Causal Effects of Judicial Sentiment:
Methods and Application to U.S. Circuit Courts

Sergio Galletta¹, Elliott Ash¹, Daniel L. Chen²

¹ETH Zürich
²Toulouse Institute for Advanced Study

Abstract

This paper provides a general method for analyzing the causal effects of sentiments expressed in the language of judicial rulings, with an application to the effect on social attitudes. We apply natural language processing tools to the text of U.S. appellate court opinions to extrapolate judges’ sentiments toward a number of specific target groups. Exogenous variation in those sentiments comes from an instrumental variables approach, which exploits the random assignment of judges to cases (and the fact that judge characteristics provide good cross-validated predictors of expressed sentiments). Our estimates are consistent with a backlash effect from judge sentiments to social attitudes. This effect does not persist over time and is heterogeneous depending on the target group considered.
1. Introduction

An increasing number of studies show that individuals’ preferences (or attitudes) respond to the surrounding cultural and institutional environment (see, for example, Alesina and Fuchs-Schündeln, 2007; Campa and Serafinelli, 2019; Eugster et al., 2011; Galletta, 2018; Kim et al., 2016). An important institutional channel for the inculcation of norms is the judiciary. But the previous literature on the determinants of social attitudes has mostly focused on other factors.

While judges can set norms in many ways, one important way is in the sentiments expressed in their rulings. These sentiments could have many impacts upon society, including the social attitudes of the population of the local jurisdiction. But getting at these types effects is difficult empirically because one has to measure sentiments in judicial opinions and obtain exogenous variation in those sentiments.

This paper provides a method for analyzing the impacts of judicial sentiments. This draws on two key ingredients of the judicial empirical context – on our case, U.S. Circuit Courts, 1964-2008. First, we have the full texts of published opinions in these courts during this period, from which we compute text-based sentiment scores toward different social groups. Second, there is random assignment of judges to cases, which can be used to obtain exogenous variation in the sentiments expressed in opinions. Specifically, we have a rich set of biographical characteristics on the judges assigned to these cases, from which we construct a high-dimensional matrix of instruments for use in the regression analysis.

To illustrate the usefulness of the method, we take on the following research question: When judicial rulings express support or opposition for various social groups, how does that influence attitudes toward these groups? Do preferences shift towards what the law indicates? Or is there a backlash and shift in the opposite direction?

The theory on this question is mixed. Laws can be expected to backfire when they are in conflict with social norms or when they pass signals that affect the stigma or honor of law-breakers (Acemoglu and Jackson, 2017; Benabou and Tirole, 2011). With respect to court decisions, two competing theoretical models have been widely accepted. On the one hand, thermostatic models predict preferences to backlash against court decisions (Chen et al., 2016; Ura, 2013). On the other hand, legitimation models suggest

1 Alesina and Giuliano (2015) provide an extensive survey of recent findings.
a mechanism by which preferences may shift toward the position of the court (Caldeira and Gibson, 1992).

In light of the competing theoretical predictions, empirical evidence is needed. But at present there is limited causal evidence on the relationship between judicial rulings and social attitudes.\(^2\) Without this evidence, judges and policymakers might be misgauging the societal impacts of judicial rulings.

For this empirical application we pair the data on judicial opinion texts and judicial characteristics with state-level biennial data on U.S. citizen opinions from the American National Election Survey (ANES). We focus on the "feeling thermometer" questions, which ask respondents to provide a rating, from zero to one hundred, of their warmth/attitude toward twenty different groups (African Americans, big business, Catholics, labor unions, the military, etc.). These survey measures are the social attitudes which may be affected by expressed judicial sentiments.

A first contribution of this paper is the method used to infer judges’ preferences towards specific target groups. Rather than focusing on the direction of decisions (for/against a particular group), we apply natural language processing techniques to the text of U.S. Circuit Court opinions. In particular, we draw upon recent embedding methods, which vectorize words and documents in a relatively low-dimensional space, where locations and directions encode meanings and associations. At a sentence level, our algorithm measures both the relevance to each of the twenty groups, and the level of sentiment (positive/warm or negative/cold). From these sentence-level measures we compute the relative sentiment in a case by the correlation between group associations and sentiment associations. This flexible and informative solution to measuring judicial attitudes highlights the growing literature using text to understand biases and preferences (Caliskan et al., 2017).

The paper’s second contribution is to address the empirical challenge that judge sentiments do not vary randomly over time and space and therefore this variable is likely to be considered endogenously determined in many contexts. For instance, in our application on social attitudes, we would expect that areas where social groups are less popular would also be treated differently by judges in the opinions, due to unobserved

\(^2\)This link is potentially bidirectional, meaning that judge rulings can be expected to either affect or be affected by citizens’ opinion (see, for example, Casillas et al., 2011; Giles et al., 2008). In this analysis we focus on the former direction of the relationship.
time-varying confounders. OLS estimates comparing sentiments to attitudes would be biased. To address this issue, we apply an instrumental variables approach that exploits the random assignment of judges to cases, along with the fact (which we document) that judges vary systematically in their writing sentiment tendencies toward various groups.

Unlike the literature that instruments for judicial decisions using judge leniency (e.g. Dobbie and Song, 2015), there is no straight-forward way to instrument for sentiment expressed in text. Instead, we apply machine learning tools to extract predictive power in the first stage from a high-dimensional set of instruments describing the biographical characteristics of judges assigned to these cases. Our approach extends the literature on sparse optimal instruments using cross-fitting techniques (Belloni et al., 2012; Chernozhukov et al., 2017). Specifically, we apply elastic net regression to the standardized judge characteristics and construct cross-validated instruments using out-of-fold data. The predictions from these estimates are then gathered together to use as instruments in the second stage.

The second stage provides the main results of our application, as we regress U.S. citizen attitudes on instrumented judge writing sentiment towards each social group. These estimates show a negative and statistically significant effect of judges’ sentiments on individuals’ attitudes. In other words, our results tend to confirm the predictions of the thermostatic model of public responsiveness, as citizens’ opinions shift in the opposite direction to those expressed by judges.

We show, however, that this effect does not persist in the long run. It becomes insignificant after 4 years. Moreover, we highlight the presence of heterogeneous effects depending on the target considered (though with less statistical power), where the largest attitudinal backlashes are seen against sentiments toward women and the elderly.

This research is most closely related to Chen et al. (2016) and Ura (2013). Chen et al. (2016) use a similar empirical strategy to estimate the impact of abortion jurisprudence on several outcomes, including abortion attitudes. Their results are in line with our findings, as they show a negative relationship between abortion jurisprudence and abortion preferences. Ura (2013) studies the effect of court decisions on public opinion, focusing on the time variation of Supreme court decisions. He shows in the time series a backlash in the short-term, while Supreme Court decision-making and public mood are positively associated in the long-run. Overall, his finding is more coherent with previous studies showing that public opinion follows court rulings (Clawson et al., 2001; Hoekstra and Segal, 1996; Stoutenborough et al., 2006) rather than studies favoring the opposite

Our results also relate to the literature in experimental economics that studies how changes in the rules affect individuals’ preferences (Dal Bó et al., 2010; Galbiati and Vertova, 2008). Further, we add to existing studies that exploit random assignment of judges (Di Tella and Schargrodsky, 2013; Galasso and Schankerman, 2014; Kling, 2006; Maestas et al., 2013). Finally, we are contributing to the growing literature that uses embedding models to study legal language (Ash and Chen, 2017).

The remainder of the paper is organized as follows. Section 2 describes the institutional background. Section 3 describes the data-set. Section 4 details the estimation strategy while Section 5 provide our results. Section 6 concludes.

2. Institutional background

The U.S. has a common law system. The main feature of this legal system is that decisions taken by judges become precedents for future cases. The Federal Courts system is organized on three levels: the national level (Supreme Court), intermediate level (Circuit Courts) and local level (District Courts). The Circuit Courts play a crucial role as their judges decide whether the decisions taken by the District Court were erroneous.

There are 12 regional U.S. Circuit Courts. Each of these courts is responsible for 3-9 states (see Figure 1). Importantly, for each case there are assigned three life-tenure judges. On average a Circuit has 17 judges, with a minimum of 8 and maximum of 40. Interestingly, both anecdotal and empirical evidence suggest that judges’ attributes are good predictors of their voting behavior as well as the voting behavior of the other judges in the selected panel (Berdejó and Chen, 2014; Boyd et al., 2010; Chen and Spamann, 2016; Fischman, 2011; Klarman, 2004). And critically, cases are randomly assigned to judges.

Circuit court judges are powerful forces in U.S. politics and culture. A large majority of appeals terminate at this stage, and those decisions are binding precedent within the circuit. Therefore judicial decisions have the force of law, and become official articulations of legal and social norms. Unsurprisingly, then, these decisions and the associated opinions are the target of significant attention by elites in government and media.

Evidence of elite response to court opinions includes Weinrib (2012), who documents the response by ACLU attorneys to major Circuit Court decisions on free speech. The attorneys responded by mobilizing people in the media in favor of stronger free-speech
protections. Clark et al. (2018) finds significant use of Twitter after several court decisions. Bromley (1994) is an early paper documenting how journalists do research on circuit court opinions. Lim et al. (2015) document the frequent coverage of criminal decisions in newspapers.\(^3\)

Hoekstra (2000) provides a lengthy discussion of how judicial decisions reach the public. Focusing on the U.S. Supreme Court, she finds that public opinion is affected where the case originates. In those communities, many residents heard about the decision and the decision affected their view of the court. In the same vein, LaRowe and Hoekstra (2014) find the the public have significant knowledge of recent federal judicial decisions.

Beyond judges having a direct line to the public, there is an important role for elites in the process of legal dissemination. Political and cultural elites might learn about the moral norms articulated in these opinions, and then those norms spread to the masses. For example, government officials often have to set up rules based on federal law to assist agencies in compliance. Some research documents how officials adjust behavior to avoid litigation after circuit court decisions (Frost and Lindquist, 2010; Pollak, 2001). Through all these channels, the views espoused in legal opinions could spread to the public.

---

\(^3\)To further check this in our context, we ran a search for articles about federal circuit courts in a news search engine (newslibrary.com), and found many results. For example, we got 35,284 hits for the year 2012, and 5,607 articles specifically mentioned the written opinion.
3. Data

This section describes the data, which have been assembled from a range of sources. We have outcomes (social attitudes), treatments (expressed judge sentiment), and instruments (judge biographical characteristics).

3.1. Social Attitudes Data

First, we use data from the American National Election Survey (ANES) to measure individuals’ preferences towards a set of different target groups. ANES is a survey conducted every two years since 1948 and provides information about citizens voting behavior, as well as their attitudes. For this paper, we focus on the feeling thermometer questions. In these types of questions, individuals are asked to report their attitudes towards a specific object (target group) by choosing a value from 0 to 100 (see Appendix Figure A.1). The respondent assigns a value around 50 if she lacks opinion or knowledge about the target group. A value between 51 and 100 reveals that the respondent feels warmly or favorably towards the target group. On the contrary, a value between 0 and 49 reveals that the respondent has cold or unfavorable feelings towards the target group. This kind of question has been included in the survey since 1964. For this analysis, we select citizen’s opinions for 19 different target groups, listed in Appendix Table A.1. For instance, these questions serve to identify sentiments towards racial groups (black, white) or institutions (supreme court, federal government).

We use this information and calculate our main outcome variables $Y_{ckt}$, which is defined as the average thermometer score for all respondents in the ANES by circuit $c$, target $k$ and year $t$. Appendix Table A.1 reports summary statistics. On average the highest favourable feeling is toward elderly people, while the lowest is towards illegal immigrants. Appendix Figure A.4 shows variation in the country’s average sentiment over time for each target group, while Appendix Figure A.5 displays the average ANES thermometer measure by circuit.

3.2. Constructing Judicial Sentiment

Our second data source is the complete collection of United States Courts of Appeals opinions from 1964 to 2008. The corpus includes all published cases and comes from

4Questions are not asked in each survey wave and therefore we have an unbalanced panel.
Bloomberg Law. We parse the raw text into Python and use the Python module `nltk` to tokenize sentences.

Next, we map sentences into vectors using Doc2Vec (Mikolov et al., 2013; Le and Mikolov, 2014). This algorithm represents words and sentences in a shared vector space (in our case, 200 dimensions). Words that tend to have similar contexts are located near each other (we used window size of five). Sentences with similar language tend to locate close to each other, and tend to locate close to words contained in the sentence. Dai et al. (2015) illustrate the use of Doc2Vec to analyze similarities and analogical relations between documents.

With the trained Doc2Vec model in hand, we also obtain vectors for each of the ANES targets as the average of a set of words for each target. Blacks, for example, are identified off of black, blacks, african, african, african-american, african-americans, negro, and negroes (see Appendix section A.2 for full lists for each target).\(^5\) We then compute the cosine similarities of the sentence vectors to each of the targets. This procedure serves to estimate the degree of association between each sentence with each specific target group. Let \(W_{i}^{k}\) represent the similarity of sentence \(d\) in case \(i\) to target \(k\).

In Appendix Figure A.7, we provide word clouds that report the words most associated with each target.

Next, for each sentence we compute a metric for positive and negative sentiment in each sentence. We use a dictionary of positive words (e.g., “warm”, “favorable”, “good”) and negative words (e.g., “cold”, “unfavorable”, “bad”) (see Appendix section A.2). We find the average vector for these word sets, and then compute the cosine similarity of each sentence to the averaged vector. Figure 2 shows the words most associated to the positive and negative vectors. We can see that the positive list (left) includes intuitively positive-slanted words such as candid, certainly, confident, perfectly, doubtless, and sincere. The negative list (right) includes very different words, such as undesirable, disruptive, unfounded, intolerable, disturbing, unpleasant, and difficult. We define the sentiment \(S_{i}^{d}\) for sentence \(d\) in case \(i\) as the cosine similarity to the positive vector, minus the cosine similarity to the negative vector.

Next, we aggregate these sentence level statistics to the case level. Let \(W_{i}^{k}\) denote

\(^5\)Based on these word clouds, we ran the analysis for a subset of well-defined targets (business, congress, federal, elderly, illegal, labor unions, military, police, supreme court and young), and obtained similar results.
Figure 2: Positive and Negative Sentiment Language

(a) Positive Sentiment

(b) Negative Sentiment

Notes: Most similar words in the embedding space to the average vector for the lexicon of positive words (left) and negative words (right). See text for details.

the vector of sentence similarities to target \( k \) in case \( i \) and let \( S_i \) denote the vector of sentence sentiments in case \( i \). We construct the case-level sentiment towards target \( k \) as \( S_i^k = S_i \cdot W_i^k \), the dot product of these two vectors.

Finally, we aggregate to the circuit-year level. Let \( C_{ct} \) be the set of cases filed in circuit \( c \) during year \( t \). We average \( S_i^k \) across all cases \( i \) for each target \( k \). The main treatment regressor is, therefore,

\[
S_{ckt} = \frac{1}{|C_{ct}|} \sum_{i \in C_{ct}} S_i^k,
\]

the average case-level sentiment toward \( k \) for each case in circuit-year \( ct \). In Appendix Figures A.4 and A.5 we display our measure of judicial sentiment for each target group over time and by circuit, respectively.

3.3. Judge Characteristics Instruments

Finally, we collected the biographical information of judges that have been assigned to at least one case in the period 1964 – 2008. We match each judge with data from the Federal Appeals and District Court Attribute Data.\(^6\) We integrate this information with data from the Federal Judicial Center’s biographies of judges and previous data collection.

\(^6\)http://www.cas.sc.edu/poli/juri/attributes.htm
We have a total of 61 variables that refer to judges’ biographical characteristics. For instance, these variables include: age, geographic history, education, occupational history, governmental positions, military service, religion, race, gender, and political affiliations.

We assign judge characteristics to cases as follows. Let $J_i$ give the average characteristics for the three judges assigned to case $i$. Next, let $W^k_i$ be the average similarity of case $i$ to target $k$. Then, the vector of judge characteristics randomly assigned to target $k$ in circuit $c$ during year $t$ is

$$J_{ckt} = \frac{1}{|C_{ct}|} \sum_{i \in C_{ct}} W^k_i J_i,$$

the vector of judge characteristics, weighted by the similarity to target $k$ of the cases to which the judges are assigned. Exogenous variation in these characteristics due to random assignment of judges to cases is used in our empirical strategy.

4. Empirical strategy

We would like to procure causal estimates of judge sentiment on people’s social attitudes. A simple OLS regression would have endogeneity issues and resulting estimates would likely be biased. This source of endogeneity can be due to both omitted variable bias and reverse causality. For instance, an omitted variable bias would threaten our estimates if there are unobservable factors simultaneously affecting individuals’ and judges’ preferences towards a certain target group in a given year and circuit. A reverse causality issue would appear because of the influence that citizens (preferences) have on policies and laws which, in turn, could directly or indirectly affect judge opinions.

To address these endogeneity concerns we take an instrumental variables approach. We exploit the random assignment of judges to federal circuit courts as a source of exogenous variation. Importantly, the characteristics of the assigned judges to a case are as good as random once conditioned on their distribution in a given circuit-year.

$J_{ckt}$ contains a large number of potential instruments (61 of them). Therefore standard 2SLS has a weak instruments problem with a low F-statistic in our context. We address this issue by drawing on recent developments in machine learning, to extract more predictive power from the instruments while avoiding over-fitting (Chernozhukov et al., 2017).
Our approach is to use regularized regression to construct instruments from first-stage cross-validated predictions. As preparation, we residualize the judge characteristics \(J_{ckt}\), as well as the circuit-target-year sentiment \(S_{ckt}\), on year-circuit fixed effects and then standardize to variance one. We then train an elastic net regression to predict sentiment using the judicial characteristics. Elastic net is a linear regression with a penalized cost function to shrink coefficients toward zero and avoid over-fitting (Zou and Hastie, 2005a). Using 10-fold cross-validation, we learned the cost-minimizing penalties: \(L1 = 0.0028\) and \(L2 = 0\) (equivalent to lasso regression with a mild penalty).\(^7\)

Let \(Z_{ckt}\) be the cross-validated prediction for \(S_{ckt}\) using the randomly assigned judge characteristics. It is a "clean" prediction in the sense that the coefficients are trained on out-of-fold data. Following Galasso and Schankerman (2014) and Sampat and Williams (2019), we use the predicted endogenous regressor \(Z_{ckt}\) as the instrument in our two-stage least-squares regressions. It is strongly predictive of sentiment, but far from collinear \((R^2 = 0.186)\), as shown in Figure 3’s scatter plot.\(^8\)

Figure 3: First stage relationship

Notes: Binscatter diagram for the first stage relationship (Coef. = 1.117, st. err. = 0.111, \(R^2 = 0.186\))

\(^7\)We produced our main results using ridge regression \((L2 but not L1 penalty, which does not impose sparsity) to produce the instrument, and got similar sign and significance of coefficients.

\(^8\)In unreported estimates we reach similar results using LASSO to select the optimal set of instruments (Belloni et al., 2012; Zou and Hastie, 2005b).
We define the first-stage equation as:

\[ S_{ckt} = \gamma_k + \gamma_{ct} + \gamma Z_{ckt} + \eta_{ckt} \]  

(3)

where \( S_{ckt} \) is the weighted average sentiment toward target \( k \) in cases published in circuit \( c \) during year \( t \). \( Z_{ckt} \) is the machine-learning-predicted instrument. \( \gamma_{ck} \) is a set of dummy variables (fixed effects) for each circuit-year and \( \gamma_k \) is a set of dummy variables (fixed effects) for each target. \( \eta_{ckt} \) is the error term.

The second-stage estimating equation is:

\[ Y_{ckt} = \alpha_k + \alpha_{ct} + \beta \hat{S}_{ckt} + \epsilon_{ckt} \]  

(4)

where the \( \alpha \)'s are fixed effects, as previously defined. \( \hat{S}_{ckt} \) is the predicted target sentiment as computed from the first stage – equation (3). \( Y_{ckt} \) is the thermometer response from ANES. \( \beta \) is our coefficient of interest as it gives the average effect of judge writing sentiment on individuals’ attitudes.

5. Results

5.1. Main results

<table>
<thead>
<tr>
<th>Table 1: Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>OLS</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Judges’ sentiment</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Year FE</td>
</tr>
<tr>
<td>Circuit FE</td>
</tr>
<tr>
<td>Year FE X Circuit FE</td>
</tr>
<tr>
<td>Target FE</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>N observations</td>
</tr>
</tbody>
</table>

Notes: The dependent variable is the thermometer score for all respondents in the ANES by circuit-target-year. **Judges’ sentiment** is the text-based average sentiment by circuit-target-year. All variables are centered and standardized by target. Standard errors clustered by circuit-year in parenthesis. * \( p < 0.1 \), ** \( p < 0.05 \) and *** \( p < 0.01 \).

In Table 1 we present our main findings. The first three columns show results using OLS regressions while the last three are the results from the instrumental variables...
approach. The OLS estimates report negative and statistically significant coefficients which are stable to the inclusion of different sets of fixed effects. The 2SLS coefficients are also of negative sign and significant. Interestingly, their sizes do not deviate too much from those estimated with OLS regression. The reported F-stat confirms the relevancy of our instruments in all three specification. Overall, we find robust evidence of a negative and significant effect of judges sentiment on individuals attitudes. The 2SLS coefficients move from a maximum of -0.167 (st. error= 0.051) in column (5) and a minimum of -0.122 (st. error= 0.058) in column (6). Therefore, a positive shift in the opinion of judges towards a given target group of one standard deviation decreases citizen opinion towards that same group in a range of 12-16% of a standard deviation. These results indicate that reactions to changes in judges opinion are in line with predictions from the thermostatic model rather than legitimization theory, similarly to that highlighted by Chen et al. (2016) and Ura (2013).

5.2. Effect dynamic

Next we are interested in a potential time dynamic of the effect. First we would like to check whether there are significant pre-trend, and secondly how long the effect lasts. To this end we replicate our estimates, keeping the specification from column (6) of Table 1, but using leads and lags of values of our outcome variable. As the ANES survey is conducted every two years, we use biennial lags.

Table 2 provides evidence on the dynamic effect. First, in column (1) we report the coefficient for a regression that uses as a dependent variable the 2-year lag of the outcome. The coefficient is negative and insignificant, which suggests the unlikely presence of a confounded pre-trend in our setting. Column (2) reports the main result already discussed in the previous section. Next, Column (3) shows the estimated coefficient with

---

9All variables are centered and standardized by target.
10We provide additional evidence on the presence of relevant variation in our data. Specifically, Appendix Figure A.6 reports histograms of the distribution of target similarity scores for each target group residualized on circuit-year fixed effects (the relevant randomization block).
11We confirm this negative effect when using a reduced sample which includes only target groups for which word associations performed better. If we focus only on: business, congress, federal, elderly, illegal, labor unions, military, police, supreme court and young, the 2SLS coefficient equal -0.229 (st. error=0.123) and F-stat=26.9.
12Not all questions appear in all issues, so the results are limited to those target group for which we have the respective leads/lags values.
13For completeness, we re-estimated column (1) of Table 2 by using each time a different set of FE. The exclusion of FE tends to increase the confidence intervals making the coefficient significantly different from 0 every time we include none or just one of the fixed effects.
Table 2: Results – Leads/Lags of Dep. variable

<table>
<thead>
<tr>
<th></th>
<th>2 years before</th>
<th>Same year</th>
<th>2 years after</th>
<th>4 years after</th>
<th>6 years after</th>
</tr>
</thead>
<tbody>
<tr>
<td>Judges' sentiment</td>
<td>-0.113 (0.083)</td>
<td>-0.122** (0.058)</td>
<td>-0.215** (0.085)</td>
<td>-0.094 (0.092)</td>
<td>-0.051 (0.189)</td>
</tr>
<tr>
<td>Year FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Circuit FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Year FE X Circuit FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Target FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>F-stat</td>
<td>65.546</td>
<td>101.201</td>
<td>77.585</td>
<td>46.285</td>
<td>29.677</td>
</tr>
<tr>
<td>N observations</td>
<td>1687</td>
<td>2678</td>
<td>1684</td>
<td>1322</td>
<td>1004</td>
</tr>
</tbody>
</table>

Notes: The dependent variables are the leads and lags of the thermometer score for all respondents in the ANES by circuit-target-year as reported in columns head. Judges’ sentiment is the text-based average sentiment by circuit-target-year. All variables are centered and standardized by target. Standard errors clustered by circuit-year in parenthesis. * p < 0.1, ** p < 0.05 and *** p < 0.01.

a 2-year lead of the outcome variable. Indeed, we see that the effect is persistent in the next period and actually it is a somewhat larger coefficient of 0.215 (st. error=0.085). The last two columns, leading the dependent variable by 4 and 6 years respectively, report insignificant coefficients. Further, we test differences across the coefficients reported in the first three columns of Table 2. We find that the effect on the attitude of the general public today is not significantly larger than the effect on the attitude of the general public two years ago, but they both are significantly different from the effect on population attitude reported in the next two years.

These results are consistent with a causal effect of judicial sentiments on social attitudes, which strengthens over the next two years, and then fades out in the following years. Again, these results are line with previous evidence suggesting that the backlash from court rulings to public opinion does not persist in the long run (Chen et al., 2016; Ura, 2013).

5.3. Effect by target group

In the previous regressions we pooled together different target groups. A potential concern is that we are not accounting for the fact that citizens might have different reactions, either in the intensity or the direction, depending on the target of interest. In addition, judges’ characteristics might have a heterogeneous effect on sentiment depending on the topic under discussion (violating monotonicity). Therefore the selection of instruments could differ from one target group to another.
This section seeks to address these concerns by running the 2SLS analysis separately for each target group. The empirical strategy is the same as that articulated above. The difference is that the instrument construction and subsequent analysis are limited to particular targets – so we now have circuit-year level data. Given we are focusing on single targets independently, we cannot include year-circuit fixed effects. Instead, our specification includes year fixed effects and circuit fixed effects (not interacted).

The 2SLS estimates by target are summarized in Figure 4.\textsuperscript{14} In the figure we report the coefficients, and confidence intervals, of 19 different 2SLS estimates. Interestingly, we find that the effect of judge sentiments on people’s attitudes varies by target group. Consistent with the aggregate estimates, most of the coefficients are negative (12 out of 19) and none of the positive coefficients is significantly different from 0. The F-statistics reported in the table reveal that the instruments are relevant (F-stat>10) for 14 of the 19 regressions. We find significant backlash effect when judges write about (in order of intensity) women (although the first stage is weak), the Supreme Court, Congress,

\textsuperscript{14}For regression table output see Appendix Table A.2, which reports OLS and 2SLS estimates for each target
conservatives, the elderly, police, and white people.

6. Conclusion

In summary, this paper has combined natural language processing, machine learning, and causal inference techniques to provide a method for analyzing the impacts of judicial sentiments. There are many research opportunities opened up by this method. Our approach could be used to develop sentiment metrics in other corpora, such as political speeches or news articles, and toward other targets (not just social groups but also concepts such as democracy or inequality for example). The cross-validated instruments approach could be applied in other circumstances with many weak instruments that are predictive of treatment. Random assignment of judges, along with judicial texts, could be used to analyze causal impacts of other features of legal language.

In our application, we study the expressive impacts of judge writing on social attitudes. We find that judge writing sentiment does have an impact on how people view the social groups that judges write about. Citizen opinions tend to move in the opposite directions of the ones expressed in judge rulings. These results add to the existing literature confirming the thermostatic model of public opinion. There are still many open questions raised by these results. What is the mechanism by which language in judicial opinions affects the political attitudes of members of the mass public? While most people don’t read opinions, organizations like the ACLU do mobilize media attention to important cases (Weinrib, 2015). Tracing these channels is an important area for future work.
Bibliography


Alesina, A. and N. Fuchs-Schündeln (2007, September). Goodbye lenin (or not?): The  
1528.


Ash, E. and D. L. Chen (2017). Judge embeddings: Toward vector representations of  

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012). Sparse models and  
methods for optimal instruments with an application to eminent domain.  
*Econometrica* 80(6), 2369–2429.

of Economic Research.


Boyd, C. L., L. Epstein, and A. D. Martin (2010). Untangling the causal effects of sex  

Bromley, R. V. (1994). Journalists assess computers’ value in covering us courts of  


from language corpora contain human-like biases.  
*Science* 356(6334), 183–186.

Workers under State Socialism.  
*The Review of Economics and Statistics* 101(2), 233–  
248.


Evidence from Random Variation in Abortion Jurisprudence.  
TSE Working Papers 16-723, Toulouse School of Economics (TSE).


Figure A.1: Example Thermometer Question - ANES 2012

- 100° Very warm or favorable feeling
- 85° Quite warm or favorable feeling
- 70° Fairly warm or favorable feeling
- 60° A bit more warm or favorable feeling than cold feeling
- 50° No feeling at all
- 40° A bit more cold or unfavorable feeling than warm feeling
- 30° Fairly cold or unfavorable feeling
- 15° Quite cold or unfavorable feeling
- 0° Very cold or unfavorable feeling
Figure A.2: Sentiment ANES over time
Figure A.3: Sentiment ANES, by circuit
Figure A.4: Judicial Sentiment, over time
Figure A.5: Judicial Sentiment, by circuit...
Figure A.6: Residuals plots
Figure A.7: Word Clouds for Target Groups

(a) Black
(b) Business
(c) Catholic
(d) Congress

(e) Conservative
(f) Democrat
(g) Elderly
(h) Federal Gov.

(i) Illegal
(j) Labor Unions
(k) Liberal
(l) Military

(m) Police
(n) Protestant
(o) Republican
(p) Supreme Court

(q) White
(r) Woman
(s) Young
Table A.1: Summary statistics ANES

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>st. dev</th>
<th>n. waves</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>65.468</td>
<td>5.251</td>
<td>21</td>
</tr>
<tr>
<td>Business</td>
<td>53.998</td>
<td>5.553</td>
<td>16</td>
</tr>
<tr>
<td>Catholic</td>
<td>66.108</td>
<td>5.290</td>
<td>12</td>
</tr>
<tr>
<td>Congress</td>
<td>56.152</td>
<td>4.557</td>
<td>9</td>
</tr>
<tr>
<td>Conservative</td>
<td>58.165</td>
<td>5.416</td>
<td>21</td>
</tr>
<tr>
<td>Democrat</td>
<td>62.942</td>
<td>6.353</td>
<td>21</td>
</tr>
<tr>
<td>Elderly</td>
<td>80.708</td>
<td>4.001</td>
<td>10</td>
</tr>
<tr>
<td>Fed. Gov.</td>
<td>54.375</td>
<td>5.967</td>
<td>8</td>
</tr>
<tr>
<td>Illegal immigrants</td>
<td>37.307</td>
<td>8.172</td>
<td>5</td>
</tr>
<tr>
<td>Labor union</td>
<td>55.058</td>
<td>5.475</td>
<td>19</td>
</tr>
<tr>
<td>Liberal</td>
<td>52.531</td>
<td>6.500</td>
<td>21</td>
</tr>
<tr>
<td>Military</td>
<td>71.251</td>
<td>7.050</td>
<td>15</td>
</tr>
<tr>
<td>Police</td>
<td>74.831</td>
<td>5.097</td>
<td>7</td>
</tr>
<tr>
<td>Protestant</td>
<td>72.142</td>
<td>6.550</td>
<td>7</td>
</tr>
<tr>
<td>Republican</td>
<td>55.223</td>
<td>7.902</td>
<td>21</td>
</tr>
<tr>
<td>Sup. court</td>
<td>62.882</td>
<td>3.927</td>
<td>8</td>
</tr>
<tr>
<td>White</td>
<td>73.888</td>
<td>5.860</td>
<td>19</td>
</tr>
<tr>
<td>Woman</td>
<td>78.106</td>
<td>3.978</td>
<td>4</td>
</tr>
<tr>
<td>Young</td>
<td>76.454</td>
<td>3.777</td>
<td>5</td>
</tr>
</tbody>
</table>
Table A.2: Results, by target group

<table>
<thead>
<tr>
<th></th>
<th>OLS 2SLS</th>
<th>OLS 2SLS</th>
<th>OLS 2SLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>Business</td>
<td>Catholic</td>
</tr>
<tr>
<td>Judges’ sentiment</td>
<td>-0.005</td>
<td>0.100</td>
<td>-0.147</td>
</tr>
<tr>
<td></td>
<td>(0.087)</td>
<td>(0.175)</td>
<td>(0.127)</td>
</tr>
<tr>
<td>F-stat</td>
<td>24.038</td>
<td>30.873</td>
<td>15.778</td>
</tr>
<tr>
<td>N observations</td>
<td>226</td>
<td>226</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>Congress</td>
<td>Conservative</td>
<td>Democrat</td>
</tr>
<tr>
<td>Judges’ sentiment</td>
<td>0.041</td>
<td>-0.488**</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(0.138)</td>
<td>(0.223)</td>
<td>(0.092)</td>
</tr>
<tr>
<td>F-stat</td>
<td>37.415</td>
<td>28.020</td>
<td>23.358</td>
</tr>
<tr>
<td>N observations</td>
<td>103</td>
<td>103</td>
<td>226</td>
</tr>
<tr>
<td></td>
<td>Elderly</td>
<td>Federal Gov.</td>
<td>Illegal</td>
</tr>
<tr>
<td>Judges’ sentiment</td>
<td>-0.059</td>
<td>-0.374***</td>
<td>-0.164</td>
</tr>
<tr>
<td></td>
<td>(0.121)</td>
<td>(0.136)</td>
<td>(0.125)</td>
</tr>
<tr>
<td>F-stat</td>
<td>51.847</td>
<td>0.229</td>
<td>11.422</td>
</tr>
<tr>
<td>N observations</td>
<td>110</td>
<td>110</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>Labor unions</td>
<td>Liberal</td>
<td>Military</td>
</tr>
<tr>
<td>Judges’ sentiment</td>
<td>-0.067</td>
<td>0.322</td>
<td>-0.131</td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td>(0.253)</td>
<td>(0.157)</td>
</tr>
<tr>
<td>F-stat</td>
<td>10.146</td>
<td>35.960</td>
<td>9.681</td>
</tr>
<tr>
<td>N observations</td>
<td>205</td>
<td>205</td>
<td>226</td>
</tr>
<tr>
<td></td>
<td>Police</td>
<td>Protestant</td>
<td>Republican</td>
</tr>
<tr>
<td>Judges’ sentiment</td>
<td>-0.155</td>
<td>-0.282*</td>
<td>-0.254**</td>
</tr>
<tr>
<td></td>
<td>(0.124)</td>
<td>(0.155)</td>
<td>(0.108)</td>
</tr>
<tr>
<td>F-stat</td>
<td>22.237</td>
<td>9.418</td>
<td>11.036</td>
</tr>
<tr>
<td>N observations</td>
<td>69</td>
<td>69</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Supreme Court</td>
<td>White</td>
<td>Women</td>
</tr>
<tr>
<td>Judges’ sentiment</td>
<td>-0.363**</td>
<td>-0.524***</td>
<td>-0.129**</td>
</tr>
<tr>
<td></td>
<td>(0.142)</td>
<td>(0.144)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>F-stat</td>
<td>92.022</td>
<td>33.803</td>
<td>2.735</td>
</tr>
<tr>
<td>N observations</td>
<td>89</td>
<td>89</td>
<td>204</td>
</tr>
<tr>
<td></td>
<td>Young</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judges’ sentiment</td>
<td>-0.015</td>
<td>-0.505</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.156)</td>
<td>(0.485)</td>
<td></td>
</tr>
<tr>
<td>F-stat</td>
<td>8.202</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N observations</td>
<td>52</td>
<td>52</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Circuit and year fixed effects included. Standard errors clustered by circuit-year in parenthesis. * p < 0.1, ** p < 0.05 and *** p < 0.01.
A.2. List of words to identify the attributes and target groups

Attributes

Negative: cold, unfavorable, bad, adverse, antagonistic, calamitous, damaging, destructive, disadvantageous, hostile, negative, objectionable, ominous, troublesome, unfriendly, contrary, discommodious, ill, ill-advised, improper, inadvisable, inauspicious, inconvenient, inexpedient, infelicitous, inimical, inopportune, late, low, malapropos, opposed, poor, regrettable, tardy, threatening, unfit, unfortunate, unlucky, unpromising, unpropitious, unseasonable, unseemly, unsuited, untimely, untoward, wrong.

Positive: warm, favorable, good, agreeable, benign, encouraging, positive, supportive, sympathetic, acclamatory, affirmative, amicable, approbative, approbatory, assenting, benevolent, benignant, commending, complimentary, enthusiastic, inclined, kind, kindly, laudatory, okay, praiseful, predisposed, reassuring, recommendatory, understanding, welcoming, well-disposed, well-intentioned.

Targets

Black: blacks, black, african, african-american, african-americans, negro, negroes
Business: business, businesses, corporation, corporations, factory, firm, market, organization, partnership, shop, store, venture
Catholic: catholics, catholic
Congress: congress, parliament, legislature, senate, house, representative, senators, representatives
Conservative: conservatives, conservative
Democrat: democrat, democrats
Elderly: elderly, aged, old
Federal government: federal, government, executive
Illegal: illegal, immigrants, undocumented
Labor unions: labor, unions, union, trade-union
Liberal: liberals, liberal
Military: military, army
Police: policemen, police, policeman
Protestant: protestant, protestants
Republican: republican, republicans
Supreme Court: supreme, court
White: whites, white, caucasian, caucasians
Woman: woman, women
Young: youngster, youth, budding, adolescent