towards a target, it has a reverse effect on the attitudes of the public towards that target.

The coefficients learned on the judge sentiment in the second stage of 2sls, using IV2SLS lies between -229.43 and -56.70 with a confidence interval of 95%. The high F-Statistic for our first stage regression, combined with the high negative coefficient value on the judge sentiment gives us enough evidence to conclude this.

FUTURE WORK

There are many different algorithms and topics to explore in feature selection and causality. It would be interesting to analyze the estimation of heterogeneous effects using random forests developed by Athey-Wager.

We can implement lagged effects using the weighted features from many previous years instead of just one year, to factor in its effects and see the change in our result and conclusion.

Additionally, we could perform robust casual inference if we can get the public opinion data towards the thermometer variables for every year from 1880 to 2012. The constraint on availability of public opinion data restricts our research.

CHALLENGES

Memory and number of file limitation on NYU HPC Prince cluster was of hindrance.

Common features being selected through various feature selection settings limited the results and hence conclusion.

Due to large data size, it couldn't be loaded in memory and hence had to be split in chunks.

Word2vec had to be enhanced to Phrase2vec to account for phrases as thermometer variables.

ACKNOWLEDGEMENT

We would like to thank Dr. Daniel L. Chen and Dr. Elliot Ash for providing the data source and guidance on the project.

We would like to extend our gratitude to Prof. David Rosenberg and Brett Bernstein for their explanations of Machine Learning principles and concepts in the course.

REFERENCES

[1] http://users.nber.org/~dlchen/papers/Policies_Affect_Preferences.pdf
[2] http://web.mit.edu/14.771/www/emp_handout.pdf
[3] http://www.jblumenstock.com/files/courses/econ174/FEModels.pdf