Nationalized Elections, Localized Campaigns? A Text Analysis Approach*

Derek Holliday[†]

April 03, 2022

Abstract

Recent research documents the "nationalization" of election results in the United States; national- and state-level vote shares for candidates of the same party are increasingly correlated and election winners tend to be copartisans. Does the nationalization of electoral results extend to campaign rhetoric itself? In this paper, I utilize quantitative text analysis techniques to determine the degree to which the rhetoric of gubernatorial candidates can or cannot be distinguished from the rhetoric of their national counterparts. The results show a more nuanced picture of nationalized rhetoric; in some forums, such as televised debates, gubernatorial speech is easily distinguishable from the speech of presidential candidates. In less structured formats, such as advertisements and social media, gubernatorial speech is still mostly relevant to state topics, but is slightly more likely to reference national topics. These results speak to the complex information environment voters may encounter in downballot contests.

Introduction

A growing literature in political science focuses on the "nationalization" of U.S. politics. Generally, this phenomenon refers to national political actors and issues influencing state and local political activity. Perhaps the most prominent results in this literature point to the increasing correlation between vote shares of presidential and downballot candidates of the same party (Hopkins 2018). This nationalization of election results has potentially problematic effects on the quality of representation from the winners of such downballot contests. If voters are evaluating candidates not based on job demands specific to the elected office (which at the state and local level is in many cases orthogonal to the contours of national politics) but simply on the partisan affiliation of the candidates, then holding an office-holder accountable becomes significantly more difficult.

A large portion of the nationalization literature links nationalized election results to a lack of information specific to downballot races. If voters have access to distinct downballot political information, they are more likely to make voting decisions less directly influenced by copartisanship with national candidates (G. J. Martin and McCrain 2019; Moskowitz 2021). While these studies focus predominantly on the media environment, one source of political information remains understudied; the content of political campaigns themselves. A growing narrative has emerged within popular media that the content of downballot political campaigns has grown more similar to the campaigns of candidates for national office. By referencing the talking points of prominent national candidates, downballot candidates are perhaps able to easily identify themselves to voters who are already more familiar with the national candidates.

In this paper, I evaluate the extent to which the rhetoric of gubernatorial candidates reflects the topics referenced by their national counterparts. In doing so, I approach the potential nationalization of political rhetoric as a text classification problem by constructing a classification model based on topic representations of speech known to be of state or national providence. I then apply this classification model to a diverse corpus of political text; televised gubernatorial and presidential debates from 2000 to 2018, televised gubernatorial and presidential advertisements from 2008, and the Twitter activity of incumbent Governors and Members of Congress in 2018. The results show a more nuanced picture of gubernatorial campaign rhetoric; gubernatorial candidates overall tend to reference topic distinct from their national counterparts, but are more likely to "nationalize" their speech in televised ads and social media. These results have significant consequences for voter behavior; if voters are effectively exposed to gubernatorial campaign messaging, they are likely able to make voting decisions that aren't as susceptible to nationalized pressures. However, voters are also susceptible to appeals that *are* more nationalized, and if these messages are the ones highlighted by media outlets, then nationalized voting behavior may follow.

Nationalized Elections

The correlation between Presidential and downballot vote shares of copartisan candidates has steadily and steeply risen since the 1970s. Specifically, Hopkins (2018) shows the correlation between Presidential and gubernatorial candidates in midterm elections growing from a low of less than 0.2 in 1970 to a high of over 0.8 in 2010. Elections of presidents and governors occuring in the same year show a similar but less dramatic rise of less than 0.3 in 1968 to around 0.7 in 2012. Sievert and McKee (2019) similarly find the rate at which the same party won both the Presidential and Senatorial contests in a given state rose from 52% in 1980 to 84% in 2012, with Jacobson (2015) finding similar trends in U.S. House elections. Examining Supreeme Court elections, Weinschenk et al. (2020) find a nearly 1-1 relationship between county-level Democratic Presidential and state Supreme court vote shares from 2000 t0 2018 in partisan elections.

A number of mechanisms have been theorized regarding the nationalization of American elections, which roughly fall under two (non-mutually exclusive) categories: identity and information. The identity mechanism explains nationalization as an extension of partisanship and polarization; that is, partisanship is an affective, expressive identity, so we should expect voters to vote for their preferred party (or against their non-preferred party) in any context (national, state, or local) as an expression of in-group solidarity and/or out-group antipathy (Huddy and Bankert 2017; Iyengar, Sood, and Lelkes 2012). For example, Abramowitz and Webster (2016) note the increase in straight-ticket party voting is associated with high levels of out-party antipathy, also known as negative partisanship. Webster (2020) finds a similar association between straight-ticket voting and anger. However, the extent to which voters are willing to make voting decisions purely on party identification may be bounded. Using a conjoint design, Mummolo, Peterson, and West-wood (2021) find voters punish excessive deviation from preferred positions on salient policies by co-partisan candidates. Costa (2021) also uses a conjoint design to find voters prefer candidates who provide substantive representation and constituency service over partisan affect. So, while partisanship contains information about ideology, it is not a substitute for ideology in the eyes of voters.

The information mechanism proposes that decreasing access to meaningful information about state and local political contests drives increased nationalization. Without specific policy information to evaluate the candidates, voters simply default to the candidate of their preferred party. Hayes and Lawless (2018) note the steep decline in access to local news in the last decade, with a 10% reduction in issue coverage and a 33% reduction in the coverage of candidate traits in U.S. House of Representatives contests between 2010 and 2014. Since 2004, one in five newspapers has closed (Abernathy 2018). This decline in access to local news is associated with increase in nationalized news content and voting behavior (G. J. Martin and McCrain 2019; Moskowitz 2021). The more information voters have besides the party identification of the candidates, the

more likely they are to make split-ticket decisions.

Both the identity and information hypotheses stress the importance of voters using available signals, including partisanship and policy stances, to make a judgment on candidate type. Even in the state context, nationalized signals are not devoid of information on candidate type. Platforms have homogenized and national and state parties are seen as more singular than separate (Caughey, Dunham, and Warshaw 2018; Hopkins 2018). The dimensions of state politics largely mirror the left-right contours of national politics (Caughey and Warshaw 2016; Shor and McCarty 2011).

The core concern of the nationalization literature is the pressure the phenomenon puts on political representation under a federal system. While national and state institutions often overlap and trade jurisdiction over policies areas from year to year, there still remain areas of functional responsibility unique to each (Beer 1978; Kousser 2014). The threat to representation occurs when voters make state-level decisions based on national-level criteria or information, potentially struggling to hold state officials accountable for functions of their offices.

Nationalized Campaigns?

The mechanisms discussed above largely focus on voter-level factors that influence the electorate's propensity to vote in a nationalized manner. While elite behavior certainly pushes the trend in a similar manner through cross-state homogenization of party platforms and the more rigorous sorting of partisans along ideological lines, the effects that candidates themselves can have in any given election remains understudied.

Media portrayals of gubernatorial campaigns have stressed their "nationalized" content. For example, national media outlets characterized the 2019 gubernatorial races in Kentucky, Louisiana, and Mississippi as being nationalized due to Donald Trump's personal involvement in some of the races and an emphasis on impeachment of Trump as a campaign issue (Manchester 2019; J. Martin 2019; Rojas and Alford 2019). Other outlets gave similar appraisals of other races, including Washington in 2016 ("Inslee... was happy to nationalize the governor's race, sounding at many events like he was running against Trump"), West Virginia in 2011 (where the Republican Governor's Association spend $3.5 million in ad buys in an attempt to link the democratic candidate to Obamacare), and Texas in 2010 ("Mr. Perry turned the race into a referendum on federal spending") (Brunner 2016; Catanese 2011; McKinley Jr. 2010). Some Governors have engaged in nationalized rhetoric themselves, with Governor Gavin Newsom of California characterizing supporters of the 2021 recall election as "a partisan, Republican coalition of national Republicans, anti-vaxxers, Q-Anon conspiracy theorists and anti-immigrant Trump supporters." At the very least, candidates for state office do not feel bound only address or espouse policies, individuals, and organizations exclusive to their own states.

4

Some empirical evidence exists to suggest political rhetoric has broadly nationalized in the same manner elections have. Das et al. (2021) analyze the tweets of incumbent Members of Congress, Governors, and mayors in 2018 utilizing a topic modeling approach to ascertain the level of semantic similarity between the different office holders. They find Members of Congress and Governors are almost indistinguishable in terms of their topical similarity, while mayors still seem to tweet about distinct topics. These findings give pause to the "all politics is national" hypothesis, at least at the local level, but still suggest gubernatorial rhetoric has nationalized parallel to the nationalization of electoral results. It is important to note, however, that the Twitter activity analyzed by Das et al. (2021) is not specific to campaigning, focuses on sitting incumbents, and may also contain content that is apolitical in nature. For example, New Jersey Governor Phil Murphy has, since February 2022, tweeted about his daily Wordle score, the Saint Peter's University men's basketball team, and changing the state bird of New Jersey to the middle finger for April Fools.

What incentives exist to nationalize campaign rhetoric in gubernatorial contests? Both the identity and information mechanisms prove potentially useful in this context. If voters are predominantly motivated by simple party identification, nationalizing one's campaign appeals in gubernatorial contests may boost signals of partisan type by linking candidates to more traditional, national-level policy positions. This makes candidates more "identifiably" Republican or Democrat. Alternatively, nationalizing campaign appeals may have the effect of diluting the pool of locale-specific information available to voters, instead focusing the information environment on national signals of partisan type.

While there are many plausible reasons to nationalize a gubernatorial campaign, there are equally plausible reasons to keep a campaign localized. The most obvious reason is voters may recovnize a candidate running on nationalized appeals has no jurisdiction over the issue being discussed. Current research is divided on the extent to which voters hold politicians accountable for conditions under their jurisdiction; Arceneaux (2006) finds survey respondents tend to attribute credit/blame to offices which they (fairly accurately) assign functional responsibility to, whereas Brown (2010) finds partisanship moderates the attribution of functional responsibility and subsequent credit/blame. Benedictis-Kessner and Warshaw (n.d.) find some evidence for both conclusions suing time series, cross-sectional models; voters routinely hold the president's party responsible for local economic conditions, but also hold governors accountable for such conditions. Therefore, the incentives for candidate to nationalize gubernatorial campaigns seems mixed.

Methodological Approach

I approach the potential nationalization of political rhetoric during campaigns as a text classification problem. That is, the collection of words spoken or otherwise disseminated during a campaign can be categorized as

having either national or state content. In doing so, I leverage a supervised machine-learning approach and quantitative text-analysis techniques. The workflow involves (1) defining a training dataset where document "class" (national or state providence) is known, (2) quantitatively representing text using a topic-modelling approach, (3) fitting a classification model using the training data and quantitative text representation, and (4) predicting the class of a testing dataset using the trained classification model. I describe each step of the workflow in more detail below. The result of this workflow is a state/national classification rate of political campaign rhetoric at the state level, where a high national classification rate signifies campaign activities sharing similar characteristics to national-level rhetoric and where a high state classification rate signifies activities sharing similar characteristics to state-level rhetoric.

Training Data

A classification approach requires a model to be fitted with data of known classes (state or national). This precludes the possibility of using words from campaign activities as part of the training process; by design, these activities are of ambiguous "class," being potentially more state or national in content than their providence would suggest. Additionally, the training data must be substantively representative of the classes I aim to predict. To this end, I synthesize a corpus of Presidential speeches (State of the Union addresses and opposition responses, inaugural addresses, official statements, and national party platforms) and gubernatorial speeches (State of the State addresses and budget addresses) representing national and state political content, respectively. The final training corpus contains 1,038 speeches and documents, 227 national and 811 state, spanning 2000-2018.

This training corpus is meant to distinguish between national and state political content via the policy discussions in each respective sphere. State of the Union/State speeches are particularly useful in this context, as they often involve explicit references to policy accomplishments and goals. However, this does not prevent certain words and phrases from existing in either the state or national contexts that are highly predictive of a particular class but devoid of policy content. For example, most gubernatorial State of the States addresses include the state names themselves and the names of residents for those states (such as "Californians" or "Hoosiers"). Including these words during the model fitting process would potentially allow for the model to "cheat" and accurately predict class not from the policy content of a speech but from these cheap signals of state providence. Therefore, during standard text preprocessing of the training data (stopwording, lemmatization, removal of very short or very rare words), I also remove all state names, names of state residents (including nicknames), references to the level of office (besides presidential), common audio transcription tags (laughter, applause), and common words without policy meaning (year, will, thank, etc.).

Quantitative Text Representation

To predict class, the text of any document needs to be quantitatively represented. Many options exist for doing so, such as raw frequency of words, term frequency-inverse document frequency, bigrams, trigrams, and more. In classic supervised learning approaches, the choice of quantitative text representation is driven by best out-of-sample prediction accuracy. In this particular application, however, the substantive meaning of the text representation is equally important, as I want to interpret the classification rate as a meaningful indicator of overall national content. Therefore, I represent all text in this paper as estimated topic proportions using a structural topic model (STM) approach.

STM treats texts as "bags of words." Like the latent Dirichlet allocation (LDA) approach, STM assigns words to topics and topics to documents probabalistically (Roberts, Stewart, and Airoldi 2016). The output gives word probabilities associated with each topic adn the topic proportions for each document. STM builds upon LDA by allowing topic probabilities to vary according to researcher-specified covariates, allowing for resulting topics to more closely approximate the theorized data-generating process. In this application, I allow topics to vary as a function of providence (state or national) and year (binned in two-year intervals). Functionally, this means every document x_i is represented by a length k vector of topic proportions $\theta_{1 \ldots k}$, which are then used as the features of the classification model[1].

The application of the STM process in this paper can be thought of as a text-as-data manifestation of other dimension-reduction techniques (such as principal components analysis) in machine learning. Methodologically, the process alleviates problems resulting from high-dimensional data such as data sparsity, computational complexity, and overfitting. Substantively, STM provides a more interpretable output than raw term frequencies and more closely captures the theoretical thrust of the nationalization hypothesis; certain collections of words are more "national" in nature than others. This is significantly more theoretically meaningful than any single word being an indicator of state or national providence.

Classification Model

I train an assortment of classification models using the STM-generated topic proportions of the training data: logistic regression, naive Bayes, regularized logistic regression (lasso), support vector machine (SVM), and boosted gradient descent (XGBoost). Before fitting the models, I randomly hold out 20 percent of the training data as a validation set to evaluate out-of-sample model performance. Models requiring hyperparameter tuning are first evaluated for performance using area under the ROC curve (AUC) with 10-fold cross-

---

[1]k is a researcher-defined hyperparameter determining the number of documents. The results presented in this paper use $k = 40$, which was determined to have the best balance of semantic coherence, held-out likelihood, and minimization of residuals. See appendix A1 for details.

validation. Furthermore, because of the class imbalance within the training data (significantly more state documents than national documents), I use a bootstrap-based synthetic oversampling technique called ROSE (Random Over-Sampling Examples) to prevent models from always predicting the majority class. The performance of each model for predicting the class of the documents in the validation set are shown in table 1.

Table 1: Classification Model Performance on Heldout Documents

Model	National Documents		State Documents		Accuracy	AUC
	Correct	Incorrect	Correct	Incorrect		
Logistic Regression (Nonpenalized)	39	1	166	2	0.9856	0.9987
Naive Bayes	37	3	162	6	0.9567	0.9903
Penalized Logistic Regression (Lasso)	39	1	165	3	0.9808	0.9990
Boosted Gradient Descent (XGBoost)	27	13	168	0	0.9375	0.9930
Support Vector Machine	39	1	167	1	0.9904	0.9972

All the models perform remarkably well, indicating there is sufficient textual differentiation between state and national rhetoric to perform classification in this manner. This also alleviates a potential concern that the state documents of the training data may themselves contain nationalized rhetoric. While such rhetoric may exist, the models are able to successfully determine which topics are most associated with state and national origin. I use the unpenalized logistic regression model as the final prediction model for the remainder of this paper because it has the highest AUC and easily interpretable coefficients, but see appendix A2 for prediction results from the other models.

Figure 1 shows the log odds coefficients associated with each of the topics from the unpenalized logistic regression. Negative coefficients indicate predictors of national relevance, while positive coefficients indicate predictors of state relevance.

Figure 1: Logistic Regression Coefficients (Unpenalized)
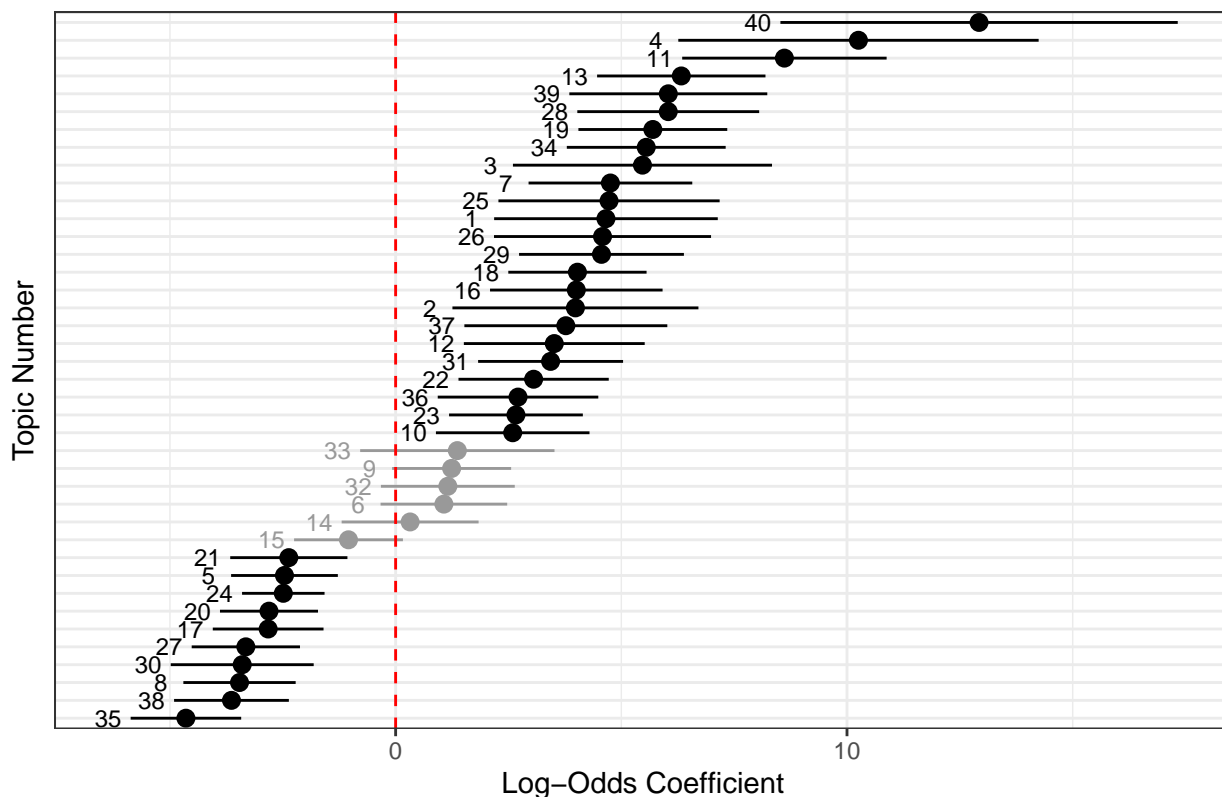
Prediction Results

The final step of the workflow is to fit the trained model to text from political campaigns to predict the class of each document. I consider three different mediums through which political candidates communicate with voters: televised debates, TV advertisements, and social media (Twitter). There are both theoretical and technical reasons to believe classification performance should vary by medium, which I will discuss in greater depth below.

Televised Debates

I first consider the potential nationalization of rhetoric during televised political debates. Specifically, I analyze an original corpus of 397 electoral debates (86 presidential and 311 gubernatorial) between 2000 and 2018 retrieved from closed-captioned transcripts from the C-SPAN video archives, which were originally broadcast either directly on C-SPAN or through local public affiliates[2].

Research on gubernatorial debates is rare, but the few studies that have been conducted conclude candi-

---

[2]Transcripts were retrieved through a combination of headlines web browsing and scraping. Transcripts for non-closed captioned videos are not available.

dates largely focus on policy positions rather than character (Benoit, Brazeal, and Airne 2007) and viewers of debates are often able to correctly identify the eventual winner of the contest (Benjamin and Shapiro 2009). Research on the effects of presidential debates largely conclude such events have some short-term effect on candidate preference (Hillygus and Jackman 2003) and issue knowledge/salience (Benoit, Hansen, and Verser 2003). Practically, the debate context helps control for candidate-level confounders such as ideology, campaign resources, or campaign activity level that may bias results in a different context (such as television advertisements or social media).

While there is no particular reason why the messaging content of debates would deviate substantially from other mediums, there are reasons to believe there would be a high hurdle to find evidence of nationalization. The length of debates (typically at least an hour long) allows for greater depth and breadth of discussion across policy issues, making state-specific content perhaps more likely to appear in gubernatorial contests. The moderator of the debate (typically a member of the local media) may push candidates to give positions on more local issues of interest to that media market. The candidates themselves may believe the audience of such debates to be fairly well-informed, makes less detailed or policy-oriented appeals less effective. Still, media coverage of the debates the next day may focus on the headline-grabbing nationalized appeals made during the debates, and candidates are easily able to answer the questions they want to answer instead of the questions that are asked of them.

Figure 2 shows the results of applying the trained classification model to the C-SPAN debates corpus. The left panel shows the confusion matrix of the classification model, the upper right panel shows the average predicted probabilities of presidential and gubernatorial debates of being of national class over time, and the lower right panel shows the predicted class counts for just gubernatorial debates over time.

Figure 2: C–SPAN Debate Predictions
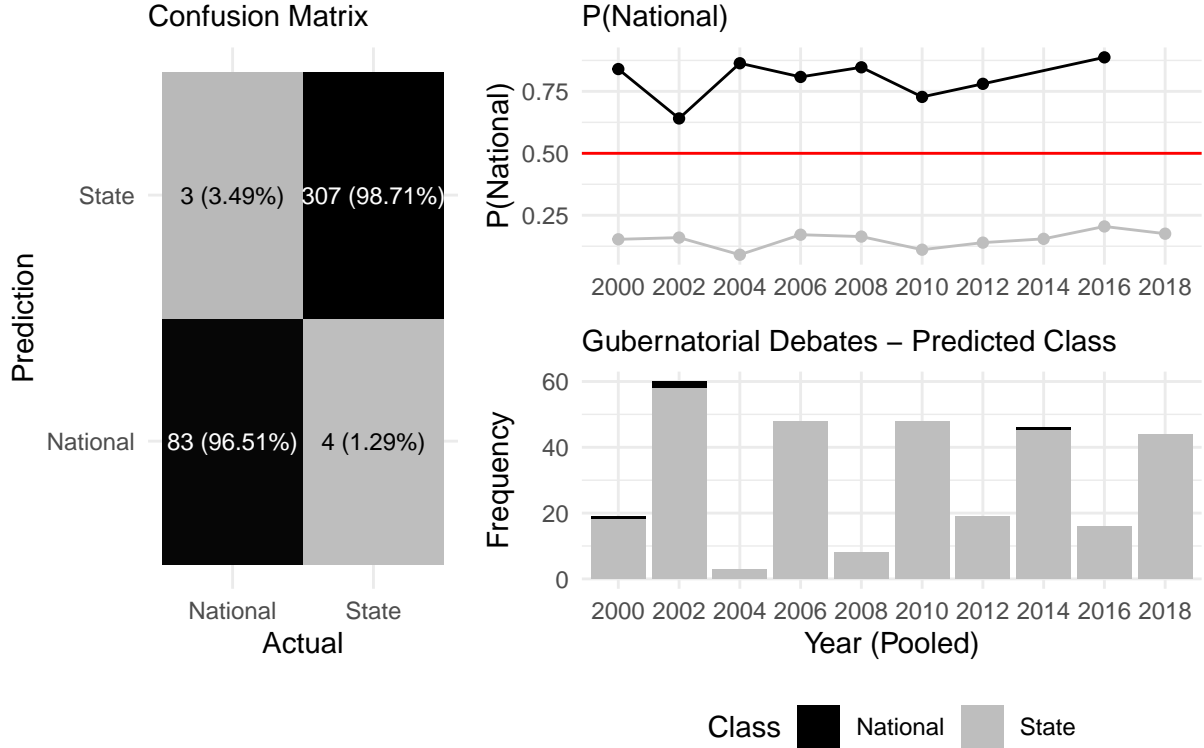
These results give a consistent picture of rhetoric in debates; candidates predominately discuss topics germane to their jurisdictions. In purely statistical terms, collections of words more indicative of state (national) content are significantly more common in gubernatorial (presidential) debates, and the prevalence of those topics has not changed significantly over time.

It is important to note that these results, particularly the predicted probability of national origin, are not meant to indicate the exact proportion of content in a debate that reflects state or national topics. The debates themselves *are* quantitatively represented as different proportions of topics, but the weights with which those topic proportions are translated into predictions of class are not uniform. An alternative method of quantitatively representing texts may break debates into more granular pieces (such as sentences), classify each individual sentence as either state or national, then tally the number of sentences in each predicted class. The advantage of the method I use in this paper is the overall sense of content for each document; in general, are the topics discussed in the document more consistent with documents of known origin?

For example, consider the 2002 New Mexican gubernatorial debate between Republican John Sanchez, Democrat Bill Richardson, and Green Party candidate David Bacon. The classification model gave this debate a predicted national probability of 16.9%, but this of course masks the full heterogeneity of topics covered during the debate, which touched on national topics such as NAFTA and the war in Iraq. Figure

3 gives a more detailed view of the debate as a treemap of estimated topic proportions, with the size of each tile representing the size of the proportion and the color representing the model estimate of its relative "state-ness" or "national-ness" (and coefficients that weren't significant predictors of either). Here, topics that lean "national" make up a larger proportion of the total debate than the 16.9% total suggested by the predicted probability, but the predictive weight of the state topics lowers shifts the prediction to the state side.

Figure 3: Topic Proportions of 2002 New Mexico Gubernatorial Debate
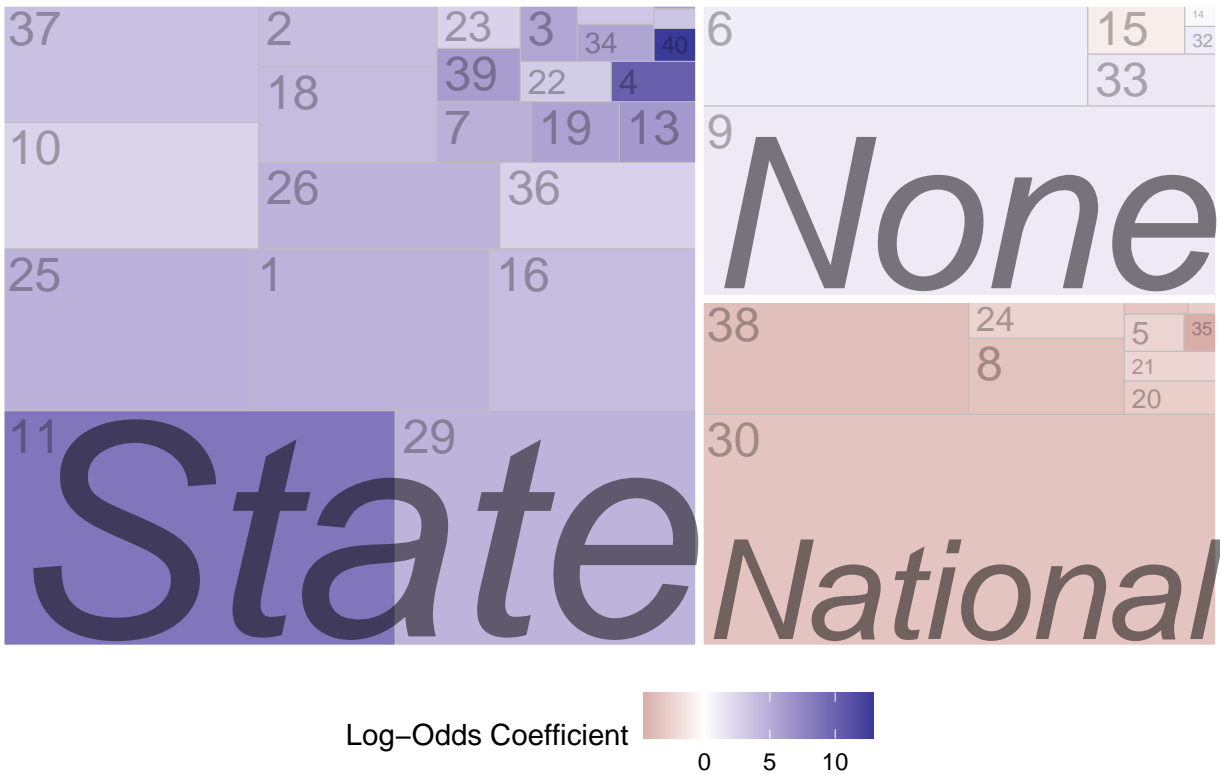
TV Advertisements

Next, I consider the potential nationalization of rhetoric in televised campaign advertisements. This medium is perhaps the modal form of campaigning in the eyes of constituents and the most commonly studied campaign messaging medium in political science. It does, however, present some unique challenges to the classification methodology utilized in this paper.

Most obviously, the quasi built-in controls for candidate ideology and campaign resources in the debate context are absent from televised ads. Candidates with larger war chests may be able to air more ads referencing a broader range of topics, whereas more cash-restricted candidates could be forced to focus their message around just a few talking points. A related problem is the unique content of ads overall; they are

much shorter than debates, tend to be more negative, and, while they still speak predominantly about policy, they do often incorporate references to general candidate character or background. Furthermore, candidates increasingly have the ability to target advertisements to particular audiences for particular purposes. Certain messages may be broadcast to swing voters as persuasive content, while other messages may be broadcast to candidates' bases to turn out the vote. This is all to say that the content of advertisements is likely substantially different from the content of debates, which has consequences for the potential for nationalized campaigning strategies.

For this paper, I analyze 1,223 televised advertisements from presidential (907 ads) and gubernatorial (316) campaigns in 2008[3]. These ads are provided by the Wisconsin media project, with transcripts scraped from PDF storyboards. These ads include those run in both the primaries and general elections and by both candidates and interest groups, with the gubernatorial ads covering 11 states (DE, IN, MO, MT, NC, ND, NH, UT, VT, WA, and WV).
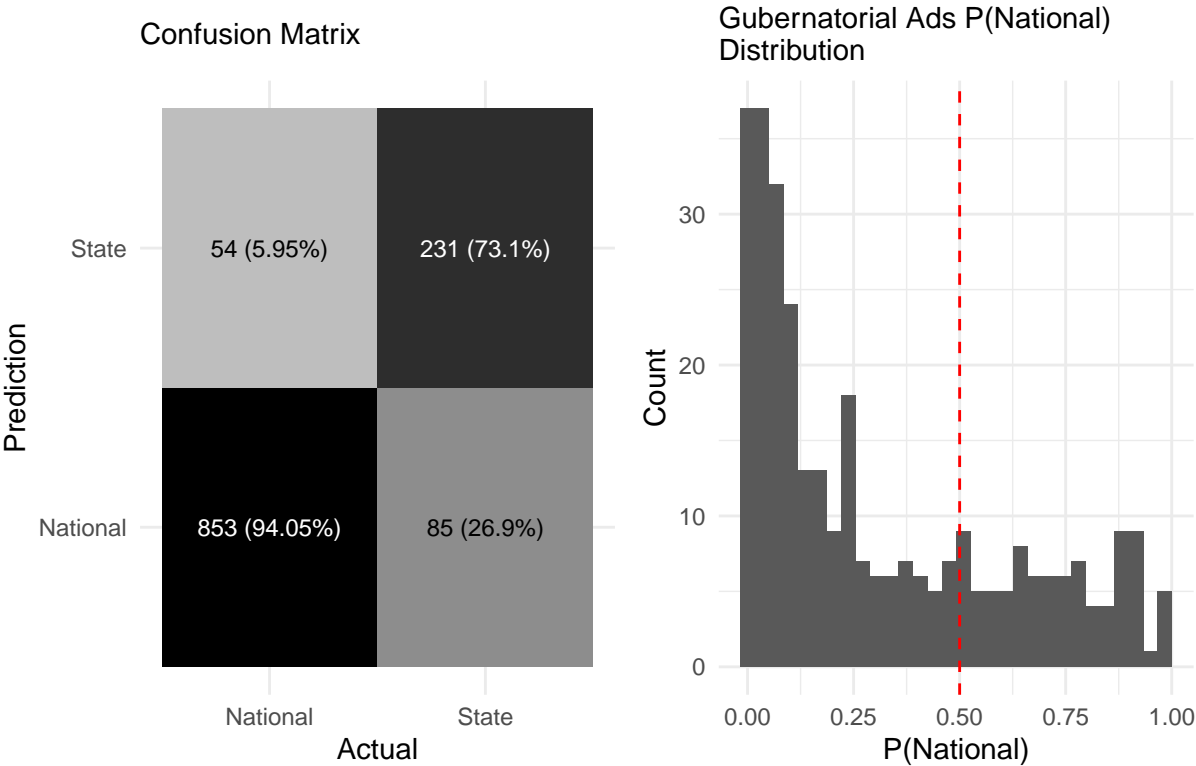
Figure 4: Ads Predictions

Figure 4 shows the results of applying the trained classification model to these televised advertisements. While advertising content still predominately consists of topics germane to candidates' jurisdictions, there is a higher proportion (26.9%) of gubernatorial ads that are classified as being national. The left panel

---

[3]I plan to expand this sample in the future utilizing automated audio-to-text transcription.

shows a similar classification rate of presidential rhetoric as in the debate context, which indicates the televised ad medium is not necessarily biased toward more state-like content. The right panel of figure XX shows the distribution of the predicted probabilities of national classification for just the gubernatorial ads. Unsurprisingly, most of the predictions are strongly in the state directly, but a large number are classified as very national in content.

To give an example of one such gubernatorial ad, the Alliance for North Carolina ran an attack ad on Pat McCrory in October 2008 that the classification model assigned an 87% probability of being national in content. The brief transcript reads:

> The big developers, energy companies, and the banking industry just love Pat McCrory and George Bush. Why? Because McCrory and Bush have the same economic philosophy. Less regulation and less oversight to help these companies make even more profit. The result, economic collapse and a Wall Street Bailout. Who ends up paying? You the middle-class. Pat McCrory, stop supporting Bush economics and start supporting more regulation and oversight of big business.

This ad clearly attempts to link McCrory to Bush policies with fairly little state-specific information, instead using terms that would be equally applicable in any other state ("Wall Street bailout" and "middle-class"). While this is a compelling example, a majority of advertisements are still classified as being predominantly state content.

Twitter

Finally, I analyze a more modern form of campaign rhetoric; social media. Specifically, I analyze the Twitter of Members of Congress and Governors in office during 2018 using the Das et al. (2021) corpus of tweets. This corpus contains 952,425 tweets from sitting Members of Congress and 101,546 tweets from incumbent governors.

This corpus is unique in this paper for many reasons. First, the tweets are not specific to the campaign timeframe, and therefore don't explicitly count as "campaigning." Second, the tweets only account for incumbents and do not include the tweets of their challengers. Third, the "national" comparison in this context is Members of Congress, not communication from Presidents. This is important because Members of Congress operate at the national stage while being beholden to district even more localized than their gubernatorial counterparts, so we might expect communication to be split between national and state topics. Finally, and perhaps most importantly, Twitter represents a fundamentally different avenue through which politicians communicate with supporters. Twitter can and is used to campaign, but can also be used for

ostensibly non-political activity, like cheering on a local basketball team or engaging in more general political hobbyism. Outside of terms of service violations, there are really no restrictions on what can or can't be said on Twitter by politicians.

Figure 5: Twitter Predictions

Confusion Matrix

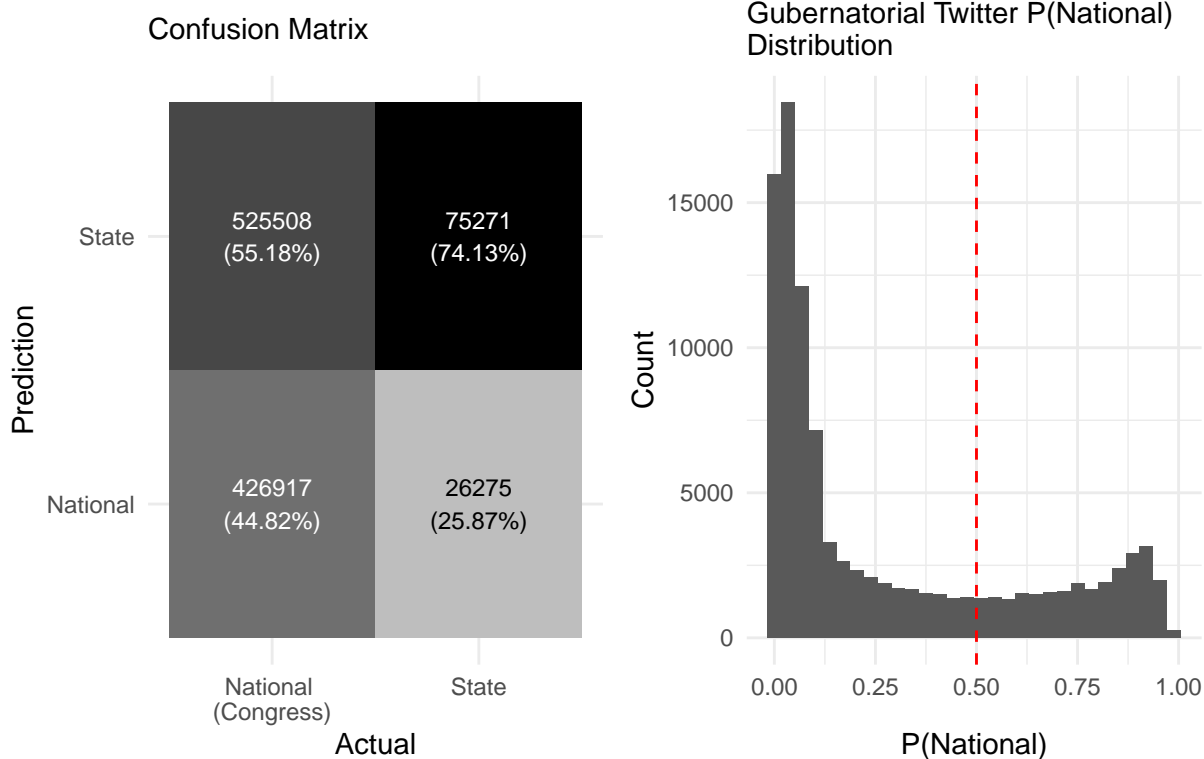Gubernatorial Twitter P(National) Distribution

Figure 5 shows the results of the Twitter analysis. Similar to rhetoric in televised ads, governors still communicate more on topics related to state politics, but engage significantly in national political topics as well. Members of Congress are more evenly split between national and state topics, which is likely a function of their accountability to district-level pressures. The right panel of Figure 5 shows the distribution of predicted probabilities for national classification of gubernatorial tweets. Most tweets show strong "state-ness," but there does seem to be a slight bimodal distribution, with a large portion of tweets being classified as heavily national.

Discussion and Conclusion

The preceding results indicate an overall picture of gubernatorial campaign rhetoric that gives pause to the "all politics is national" hypothesis. Across all mediums, a comfortable majority of communications were classified as primarily consisting of state topics. This approached almost 100% in debates, but closer to 75% in televised advertisements and social media posts on Twitter. The lower state classification rates in the

later to mediums suggest gubernatorial candidates do engage in some degree of nationalization when they have the flexibility to do so. The debate context is fairly constrained, so when those constraints are lifted and the field of possible topics expands beyond topics presented by a debate moderator, we would expect to see somewhat higher rates of nationalization.

While this paper has focused largely on the classification of single pieces of communication in isolation, it is possible that the real engine of information nationalization is the media environment reporting on, circulating, and commenting on the communications. Media plays a major role in how voters engage with campaign materials. It is possible that while most campaign messaging from gubernatorial candidates focuses on state topics, the few communications that *are* nationalized are circulated more widely by the media. During the gubernatorial race in Kentucky in 2019, for example, coverage from national outlets like the New York Times largely focused on the Republican incumbent's (Matt Bevin) allegiance with and affinity for Donald Trump. In the debate between the Bevin and his Democratic challenger Andy Beshear, the topic of impeachment *did* arise, but it was constrained to a single question. For the most part, the rest of the debate revolved around issues germane to Kentucky politics.

Further work must be done to both extend the corpus of text analyzed with the methodological approach of this paper and determine if there exists a relationship between nationalized rhetoric and nationalized results. In future work, I will extend the corpus of advertising text to include all advertisements from 2000-2018, courtesy of the Wesleyan and Wisconsin Media Projects. I will also extend the analysis to model the nationalization of election results as a function of the degree of campaign nationalization. Finally, I hope to better understand that "downstream" nationalization that may occur as a result of what topics media outlets choose to cover. Does the coverage of campaign activity skew more national than the campaign activity itself?

While preliminary, the results of this paper have consequences for how we understand voter interaction with campaigns. These campaign activities *do* offer a source of information to voters that is functionally distinct from national politics. How voters then process this information is of subsequent importance to better understand the information environment voters must navigate within a nationalized context.

References

Abernathy, Penelope Muse. 2018. *The Expanding News Desert.* Chapel Hill, NC: Center for Innovation; Sustainability in Local Media, University of North Carolina at Chapel Hill. https://www.cislm.org/wp-content/uploads/2018/10/The-Expanding-News-Desert-10_14-Web.pdf.

Abramowitz, Alan I., and Steven Webster. 2016. "The Rise of Negative Partisanship and the Nationalization of U.S. Elections in the 21st Century." *Electoral Studies* 41: 12–22. https://linkinghub.elsevier.com/retrieve/pii/S0261379415001857 (October 11, 2019).

Arceneaux, Kevin. 2006. "The Federal Face of Voting: Are Elected Officials Held Accountable for the Functions Relevant to Their Office?" *Political Psychology* 27(5): 731–54. http://doi.wiley.com/10.1111/j.1467-9221.2006.00530.x (October 10, 2019).

Beer, Samuel H. 1978. "Federalism, Nationalism, and Democracy in America." *American Political Science Review* 72(1): 9–21. https://www.cambridge.org/core/product/identifier/S0003055400154139/type/journal_article (October 11, 2019).

Benedictis-Kessner, Justin De, and Christopher Warshaw. "Accountability for the Local Economy at All Levels of Government in United States Elections." *American Political Science Review* 114(3): 660–76. https://www.cambridge.org/core/journals/american-political-science-review/article/accountability-for-the-local-economy-at-all-levels-of-government-in-united-states-elections/6B61FD69CF47F0DA6E55341A5FF12454 (March 23, 2020).

Benjamin, Daniel J, and Jesse M Shapiro. 2009. "Thin-Slice Forecasts of Gubernatorial Elections." *The Review of Economics and Statistics* 91(3): 523–36. https://doi.org/10.1162/rest.91.3.523 (January 9, 2020).

Benoit, William L., LeAnn M. Brazeal, and David Airne. 2007. "A Functional Analysis of Televised U.S. Senate and Gubernatorial Campaign Debates." *Argumentation and Advocacy* 44(2): 75–89. https://doi.org/10.1080/00028533.2007.11821679 (January 9, 2020).

Benoit, William L., Glenn J. Hansen, and Rebecca M. Verser. 2003. "A Meta-Analysis of the Effects of Viewing U.S. Presidential Debates." *Communication Monographs* 70(4): 335–50. https://nca.tandfonline.com/doi/abs/10.1080/0363775032000179133 (January 9, 2020).

Brown, Adam R. 2010. "Are Governors Responsible for the State Economy? Partisanship, Blame, and Divided Federalism." *The Journal of Politics* 72(3): 605–15. https://www.journals.uchicago.edu/doi/abs/10.1017/S0022381610000046 (March 23, 2020).

Brunner, Jim. 2016. "Gov. Jay Inslee Defeats Republican Challenger Bill Bryant." *The Seattle Times.* https://www.seattletimes.com/seattle-news/politics/washington-governor-jay-inslee-bill-bryant/.

Catanese, David. 2011. "Tomblin Wins W.Va. Gov Race." *Politico.* https://www.politico.com/story/2011/10/tomblin-wins-wva-gov-race-065172.

Caughey, Devin, James Dunham, and Christopher Warshaw. 2018. "The Ideological Nationalization of Partisan Subconstituencies in the American States." *Public Choice* 176(1-2): 133–51. http://link.springer.com/10.1007/s11127-018-0543-3 (October 10, 2019).

Caughey, Devin, and Christopher Warshaw. 2016. "The Dynamics of State Policy Liberalism, 1936–2014." *American Journal of Political Science* 60(4): 899–913. https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12219 (September 10, 2021).

Costa, Mia. 2021. "Ideology, Not Affect: What Americans Want from Political Representation." *American Journal of Political Science* 65(2): 342–58. https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12571 (September 2, 2021).

Das, Sanmay, Betsy Sinclair, Steven W. Webster, and Hao Yan. 2021. "All (Mayoral) Politics Is Local?" *The Journal of Politics.* https://www.journals.uchicago.edu/doi/abs/10.1086/716945 (September 5, 2021).

Hayes, Danny, and Jennifer L. Lawless. 2018. "The Decline of Local News and Its Effects: New Evidence from Longitudinal Data." *The Journal of Politics* 80(1): 332–36. https://www.journals.uchicago.edu/doi/10.1086/694105 (October 11, 2019).

Hillygus, D. Sunshine, and Simon Jackman. 2003. "Voter Decision Making in Election 2000: Campaign Effects, Partisan Activation, and the Clinton Legacy." *American Journal of Political Science* 47(4): 583–96. https://onlinelibrary.wiley.com/doi/abs/10.1111/1540-5907.00041 (January 9, 2020).

Hopkins, Daniel J. 2018. *The Increasingly United States: How and Why American Political Behavior Na-*

tionalized. Chicago: The University of Chicago Press.

Huddy, Leonie, and Alexa Bankert. 2017. "Political Partisanship as a Social Identity." *Oxford Research Encyclopedia of Politics*. https://oxfordre.com/view/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-250 (March 21, 2020).

Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes. 2012. "Affect, Not Ideology: A Social Identity Perspective on Polarization." *Public Opinion Quarterly* 76(3): 405–31. https://doi.org/10.1093/poq/nfs038 (May 9, 2021).

Jacobson, Gary C. 2015. "It's Nothing Personal: The Decline of the Incumbency Advantage in US House Elections." *The Journal of Politics* 77(3): 861–73. https://www.journals.uchicago.edu/doi/10.1086/681670 (September 9, 2021).

Kousser, Thad. 2014. "How America's 'Devolution Revolution' Reshaped Its Federalism." *Revue française de science politique* 64(2): 265. http://www.cairn.info/revue-francaise-de-science-politique-2014-2-page-265.htm (October 11, 2019).

Manchester, Julia. 2019. "GOP Seeks to Nationalize Gubernatorial Elections." *The Hill*. https://thehill.com/homenews/campaign/468911-gop-seeks-to-nationalize-gubernatorial-elections.

Martin, Gregory J., and Joshua McCrain. 2019. "Local News and National Politics." *American Political Science Review* 113(2): 372–84. https://www.cambridge.org/core/journals/american-political-science-review/article/local-news-and-national-politics/C8EEA488A777C37C7987964F8F85AEB5 (March 18, 2020).

Martin, Jonathan. 2019. "Kentucky Governor's Race Tests Impact of Impeachment in States." *The New York Times*. https://www.nytimes.com/2019/10/29/us/politics/matt-bevin-andy-beshear-kentucky-governor.html (December 16, 2019).

McKinley Jr., James C. 2010. "Perry Re-Elected in Texas Governor Race." *The New York Times*. https://www.nytimes.com/2010/11/03/us/politics/03texas.html (March 20, 2020).

Moskowitz, Daniel J. 2021. "Local News, Information, and the Nationalization of U.S. Elections." *American Political Science Review* 115(1): 114–29. https://www.cambridge.org/core/journals/american-political-science-review/article/abs/local-news-information-and-the-nationalization-of-us-elections/4AEEA64CB7EC2CF384434AB0482E63F4 (May 9, 2021).

Mummolo, Jonathan, Erik Peterson, and Sean Westwood. 2021. "The Limits of Partisan Loyalty." *Political Behavior* 43(3): 949–72. https://doi.org/10.1007/s11109-019-09576-3 (September 10, 2021).

Roberts, Margaret E., Brandon M. Stewart, and Edoardo M. Airoldi. 2016. "A Model of Text for Experimentation in the Social Sciences." *Journal of the American Statistical Association* 111(515): 988–1003. https://doi.org/10.1080/01621459.2016.1141684 (November 12, 2019).

Rojas, Rick, and Jeremy Alford. 2019. "In Louisiana, a Narrow Win for John Bel Edwards and a Hard Loss for Trump." *The New York Times*. https://www.nytimes.com/2019/11/16/us/louisiana-governor-edwards-rispone.html (March 20, 2020).

Shor, Boris, and Nolan McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105(3): 530–51. https://www.cambridge.org/core/journals/american-political-science-review/article/ideological-mapping-of-american-legislatures/8E1192C22AA0B9F9B56167998A41CAB0 (September 10, 2021).

Sievert, Joel, and Seth C. McKee. 2019. "Nationalization in U.S. Senate and Gubernatorial Elections." *American Politics Research* 47(5): 1055–80. https://doi.org/10.1177/1532673X18792694 (December 16, 2019).

Webster, Steven W. 2020. *American Rage: How Anger Shapes Our Politics*. Cambridge: Cambridge University Press. https://www.cambridge.org/core/books/american-rage/EE2BAF959B773D58529F148782E58CF2 (September 9, 2021).

Weinschenk, Aaron et al. 2020. "Have State Supreme Court Elections Nationalized?" *Justice System Journal* 41(4): 313–22. https://doi.org/10.1080/0098261X.2020.1768187 (September 2, 2021).

Appendices

A1: K Selection

Figure A1.1 Model Diagnostics by Number of Topics
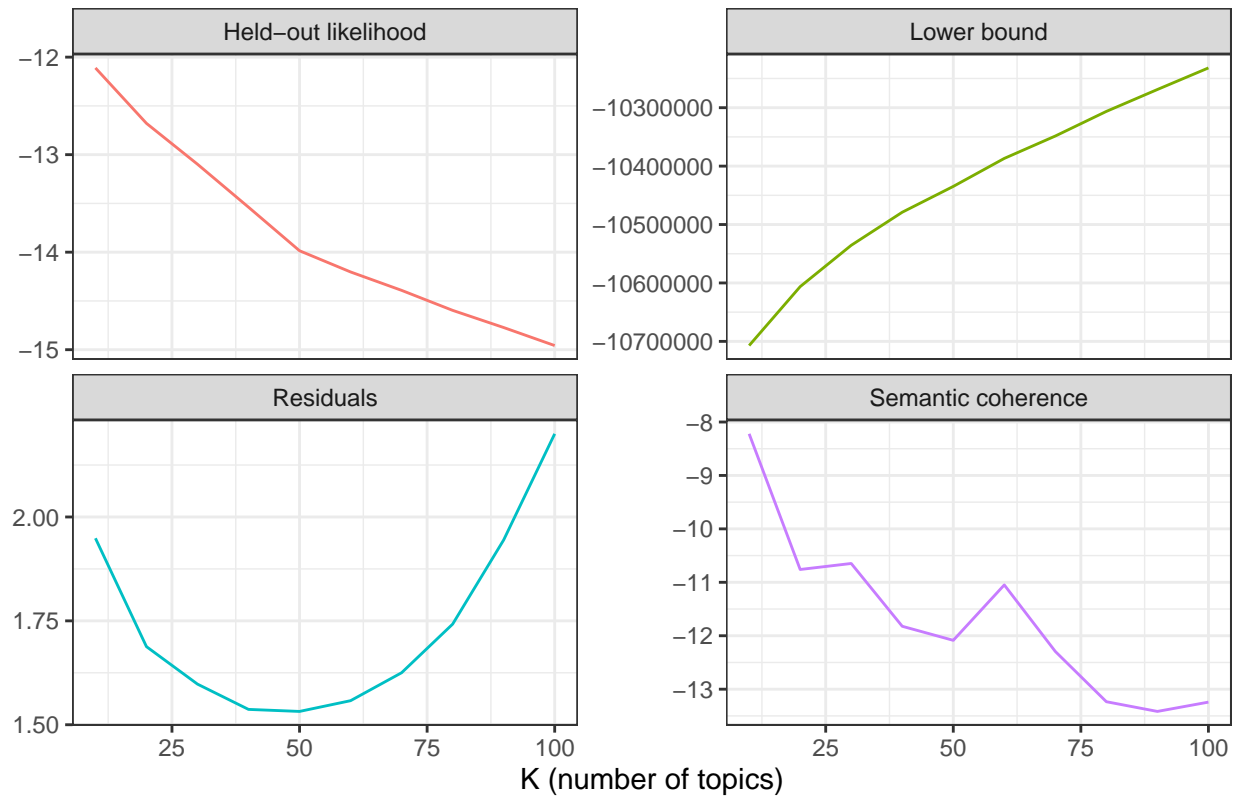
A2 Testing Data Results (All Models)

Table A2.1: Classification Model Performance on C-SPAN Debates

	National Documents			State Documents		
Model	Correct	Incorrect	Accuracy	Correct	Incorrect	Accuracy
Logistic Regression (Nonpenalized)	83	3	0.965	307	4	0.987
Naive Bayes	83	3	0.965	233	78	0.749
Penalized Logistic Regression (Lasso)	82	4	0.953	307	4	0.987
Boosted Gradient Descent (XGBoost)	64	22	0.744	310	1	0.997
Support Vector Machine	82	4	0.953	310	1	0.997

Table A2.2: Classification Model Performance on TV Ads

	National Documents			State Documents		
Model	Correct	Incorrect	Accuracy	Correct	Incorrect	Accuracy
Logistic Regression (Nonpenalized)	853	54	0.940	231	85	0.731
Naive Bayes	817	90	0.901	237	79	0.750
Penalized Logistic Regression (Lasso)	821	86	0.905	231	85	0.731
Boosted Gradient Descent (XGBoost)	524	383	0.578	310	6	0.981
Support Vector Machine	840	67	0.926	240	76	0.759

Table A2.3: Classification Model Performance on Twitter

	National Documents			State Documents		
Model	Correct	Incorrect	Accuracy	Correct	Incorrect	Accuracy
Logistic Regression (Nonpenalized)	426917	525508	0.448	75271	26275	0.741
Naive Bayes	255429	696996	0.268	85349	16197	0.840
Penalized Logistic Regression (Lasso)	353273	599152	0.371	81551	19995	0.803
Boosted Gradient Descent (XGBoost)	27933	924492	0.029	100550	996	0.990
Support Vector Machine	433479	518946	0.455	74479	27067	0.733

A3: Topics

Below are the top words for the topics most highly associated with state (40,4,11) and national (8,38,35) content. FREX words are weighted by both frequency and topic exclusivity, lift word weights are divided by the frequency of the words in other topics, and score words are weighted by dividing their intra-topic log frequency by their inter-topic log frequency.

```
## Topic 40 Top Words:
##      Highest Prob: million, budget, increase, program, education, fund, funding
##      FREX: recommend, therefore, transportation, additional, funding, session, trooper
##      Lift: dante, ruvo, squires, georgialearns.com, godwin, fury, sparks
##      Score: tonight, session, vegas, budget, trooper, recommend, 12
## Topic 4 Top Words:
##      Highest Prob: property, reform, taxis, school, work, good, government
##      FREX: property, town, zone, empire, urban, relief, taxis
##      Lift: poughkeepsie, yorkstate, wicks, binghamton, bruno, garrahy, nanoelectronics
##      Score: buffalo, property, manhattan, cuny, bruno, empire, syracuse
## Topic 11 Top Words:
##      Highest Prob: government, people, work, business, time, every, well
##      FREX: south, cabinet, taxpayer, government, greitens, honest, elect
##      Lift: caryn, messenger, proviso, rabjohn, fuse, gifts, greitens
##      Score: greitens, tonight, south, packers, alabamian, hitt, rena
## Topic 8 Top Words:
##      Highest Prob: federal, government, right, support, program, family, american
##      FREX: abortion, religious, marriage, amendment, affirm, oppose, constitution
##      Lift: americansthe, commandments, consensual, energetically, enfranchisement, euthanasia, inval
##      Score: abortion, abstinence, sonia, sotomayor, religious, alien, marriage
## Topic 38 Top Words:
##      Highest Prob: support, community, energy, democrats, american, americans, work
##      FREX: democrats, indian, puerto, rico, obama, discrimination, tribal
##      Lift: celsius, environmentalism, outsize, proselytize, servicethe, ricans, cfpb
##      Score: puerto, rico, guam, gore, islands, lgbt, democrats
## Topic 35 Top Words:
##      Highest Prob: work, president, americans, america, family, health, american
##      FREX: americans, republicans, medicare, democrats, middle, class, president
##      Lift: 1835, caregive, cherrey, quayle, budge, gephardt, warrensburg
##      Score: medicare, americans, gore, mitt, romney, democrats, republicans
```