Measuring the Issue Content of Supreme Court Opinions*

Abstract

The opinions of the U.S. Supreme Court are central to volumes of research on law, courts, and politics. To understand these complex and oft-lengthy documents, scholars frequently rely on dichotomous indicators of opinion content. While sometimes appropriate, for many research settings this simplification of opinion content systematically omits important information. Using all U.S. Supreme Court opinions from 1803 to 2010 in association with structural topic models, I instead demonstrate the value of representing the Court's attention in opinions in terms of topic proportions.

---

*All materials necessary to replicate the results presented in this article will be made available on Dataverse upon publication.

Perched atop the judiciary, the U.S. Supreme Court regularly publishes opinions which provide critical information on law and public policy. Understanding the information contained within opinions has therefore served as a principal objective for social scientists and legal scholars for generations. As is frequently recognized, the identification of this dimension is critical to our understanding of the Court, including for example debate over strategic judicial behavior (Benesh and Spaeth 2007; Friedman 2006) and the Court's policymaking role (Baird 2004, 2007; Peters 2007).

Yet identifying and measuring the content of a judicial opinion has proven remarkably controversial, as parsing voluminous and verbose documents poses a particular challenge for the researcher. At the heart of this challenge is what I term the *dimensionality* problem. Even within a case the Court is often called to at least consider if not resolve multiple issues. Simply at a minimum, questions over justiciability are addressed in all cases reaching the Court in addition to whatever specific controversy is at issue. Take, for example, the landmark abortion decision *Roe v. Wade*; beyond establishing a trimester framework for evaluating the constitutionality of abortion restrictions, the opinion addressed standing and justiciability questions at length. For the researcher, incorporating each additional dimension may yield a better and more accurate understanding of the Court's attention but it correspondingly complicates the classification task. Moreover, even if one identified the presence of multiple dimensions, the measurement would still obscure the relative attention the Court paid to each of those dimensions. Given the challenges of identifying multiple dimensions, and the relative impossibility of identifying attention proportionally through standard human-coding approaches, scholars regularly rely on single membership classification approaches, relying on measures that identify only a single, primary dimension.

For this, scholars almost exclusively rely on the high quality issue coding available from the Supreme Court Database [SCD].[1] Explicitly intended to measure the policy content

[1]Indeed, Epstein, Knight and Martin (2003) – in their review of research on judicial behavior – identify the SCD as "so dominating in our discipline that it would certainly be

(or "substantive basis") of the opinion, the coding protocol places a strong preference on assigning only a single issue for each case. In so doing, the codes reduce lengthy treatises to – in the overwhelming majority of cases – a single issue dimension which was the basis of the Court's decision. Given their high quality, accessibility, and widespread adoption, the codes have proven valuable in enhancing the comparability of a long history of research.

Yet the cost of the simplification is the potential loss of information. The preference for assigning a single membership classification potentially obscures large amounts of important information for scholars interested in the Court. While the loss may be acceptable for some research applications, it limits the opportunities for other research to fully understand the Court's attention and judicial behavior. Alternatively, the classification of issues – or topics – is precisely the sort of task for which computational approaches to text analysis are ideally suited. In this article, I demonstrate the utility of one such computational method – structural topic models (Roberts et al. 2014) – as a flexible measurement approach offering estimates of topic attention across multiple dimensions.[2] I estimate the Court's topical attention across a corpus of all Supreme Court majority opinions between 1803 and 2010. Through a comparison of the estimated topic proportions to the SCD issue variables both for specific cases and as measures of changes in the Court's agenda, I document the validity of the estimated topics. Then, I offer a series of illustrations for areas where the estimated topic proportions offer an opportunity to address substantively important questions for which dichotomous indicators are potentially insufficient. Finally, I conclude by discussing how the flexibility of the approach offers a wealth of opportunities for extensions in subsequent research.

---

unusual for a refereed journal to publish a manuscript whose data derived from an alternative source" (812).

[2]On publication, a series of pre-estimated models of opinion content across different levels of specificity will be made available on the author's Dataverse page.

Measuring Opinion Content

Identifying the issues addressed by a given court case is critically important for research on law and courts. Often, scholars restrict their analyses to particular areas of the law in order to better identify a specific effect, say of case facts or precedent, on judicial behavior (e.g., Segal 1984, 1986; Richards and Kritzer 2002; Richards, Smith and Kritzer 2006). In yet other work, scholars frequently rely on model specifications which include dichotomous covariates (fixed effects) for issue areas, seeking to address issue-specific heterogeneity. Likewise, specifying the issue at debate serves as the principal means to identify the Court's ideological output (e.g., McGuire et al. 2009) and to identify dimensions of dispute in the case (e.g., Benesh and Spaeth 2007). Even further, the identification of issues has implications for the rhetorical strategies of the justices, including whether they engage in heresthetical maneuvering (e.g., Rice 2016) or issue creation (e.g., McAtee and McGuire 2007). Moving beyond judicial behavior, measuring issue content also has important implications for cross-institutional studies, including understanding the judicial hierarchy (e.g., Baird 2004, 2007; Peters 2007), legal mobilization (e.g., Rice 2014), and the influence of the courts in the policy process (e.g., Flemming, Bohte and Wood 1997; Flemming, Wood and Bohte 1999). In all, an understanding of the issues addressed by the Court are implicated in nearly all areas of research on the Court.

To that end, the identification of the issue continues to generate discussion and disagreement among top scholars in the field, including one recent exchange on the public discussion list of the Law and Courts Section of the American Political Science Association.[3] In the first place, there is disagreement as to the very definition of issue. The Supreme Court Database, for instance, classifies based on public policy, or the "subject matter of the controversy." Even within this definition, however, there are conceptual difficulties, as categories like "Judicial Power" are generally not perceived to be public policy issues.[4] In another project,

[3]Law & Courts List-Serv Digest, August 15-16, 2016.

[4]Note, for instance, that the Policy Agendas Project does not include any references to

Shapiro (2009) classifies instead *legal* issues which transition away from a public policy focus and are instead explained as: "[i]magin[ing] that you are trying to describe the case to a first year law student. You want the student to understand what the case is about and what areas of law it implicates" (532). Here, I build on these definitions and consider issues or topics thematically. That is, I consider issues as the themes which characterize a corpus of text. This definition has the benefit of making plain that which is implicit in both of the above mentioned definitions: there are multiple potential nestings and hierarchies of topics, such that any division ultimately rests on the substantive interests of the scholars employing the measures. Moreover, this definition does not require differentiating *policy* from *legal* issues. Instead, both can be captured at least in so far as they represent important variation in themes across the corpus.

The primary difficulty of the manual coding enterprise, though, lies in the *dimensionality* problem. Court opinions are lengthy, complex documents, and therefore regularly address multiple issues. Moreover, emphasizing or prioritizing one of those issues over the others may often be a key component of the Court's decision making process (Epstein and Shvetsova 2002; Wedeking 2010; Black, Schutte and Johnson 2013; Rice 2016). Yet identifying each issue vastly complicates the coding task, requiring detailed analysis of the entire opinion. Moreover, there is no clear stopping rule for the coder; that is, what constitutes enough attention? Is a footnote adequate? Both of these problems are then magnified in that this dichotomous treatment would equivalently treat issues which are discussed in brief to those issues which formed the primary basis of the Court's decision. Finally, the proliferation of these dimensions increases the probability of discrepancies in coding decisions, as the reliability of measures decreases as the complexity of the task increases.

To overcome these challenges, researchers reasonably rely on publicly available issue measures from the SCD, which provides human-coded public policy and legal issue codes. For each case, the classification scheme addresses the dimensionality problem by explicitly

---

judicial power in their topic or subtopic categories.

identifying the single most important dimension in the Court's decision. That is, the SCD treats opinions as a single membership classification, therefore every one of the Court's opinions are therefore classified into *one* of 260 unique codes further organized into one of 14 broad categories called issue areas.[5] These codes are quite explicitly intended to reflect the public policy content of the opinion; as the coding protocol states, "[t]his variable identifies issues on the basis of the Court's own statements as to what the case is about. The objective is to categorize the case from a public policy standpoint" (42). Thus, for nearly every opinion of the Court, the issue codes identify a single policy which the Court (justice) addressed in that opinion.

In relying on these issue codes, researchers reasonably resolve three of the major challenges for identifying issue content. First, the public availability of high-quality, consistent, and reliable coding of issue attention is available for free and for all Supreme Court decisions. Thus, researchers are able to mitigate the massive resource costs necessary to identify these codes themselves. Second, by identifying a single, primary dimension for the decision, researchers explicitly avoid the equivalency problem. The coding protocol specifically advocates identifying the primary dimension on which the decision was made, and thus all subordinate issues – though potentially – are omitted. Finally, the prevalence of the codes – and the prevalence of the SCD more generally – enhances the comparability of research. That is, while two separate researchers may come to different conclusions as to the primary issue of a case, relying on a single estimate across different research settings enhances our ability to compare identified relationships.

Yet the approach carries an important cost. By relying on single membership classifications, the measure necessarily simplifies complex legal documents. Importantly, in many cases such an approach likely leads to a large-scale deletion of potentially important infor-

[5]On rare occasions, the SCD does code multiple topics for a single case. However, these instances are exceedingly rare, and only arise where the coder can not identify a single *primary* dimension.

mation. Were cases generally about a single issue, the deletion may be ignorable. But the presence of multiple issues in a case is well-recognized, as minimally the Court must address justiciability concerns in addition to whatever controversy it is being asked to decide. Consider the SCD's classification of *Wiggins v. People in Utah* (1876), which is identified as disposed on the grounds of Judicial Power, specifically jurisdiction. The entirety of the majority opinion's discussion of jurisdiction is as follows:

> Sec. 3 of the Act of Congress of June 23, 1874, 18 Stat. 254, allows a writ of error from this Court to the Supreme Court of the Territory of Utah, where the defendant has been convicted of bigamy or polygamy or has been sentenced to death for any crime. The present writ is brought under that statute to obtain a review of a sentence of death against plaintiff in error for the murder of John Kramer, commonly called Dutch John, in Salt Lake City.[6]

The remaining six pages of the majority opinion, and the entirety of Justice Clifford's dissenting opinion, focused on a description of the homicide at the heart of the case and the admissibility of witness testimony to the jury. Thus, though the Court certainly considered and was required to address the issue of jurisdiction – as they must in any case – the overwhelming focus of *Wiggins* was on a separate issue that the Court also addressed.

To understand the scale of the potential deletion, Shapiro (2009) revisited a random sample of 95 opinions and identified each of the issues addressed in the opinions. In this context, Shapiro utilized the SCD's coding protocol but *did not* give preference to assigning only a single issue. Of the 95 cases recoded, Shapiro identified 89 – or approximately 94% of the opinions – as addressing more than one issue. Moreover, for many of those cases Shapiro identified more than two issues in the case.. In contrast, only one of those 89 opinions was classified as addressing multiple issues by the SCD.

Considering the implications of deleting information for statistical inference (see, e.g., Gelman and Hill 2007), the implications of the *unidimensional assumption* are potentially profound for some research settings. Particularly as researchers seek to better identify and understand judicial behavior – to "take law seriously" (Friedman 2006) – generating and

[6] 93 U.S. at 466.

utilizing appropriate measures of opinion content are essential. In the remainder of this article, I turn to this task.

Topic Models for Measuring Opinion Content

I do so by employing tools from a well-developed area of computational social science. The potential utility of computational approaches to studies of the Supreme Court and other judicial institutions is well-demonstrated by the volume of research in the area. In the first place, researchers have looked to improve variables already included in the SCD (e.g., McGuire and Vanberg 2005; Evans et al. 2007). For example, McGuire and Vanberg (2005) use Wordscores (Laver, Benoit and Garry 2003) to measure the ideological direction of Supreme Court opinions, an approach built upon by Lauderdale and Clark (2014) who jointly model votes and issues to extract ideological dimensions within issue areas. In addition to refining existing measures, scholars have proposed expanding the coding scheme to other judicial venues (Szmer and Edwards 2011).

In this vein, the unsupervised estimation of topical content from text offers an avenue to resolve the underlying problems of the dimensionality problem in a flexible and cost-efficient way. Topic models are a suite of methods for organizing and understanding collections of documents. In this vein, topics can be understood as clusters of co-occurring words. This clustering of co-occurring words fits well with the definition of topics as themes which define the corpus. Within political science alone, increasingly sophisticated research has been done on topic models to perform unsupervised topic classification (Quinn et al. 2010; Grimmer 2010) with recent efforts at creating more flexible topic modeling forms which can adopt additional information (Roberts, Stewart and Airoldi 2016). Models uncover patterns of word use across documents and connect those documents which exhibit similar patterns (Blei and Lafferty 2009). Note that, unlike human-coded schemes, topics are explicitly determined by the documents' language. Subsequent analysis within one of these topic areas would thus be able to differentiate other dimensions of interest within an issue area, such as ideological direction (Lauderdale and Clark 2014).

Multiple algorithms are available for unsupervised topic classification. The standard, vanilla approach is latent Dirichlet allocation, or LDA. In LDA and many associated topic modeling approaches, documents are considered as arising from multiple latent topics – that is, documents can belong to multiple classifications, or mixed membership classification – an important improvement over many other clustering algorithms which classified documents into single topics – that is, single membership classification (Blei, Ng and Jordan 2003). Assuming a certain number of topics associated with a set of documents and having observed the words which make up the documents, we estimate the proportion of each document which exhibits each topic (Blei and Lafferty 2009).[7] The primary value of the approach lies in the estimation of the relative proportion of each document arising from a topic, or topic proportions. In so doing, the approach allows researchers to move past unidimensional assumptions to more granular reflections of the Court's attention.

One major advance in recent years was the introduction of Structural Topic Models (STM) by Roberts et al. (2013). The STM advances prior work by allowing users to incorporate document-level covariates affecting the estimation of topics. For instance, one common covariate affecting the estimation of topics is *time*; in our context, the Supreme Court is well-recognized as having an agenda which shifted markedly over time (e.g., Pacelle 1995). By incorporating document-level covariates for the term of the Court, one can incorporate our understanding of factors affecting agenda change while also directly and more appropriately modeling the Court's issue attention. Applications of the STM modeling approach have already yielded important insights for a diversity of relevant political science work, including survey research (Roberts et al. 2014), international bargaining (Tingley 2017), lobbying and trade (Kim 2017), and constitutional design (Law 2016). In this vein, I estimate a series of

[7]Inference is performed on the posterior probabilities through one of multiple possible methods (Blei, Ng and Jordan 2003), frequently Gibbs sampling (Griffiths and Steyvers 2004).

structural topic models to demonstrate the value of this approach, and mixed-membership topic models more generally, for understanding the issue attention of the Court.

Estimating Topic Attention

To begin, I acquired the texts of all majority opinions in the Court's history from `Justia`, an online repository of legal information.[8] In total, the corpus features 25,291 majority opinions, with the earliest opinions in 1803 and the most recent opinions in 2010. I retain the 5,000 most frequent unigrams (i.e., "constitution", "congress") and the 2,500 most frequent bigrams (i.e., "civil rights", "congressional power").[9] I include bigrams in order to capture particular legal concepts which may be muddled in a unigram only approach.

Though STM requires very few predefined parameters, it does require prior selection of the number of topics (k) before estimation. Selection of the number of topics is, asBlei and Lafferty (2009) state, "a persistent problem in topic modeling" (11) as there exists no widely-agreed method for selecting the "correct" value. The problem arises, however, from the flexibility of the model; researchers can adjust the parameter in order to obtain more general or more specific estimates of topical content, and for different research enterprises different values are potentially appropriate. Rather than settling on a single k here, I instead extract latent topic proportions across multiple specifications in the analyses to follow. Finally, I create topic names by concatenating the three most representative terms by topic.[10]

[8]Minimal pre-processing was utilized in order to prepare the texts for analysis. Speficially, I removed numbers and punctuation with the exception of intra-word dashes, and I stripped capitalization from the corpus. In recent work, Denny and Spirling (2017), these choices were some of the least influential when it came to changing the substantive content of texts while vastly improving statistical inference.

[9]I include only the most frequent as – given Zipf's law – words occurring less than this frequency are generally quite rare and thus offer little help in identifying themes while slowing computation.

[10]A number of algorithms are available for identifying the most representative terms for

[include Table 1 about here]

Measurement models are assessed by a number of validity criteria. As evidence of the face validity of the measure as well as convergent (e.g., does the measure map to other measures which it should map to) and discriminant (e.g., does the measure diverge from other measures in useful ways) validity, I look here to topics identified through a 14 topic STM. I utilize 14 topics in order to maximize comparability to the 14 issue areas identified by the Supreme Court Database. In Table 1, I present the top three words characterizing each of the estimated topics for all 14 topics. For each topic, I also include the Supreme Court opinion with the highest proportion of the opinion devoted to that topic. Finally, I include both the SCD's assigned issue area code, and issue code.

Note first the estimated topics generally reflect substantively interpretable themes based solely on the automatically generated three-word labels. For instance, "invention", "patent", and "machine" are the defining words for one topic. Other well-identified topics include a topic related to interstate commerce ("interstate, the commission, interstate commerce"), criminal activity ("criminal, conviction, crime"), and labor relations ("employer, employees, labor"). Moreover, the issue areas assigned by the SCD to these opinions consistently indicate the automated approach identifies themes that closely match human understandings. In the case of the previous examples, the assigned issue areas are Economic Activity, Economic Activity, Criminal Procedure, and Unions, respectively, while the more granular assigned issues are Patent, Regulation of Railroads, Death Penalty, and Bargaining, respectively. Importantly, higher dimensional topic models are capable of recovering these more granular representations for scholars interested in them.[11]

Three topics require additional discussion based on the assigned labels. Consider first the

---

each topic. I utilize the FREX scoring algorithm, which identifies words which are both frequent within the topic and exclusive to that topic, relative to other topics.

[11]In the data release to accompany this article, I include a 250 topic model, approximating the issue category of the SCD.

topic for which the topic label is “program, id at, at”; alternative approaches to identifying the top words for the topic identify “epa”, “erisa”, “afdc”, and “eeoc” as characteristic terms of the topic. That is, the topic directly relates to statutory language and the Court’s statutory interpretation of legislation. Likewise, for the “secretary, the act, the secretary” topic other principle terms include “postmaster”, “the treasury”, and “the collector”, indicating the association of the topic with judicial interpretation of executive actions. Finally, the “writ, appeal, writ of” topic is associated with other procedural legal actions, including “of error”, “depositions”, “the appeal”, “jurisdiction”, and “decree.” Taken together, three topics for which the assigned names prove initially difficult to interpret instead seem to reflect the underlying activities of the three branches of government and the continuing interpretive dialogue in which the Court is involved (see, e.g., Fisher 2004). In all, the estimates of the Court’s attention reflect sensible divisions of the Court’s attention, identified without imposing any structure other than a temporal component, and providing proportional (rather than exclusive) estimates of attention.

The validity of the structural topic model approach can also be seen in a plot of average attention to issues over time.[12] In Figure 1, I plot the average attention to each SCD issue area in the upper-half of the figure, and the average attention to each of the 14 estimated topics in the bottom half of the figure. While sensible changes in attention according to the SCD are to be expected, the changes in attention according to the topic model are also sensible. Note, for instance, the near mirror similarity of patterns of attention as reflected by the SCD’s Criminal Procedure category and the estimated “criminal, conviction, crime” topic.

Throughout, we see evidence the estimated models generate topics sensible to human readers and coherently associated with concepts with which they should be associated (con-

[12] I utilize average attention per term in order to avoid falsely associating changes in the Court’s overall workload with changes in the Court’s proportion of attention devoted to an issue.

vergent validity). I turn then to a series of illustrations of the value-added from such an approach for many research applications. As these illustrations demonstrate, while the prominent existing approach may at times appropriate, for many other research applications the flexibility and specificity of topic models offer a more suitable measurement approach.

Illustration One: Related Topics

One research application for which topic models and the associated topic proportions are especially valuable is when theory implicates or suggests relationships between topics. In those situations, the issue and issue area variables available from the Supreme Court Database are, given single membership classification preferences, inappropriate. Instead, the simplification obscures important variations in behavior. Consider two recent articles. First, Harvey and Woodruff (2011) suggest systematic bias in issue coding introduced by coder's unconscious beliefs about the decision's ideological direction (Harvey and Woodruff 2011). Here, confirmation bias manifests in decisions over which of many potential issue areas is primary. By manipulating the choice of the single issue classifications – as is required by the coding protocol – the coder shifts the ideological direction of a vote, thereby matching expectations of case disposition and justice ideology. Second, recent work by Rice (2016) suggests dissenting opinion authors strategically emphasize alternative issue dimensions on which the case could be decided. Though consistent with the sort of heresthetical maneuvering identified at all other stages of Court decision-making (Epstein and Shvetsova 2002; Wedeking 2010; Black, Schutte and Johnson 2013), the findings are contrary to other prior work suggesting the Court primarily divides (and decides) solely on a single issue dimension in the overwhelming majority of cases. In both articles, the researchers document how the coding protocol – and specifically the unidimensional assumption – of the dominant measure of opinion content and Supreme Court issue attention yields measures of issue attention that are inappropriate for the purposes of some empirical research settings.

Alternatively, topic models and the associated topic proportions enable researchers to directly estimate the relationships between topics, and to examine variations in topic attention within opinions. Consider the relatively simple case of correlated topics. As an example, in *NFIB v. Sebelius*, the Court considered – in addition to jurisdictional questions – three primary issues: (1) the constitutionality of the individual mandate under Congress' Commerce Clause powers, (2) the constitutionality of the individual mandate under Congress' Taxing and Spending Clause powers, and (3) the constitutionality under principles of federalism of the withholding of federal Medicaid funding to states that do not establish exchanges. Importantly, these legal issues are frequently linked to one another in cases, as claims of federal infringements on state power often follow from congressional exercises of power under the Commerce Clause or Taxing and Spending Clause. Indeed, leading textbooks for constitutional law often explicitly join the issues together (e.g., Epstein and Walker 2017).

[include Figure 2 about here]

Identifying such correlated topics – well-understood by legal scholars and social scientists but hard to identify where preferential treatment is given to assigning a single, mutually-exclusive topic – is simple with structural topic models, which permit positive correlations among the estimated topics. To see this, in Figure 2, I plot a heatmap of the correlation matrix among topics. Each entry in the heatmap reflects the correlation coefficient between the identified row and column topics. In the plot, darker shades of orange indicate increasingly negatively correlated topics, while darker shades of blue indicate increasingly positively correlated topics. The topics are ordered according to hierarchical clustering of the resulting correlations, with black boxes indicating the identified clusters.[13] Before proceeding, note that if the same plot were drawn with the single membership treatment of the SCD's issue areas, each of the off-diagonal elements would be white, as they are effectively treated as zero.

Instead, with the topics estimated from a structural topic model we can identify a series

[13]Clusters are estimated using single-linkage agglomerative clustering.

of interesting relationships in the themes characterizing the Court's opinions. First, each of the clusters offers evidence of a coherent area of policy attention on the Court. The three main clusters feature topics that are strongly linked together. Notably, the clusters identify a group of correlated topics related to civil rights and liberties and another cluster related to economic issues. Yet though the topics cluster, one can still see positive correlations outside of the clusters as well; note, for instance, that the interstate commerce topic ("interstate", "the commission", "interstate commerce") is positively correlated with a topic related to congressional power ("usc", "cong", "sess") as well as a topic related to federalism and state power ("the state", "state", "constitution"). Turning back to the example of *NFIB v. Sebelius*, one can see that across the corpus of opinions – the history of Supreme Court attention – the topics from which the Court's questions arose consistently and positively relate to one another. In other words, discussion of one of these topics frequently accompanies the discussion of another.

Of course, increasing the number of topics would permit identifying even more specific attention relationships in Supreme Court opinions. Even at this general level, though, finding that topic proportions for theoretically connected concepts correlate across the corpus provides yet further evidence of the value in considering opinions as composed of topic proportions – rather than a single issue – for some research endeavors.

Illustration Two: Case Complexity

Topic proportions also offer a robust method for addressing case complexity, a concept regularly cited in law and politics research (e.g., Collins 2008; Maltzmann, Spriggs and Wahlbeck 2000; Carrubba and Zorn 2010) but with little resulting consensus (e.g., Goelzhauser and Vouvalis 2014). Part of this arises from the fuzzy nature of the concept, with little effort in prior research to clearly explicate a clean definition of the concept of case complexity. Breaking from this general pattern, Goelzhauser and Vouvalis (2014) define complexity as referring to "the amount of information required to process before reaching a decision" (3). This definition neatly characterizes most of the treatments of the concept in prior work.

As an example, I turn to Lax and Rader (2015), in which the authors model the influence of opinion authors over the maintenance (or defection from) the majority coalition. The authors hypothesize there will be more defection from the majority coalition when cases are more complex. The underlying theory suggests that cases with more areas for disagreement generate more difficult bargaining conditions, and thus increase the probability of defection. In order to capture case complexity, they utilize a dichotomozed version of the Supreme Court Database's coding of legal provisions, with 0 indicating a case with only one legal provision and 1 indicating a case with more than one legal provision.

Alternatively, consider topic models and the associated topic proportions in light of the theoretical underpinnings of case complexity. Definitions of case complexity focus on connecting the concept to the amount of information that needs to be processed. With topic proportions, the amount of information processed is laid bare. Each opinion is divided into the proportion of attention devoted to a particular topic; when attention is concentrated within a single topic area, little other information has been processed by the justices. On the other hand, when attention is diffuse across multiple topics, the Court is directly signaling the additional information which was necessarily gathered, considered, and evaluated in reaching the decision.[14]

Estimating concentration or diffusion of proportions is a well-studied area in economics and other related fields, developed to study, for instance, market shares. A standard approach to estimating concentration is the Herfindahl - Hirschman index (H-index), which reflects the extent to which attention is devoted to a single or multiple topics and is equal to:

$$C_d^* = \frac{(C_d - 1/N)}{1 - 1/N} \tag{1}$$

[14]Of course, for many applications the causal process would require an ex ante measure of complexity. For that, topic models of litigant briefs would be ideal.

where $C_d = \sum_{i=1}^{N} \hat{\pi}_i^2$ is the standard H-index, N is the number of topics, and $\hat{\pi}_i$ is the proportion of a topic captured in a given document. In order to enhance comparability, I multiply the topic concentration by -1; thus, higher values indicate documents which address many different topics, whereas lower values indicate documents more concentrated on a single topic.

I construct measures of topic concentration across 15, 25, and 50 topic structural topic models. With the measure of complexity based on topic concentration in hand, I replicate Lax and Rader (2015). In each the analyses, the dependent variable is whether or not the member of the majority coalition defected from the majority. I retain all covariates from Lax and Rader (2015) in the analysis. Importantly, for every variable other than the case complexity measure, all substantive results persist across specifications.

On the other hand, in not one of the three replications does case complexity reach standard levels of statistical significance. To demonstrate the difference across measures of case complexity, I estimate predicted probabilities across the range of values for the relevant measures – a dichotomy of the number of legal provisions in the original article, and an alternative measure based on topic concentration. The results appear in Figure 3. Because the results across topic dispersion based measures of complexity are consistent across the number of topics, I present only the results from the 15 topic specification. In the left panel of Figure 3, the increase in the probability of defection as one goes from one legal provision to more than one legal provision is clear. On the other hand, the change in the predicted probability across the entire range of values for topic dispersion makes clear the lack of a statistically or substantively significant effect from case complexity on the probability of defection. Thus, with a more granular, continuous measure of case complexity, one finds little evidence that case complexity influences the probability of defection.

Illustration Three: Robustness

Finally, the flexibility of topic model estimation offers an avenue by which to assess the robustness of empirical results across alternative specifications. The replication crisis in scientific research is well-documented (Gerber et al. 2010; Simonsohn, Nelson and Simmons 2013). Here, identifying a single topic structure potentially obscures important variation within topic. Though there are often theoretical reasons for preferring smaller or larger numbers of topics within an individual research project, present research is frequently resigned to employ the SCD's issue area codes. Yet it is well-understood that the choice of the number of topics may influence inference; indeed, this is subject of great debate in the topic modeling community (Wallach, Mimno and McCallum 2009). Because of the ease and flexibility of estimating models, however, estimating models across values of the number of topics offers an additional robustness check for analyses.[15]

To demonstrate, I turn to research on the litigant signal model (Baird 2004, 2007; Baird and Jacobi 2009*b*,*a*), which suggests the Supreme Court – in deciding cases – signals litigants as to areas it considers a priority. In responding to these signals, litigants substantively shift the pool of available cases available to the Court for review, providing the Court with an additional agenda-setting influence and yielding a more powerful policy-making institution. The research identifies significantly more court cases two years and even more prominently five years after the increase in the Supreme Court's prioritization of an issue area.

The dependent variable in Baird's 2004 analysis is the number of cases decided by the Supreme Court during a term. The independent variable of interest is a measure of policy priority, which Baird captures by aggregating the number of salient cases within a term for each policy area. Salience is measured through the (Epstein and Segal 2000) measure of case salience, which is simply an indicator for whether or not the decision of the Court was featured on the front-page of the *New York Times* the day after the decision.

[15]For an example of this in an applied setting, see Rice (2016).

Thus, both the dependent and primary independent variables in the dataset are counts of cases *within a given policy area* during a year. In that context, defining the boundaries of policy areas is obviously critical. Baird identifies eleven issue areas, relying almost entirely on the classification scheme of the SCD but making some minor modifications (e.g., combining some categories while dropping others from the analysis). Because the exact time period necessary for the Court's agenda to respond to the additional litigant attention is unclear, Baird specifies an autoregressive distributed lag model, with six lags of the primary independent variable, controls for Rehnquist and Burger Court eras, and a lagged dependent variable to account for temporal dynamics. In so doing, the specification of the model features six independent variables all directly tied to the identification of policy, in addition to a series of fixed effects for each policy area.

To again demonstrate the utility and flexibility of structural topic models, I turn to replicating the above analysis but with topic attention instead as estimated from a series of structural topic models. Specifically, I estimate six structural topic models, with k set to values of 5, 10, 15, 20, 25, and 50. Rather than case counts for each policy area, I sum the proportions of attention devoted to the policy area for all cases – the dependent variable – and for salient cases – the independent variable – with salience again captured through the *New York Times* measure.[16] Thus, the measure more accurately captures the Court's total attention to a policy area, as subordinate attention is explicitly incorporated.

In Figure 4, I plot a visualization of the relevant model results. Specifically, for each model specification, I plot the coefficient estimate for each of the six lags of political priority and the associated 95% confidence interval. Two observations bear discussion. First, the general result persists across a series of specifications; that is, throughout models we see that increases in the saliency of the Court's attention to a policy corresponds with increases in

[16]Consistent with Baird, the models include fixed effects for policy area. Rather than Chief Justice fixed effects, I included fixed effects for October Term, which I was able to do here because I did not face the model identification problem confronting Baird's specification.

subsequent Court attention to that policy.

Second, though, the lag at which we would conclude there is a relationship varies across every estimation. The most consistent lags are the three and fifth year lags, both of which are positive and significantly related to subsequent attention in all but one specification. That inconsistency makes it difficult to conclusively settle on the particular length of time at which the Court's prioritization of a policy might generate subsequent. Indeed, at larger numbers of topics nearly every subsequent year approaches or exceeds standard levels of statistical significance, suggesting the Court's agenda-setting influence is important but temporally dispersed. As before, substantive inferences are adjusted based on incorporating the additional information available from the mixed-membership classification and doing so across a variety of specifications.

Discussion and Conclusion

In this article, I introduce a series of measures based on a now well-accepted approach from computational linguistics for capturing textual content. For many research applications, these proportions offer marked improvements over existing measures. First, they overcome data lost through single membership classification by instead providing estimates of the proportion of each document addressing each topic. In so doing, the approach better reflects opinion content in ways consistent with the movement of the field toward the strategic model and more complex assessments of judicial behavior. Moreover, they offer a flexible approach for scholars interested in studying the Court's attention, as they can be simply adjusted and estimated in order to yield more general topical frameworks or highly specific frameworks. To that end, all texts and code necessary for scholars to replicate and extend the models for their own purposes are publicly available, enabling future research to incorporate more sophisticated understandings of opinion content where necessary to improve causal inference. Moreover, for the researcher who wishes to use these topics to classify new opinions into identical topics, the estimated models are available, providing an avenue to directly compare results across updated corpora.

Though the above demonstrates the value of topic models, if anything I understate the potential of topic models, and in particular structural topic models, for empirical legal research. As noted previously, structural topic models can include covariates which help to structure and define either of topical prevalence, or how prevalent the topic is within a document, and topical content, or what words are used within a topic. This effectively allows one to analyze the influence of different covariates of interest on the topic distributions. In the context of the above models, I have estimated the effect of time on the prevalence of topics, which helps to address temporal dynamics. But consider alternative specifications, for instance a variable specifying the Rehnquist Court. Here, one can then examine the effect of those covariates on topical prevalence or content. In the case of a Rehnquist Court variable, one might expect to see a specific effect on federalism. In other areas, recent work has leveraged similar approaches in analyzing differences in open-ended survey responses (Roberts et al. 2014). For the Court, the opportunities are myriad. For instance, one could even extend the Rehnquist Court example, potentially analyzing differences precipitated by ideology on the prevalence of federalism arguments, or on the language associated with federalism arguments.

Finally, moving forward, research might look to build on the approach identified here to construct increasingly sophisticated indicators of the content of legal documents. The approach utilized here should be seen as an important first step in this process, but the effort to identify and measure opinion content should certainly continue. Legal texts are – as legal scholars have frequently reminded social scientists (Friedman 2006) – much more complex than often presumed in empirical research. Alternative methods for estimating topics from text corpora and minor improvements and adjustments to existing approaches continue to be rapidly introduced across a variety of technical and applied fields of research, and as these tools become established it may become clear they offer yet further improvements on that which I have introduced in this article. I leave this to future research.

References

Baird, Vanessa. 2004. "The Effect of Politically Salient Decisions on the U.S. Supreme Court's Agenda." *The Journal of Politics* 66(3):755–772.

Baird, Vanessa. 2007. *Answering the Call of the Court: How Justices and Litigants Set the Supreme Court Agenda*. Charlottesville: University of Virginia Press.

Baird, Vanessa and Tonja Jacobi. 2009*a*. "How the Dissent Becomes the Majority: Using Federalism to Transform Coalitions in the U.S. Supreme Court." *Duke Law Journal* 59:183–238.

Baird, Vanessa and Tonja Jacobi. 2009*b*. "Judicial Agenda-Setting through Signaling and Strategic Litigant Responses." *Washington University Journal of Law and Policy* 29:215–239.

Benesh, Sara and Harold Spaeth. 2007. "The Constraint of Law: A Study of Supreme Court Dissensus." *American Politics Research* 35(5):755–768.

Black, Ryan, Rachel Schutte and Timothy Johnson. 2013. "Trying to Get What You Want: Heresthetical Maneuvering and U.S. Supreme Court Decision Making." *Political Research Quarterly* 66:819–830.

Blei, David, Andrew Ng and Michael Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022.

Blei, David and John Lafferty. 2009. Topic Models. In *Text Mining: Classification, Clustering, and Applications*, ed. A Srivastava and M Sahami. Chapman and Hall/ CRC Press.

Carrubba, Cliff and Christopher Zorn. 2010. "Executive Discretion, Judicial Decision Making, and Separation of Powers in the United States." *The Journal of Politics* 72(3):812–824.

Collins, Paul. 2008. *Friends of the Supreme Court: Interest Groups and Judicial Decision Making*. Oxford University Press.

Denny, Matthew and Arthur Spirling. 2017. "Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do About It." *Political Analysis* .

Epstein, Lee, Jack Knight and Andrew Martin. 2003. "The Political (Science) Context of Judging." *St. Louis University Law Journal* 47:783–817.

Epstein, Lee and Jeffrey Segal. 2000. "Measuring Issue Salience." *American Journal of Political Science* 44(1):66–83.

Epstein, Lee and Olga Shvetsova. 2002. "Heresthetical Maneuvering on the U.S. Supreme Court." *Journal of Theoretical Politics* 14(1):93–122.

Epstein, Lee and Thomas Walker. 2017. *Constitutional Law for a Changing America: Institutional Powers and Constraints*. CQ Press.

Evans, Michael, Wayne McIntosh, Jimmy Lin and Cynthia Cates. 2007. "Recounting the Courts? Applying Automated Content Analysis to Enhance Empirical Legal Research." *Journal of Empirical Legal Studies* 4:1007–1039.

Fisher, Louis. 2004. *Making Policy, Making Law: An Inter-branch Perspective*. Georgetown University Press chapter Judicial Finality or Ongoing Colloquoy?, pp. 153–169.

Flemming, Roy, B. Dan Wood and John Bohte. 1999. "Attention to Issues in a System of Separated Powers: The Macro-Dynamics of American Policy Agendas." *Journal of Politics* 61(1):76–108.

Flemming, Roy, John Bohte and B. Dan Wood. 1997. "One Voice Among Many: The Supreme Court's Influence on Attentiveness to Issues in the United States." *American Journal of Political Science* 41(4):1224–1250.

Friedman, Barry. 2006. "Taking Law Seriously." *Perspectives on Politics* 4(2):261–276.

Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University.

Gerber, Alan, Neil Malhotra, Conor Dowling and David Doherty. 2010. "Publication Bias in Two Political Behavior Literatures." *American Politics Research* 38(4):591–613.

Goelzhauser, Greg and Nicole Vouvalis. 2014. Measuring Case Complexity. In *Annual Conference of the Midwest Political Science Association*. Chicago, Illinois: .

Griffiths, Thomas L. and Mark Steyvers. 2004. "Finding scientific topics." *Proceedings of the National Academy of Sciences* 101(suppl 1):5228–5235.

Grimmer, Justin. 2010. "A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases." *Political Analysis* 18:1–35.

Harvey, Anna and Michael Woodruff. 2011. "Confirmation Bias in the United States Supreme Court Judicial Database." *The Journal of Law, Economics, and Organization* .

Kim, In Song. 2017. "Political Cleavages within Industry: Firm-Level Lobbying for Trade Liberalization." *American Political Science Review* .

Lauderdale, Benjamin E. and Tom S. Clark. 2014. "Scaling Politically Meaningful Dimensions Using Texts and Votes." *American Journal of Political Science* pp. n/a–n/a.

Laver, Michael, Kenneth Benoit and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 47:215–233.

Law, David. 2016. "Constitutional Archetypes." *Texas Law Review* 95:153–245.

Lax, Jeffrey and Kelly Rader. 2015. "Bargaining Power in the Supreme Court: Evidence from Opinion Assignment and Vote Switching." *The Journal of Politics* 77(3):648–663.

Maltzmann, Forrest, James Spriggs and Paul Wahlbeck. 2000. *Crafting Law on the Supreme Court: The Collegial Game*. New York: Cambridge University Press.

McAtee, Andrea and Kevin McGuire. 2007. "Lawyers, justices, and issue salience: When and how do legal arguments affect the US Supreme Court?" *Law & Society Review* 41:259–278.

McGuire, Kevin and Georg Vanberg. 2005. "Mapping the Policies of the U.S. Supreme Court: Data, Opinions, and Constitutional Law." Prepared for delivery at the Annual Meeting of the American Political Science Association, Washington, D.C., September 1-5.

McGuire, Kevin, Georg Vanberg, Charles Smith and Gregory Caldeira. 2009. "Measuring policy content on the U.S. Supreme Court." *Journal of Politics* 71(4):1305–1321.

Pacelle, Richard. 1995. The Dynamics and Determinants of Agenda Change in the Rehnquist Court. In *Contemplating Courts*, ed. Lee Epstein. Washington, D.C.: Congressional Quarterly.

Peters, C. Scott. 2007. "Getting Attention: The Effect of Legal Mobilization on the U.S. Supreme Court's Attention to Issues." *Political Research Quarterly* 60:561–572.

Quinn, Kevin, Burt Monroe, Michael Crespin, Michael Colaresi and Dragomir Radev. 2010. "How to Analyze Political Attention With Minimal Assumptions and Costs." *American Journal of Political Science* 54:209–228.

Rice, Douglas. 2014. "The Impact of Supreme Court Activity on the Judicial Agenda." *Law & Society Review* 48(1):63–90.

Rice, Douglas. 2016. "Issue Divisions and U.S. Supreme Court Decision Making." *Journal of Politics* .

Richards, Mark and Herbert Kritzer. 2002. "Jurisprudential Regimes in Supreme Court Decision Making." *American Political Science Review* 96:305–320.

Richards, Mark, Joseph Smith and Herbert Kritzer. 2006. "Does Chevron Matter?" *Law & Policy* 28:444–469.

Roberts, Margaret, Brandon Stewart, Dustin Tingley, Christopher Luca, Jetson Leder-Luis, Shana Gadarian, Bethany Albertson and David Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* .

Roberts, Margaret, Brandon Stewart, Dustin Tingley and Edoardo Airoldi. 2013. The Structural Topic Model and Applied Social Science. In *Proceedings of the 2013 NIPS Workshop on Topic Models: COmputation, Application, and Evaluation*.

Roberts, Margaret, Brandon Stewart and Edoardo Airoldi. 2016. "A Model of Text for Experimentation in the Social Sciences." *Journal of the American Statistical Association* 111(515):988–1003.

Segal, Jeffrey. 1984. "Predicting Supreme Court Cases Probabilistically: The Search-and-Seizure Cases, 1962-1981." *American Political Science Review* 78(4):891–900.

Segal, Jeffrey. 1986. "Supreme Court Justices as Human Decision-Makers: An Individual Level Analysis of the Search-and-Seizure Cases." *American Journal of Political Science* 48:938–955.

Shapiro, Carolyn. 2009. "Coding Complexity: Bringing Law to the Empirical Analysis of the Supreme Court." *Hastings Law Journal* 60:477–540.

Simonsohn, Uri, Leif Nelson and Joseph Simmons. 2013. "P-Curve: A Key to the File Drawer." *Journal of Experimental Psychology* Forthcoming.

Szmer, John and Barry Edwards. 2011. "Future Directions in Data Analysis and Collection." *Law and Courts Newsletter* 21:16–19.

Tingley, Dustin. 2017. "Rising Power on the Mind." *International Organization* 71(1):165–188.

Wallach, Hanna, David Mimno and Andrew McCallum. 2009. Rethinking LDA: Why Priors Matter. In *NIPS*.

Wedeking, Justin. 2010. "Supreme Court Litigants and Strategic Framing." *American Journal of Political Science* 54(3):617–631.

Case	Issue Area	Issue
STM = invention_patent_machine		
Hoffheins v. Russell (1882)	Economic Activity	Patent
STM = writ_appeal_writof		
Buckingham v. McLean (1851)	Due Process	Hearing or Notice
STM = tax_income_thetax		
Kirby Petroleum v. Cmr. of Internal Revenue (1945)	Federal Taxation	Internal Revenue Code
STM = criminal_conviction_crime		
Hopper v. Evans (1981)	Criminal Procedure	Death Penalty
STM = indian_treaty_war		
Roff v. Burney (1897)	Civil Rights	Indians
STM = interstate_thecommission_interstatecommerce		
Ayrshire Colleries v. U.S. (1948)	Economic Activity	Regulation of Railroads
STM = lands_thelands_theland		
U.S. v. Heirs of Arredondo (1839)	Economic Activity	Land Claims
STM = bankruptcy_creditors_suits		
Russell v. Todd (1939)	Federalism	Federal Pre-emption
STM = thestate_constitution_state		
Mason v. Missouri (1900)	Civil Rights	Reapportionment
STM = program_idat_at		
Cuomo v. Clearing House Asso. (2008)	Federalism	Federal Pre-emption
STM = vessel_thevessel_ship		
Kennedy's Claim (1814)	Due Process	Takings
STM = bonds_thebank_railroadcompany		
Howard County v. Boonville Bank (1882)	Economic Activity	Regulation of Business
STM = secretary_theact_thesecretary		
McLean v. Fleming (1877)	Economic Activity	Trademark
STM = employer_employees_labor		
Medo Photo Supply. v. NLRB (1943)	Unions	Bargaining

Table 1: *Top cases for each estimated topic* . Each entry indicates the topic generated from a 14 topic structural topic model, and the case with the highest proportion for that topic. The second and third columns represent the broad issue area and the more specific issue, respectively, both as coded by the Supreme Court Database.

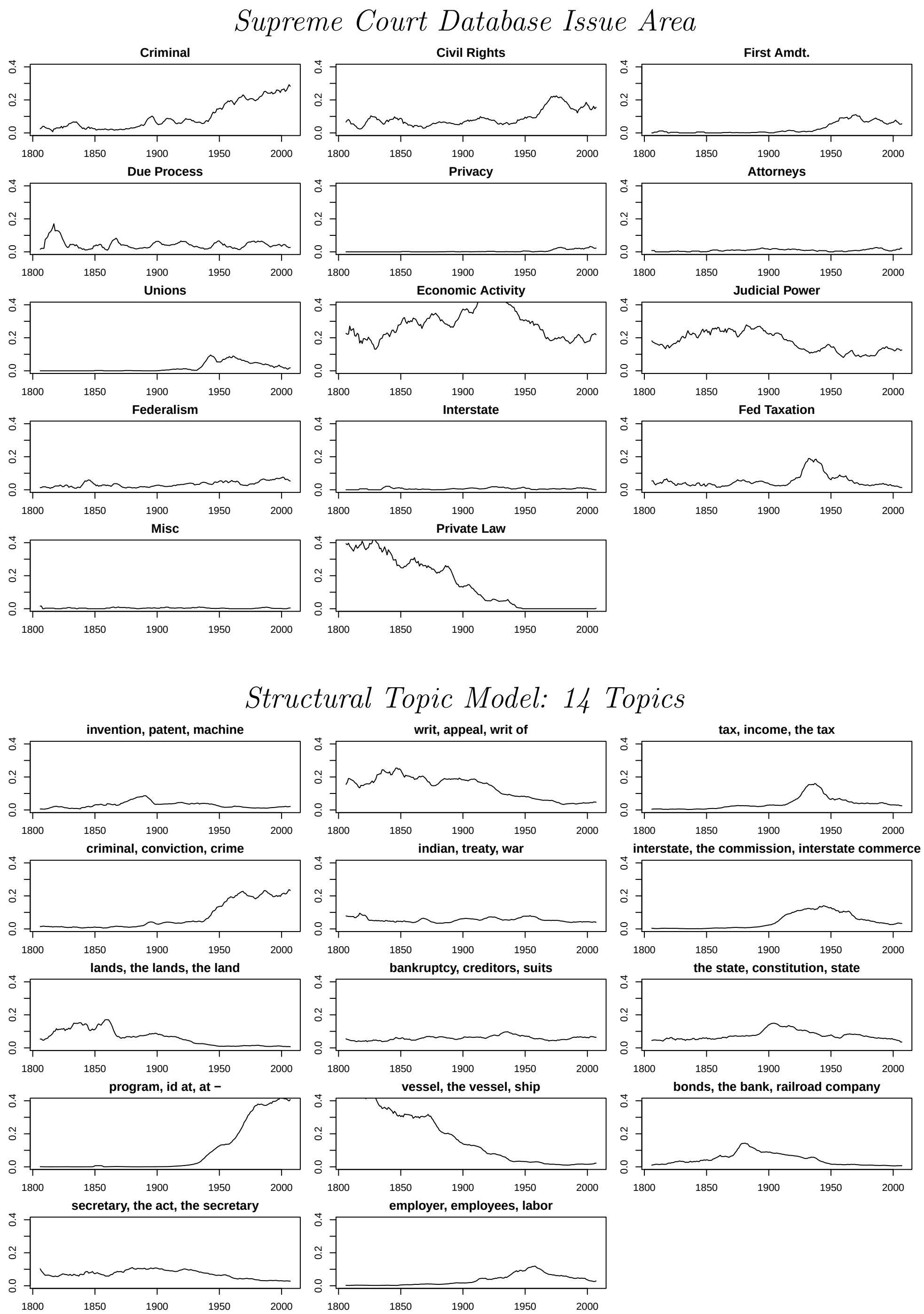

Figure 1: *Changes in Supreme Court Issue Attention Over Time* Plots of Average Issue Attention by Supreme Court Database Issue Area Codes (top plots) and 14 Topic Structural Topic Model (bottom plots).

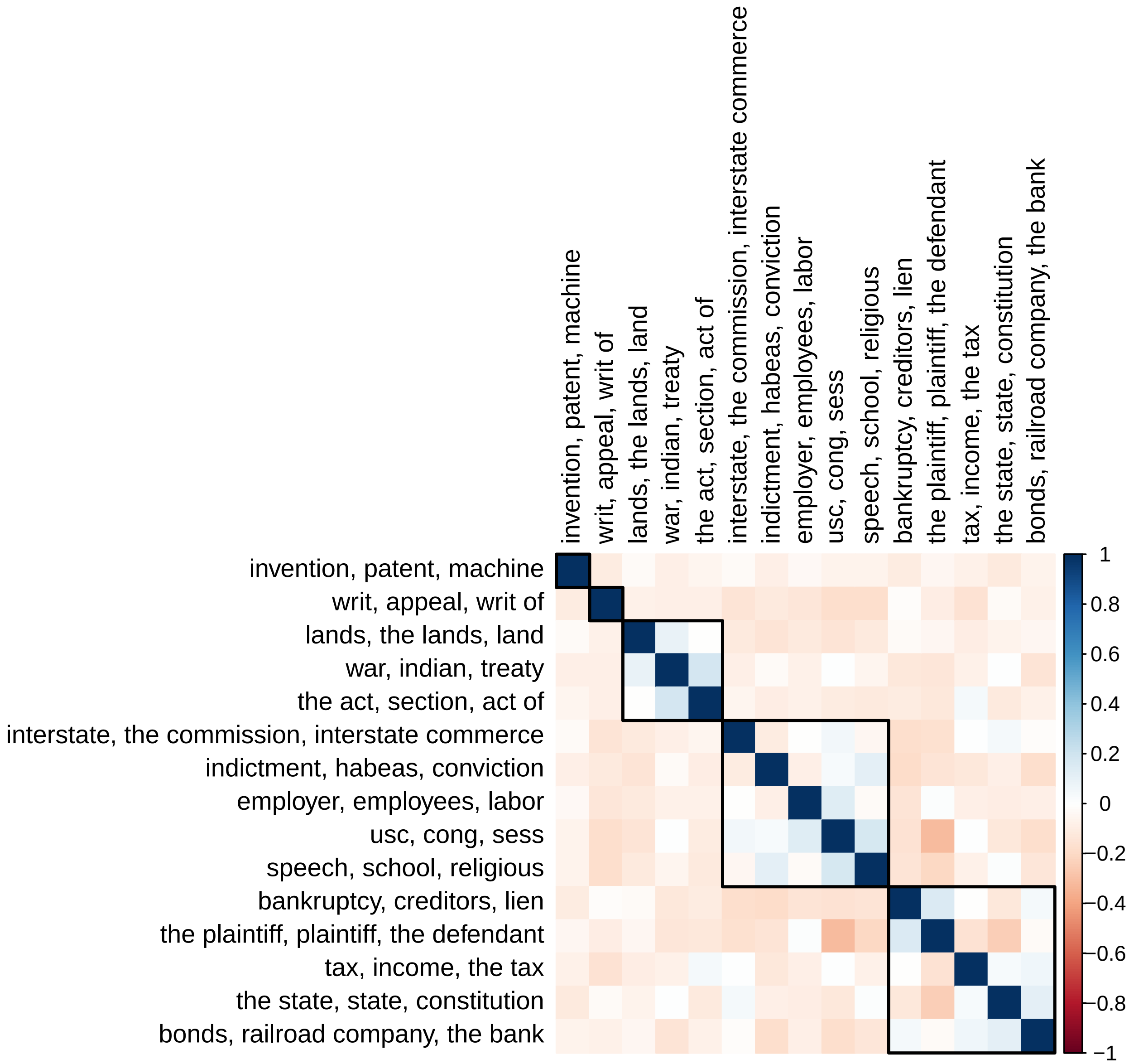

Figure 2: *Correlation of Estimated Topic Proportions* Heatmap of correlations between estimated topic proportions. Blue indicates positive correlations while orange represents negative correlations, with the darkness of the shading representing the magnitude of the correlation. Black boxes indicate hierarchical clusters estimated according to single-linkage clustering.

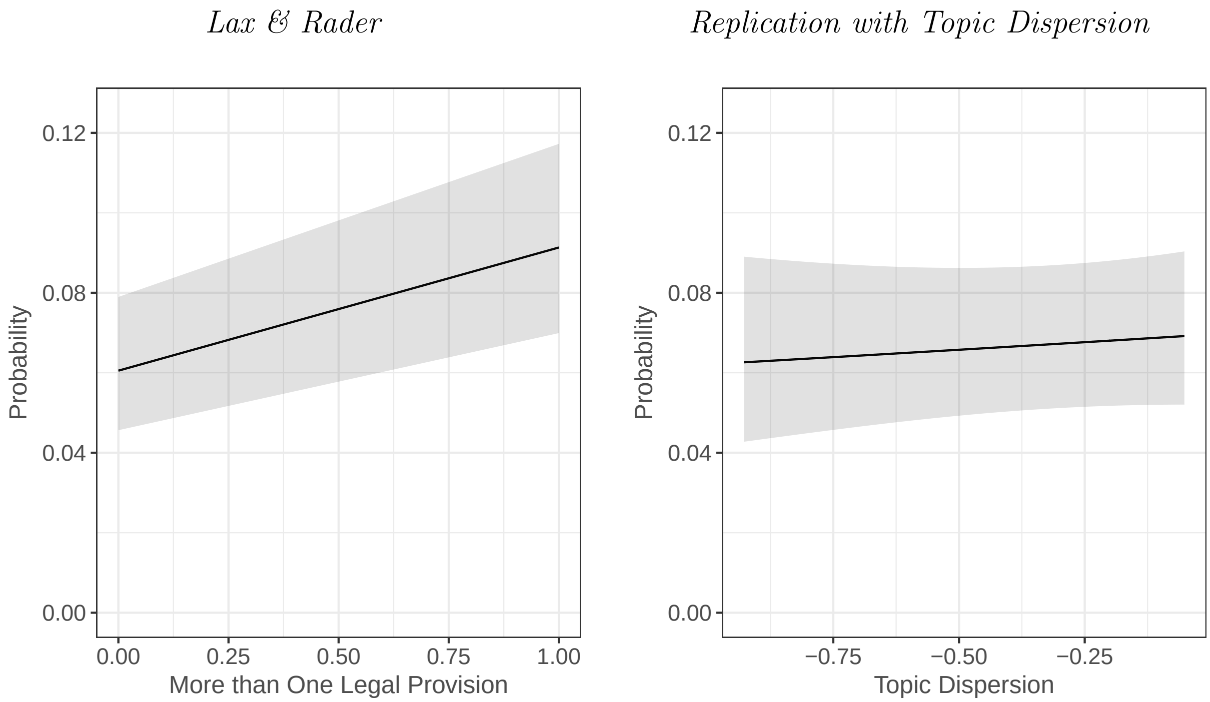

Figure 3: *Plots of Predicted Probabilities* These plots provide the probability (black line) and associated 95% confidence intervals for defection from the majority for the the legal provisions based measure (left panel) and the topic models based measure (right panel).

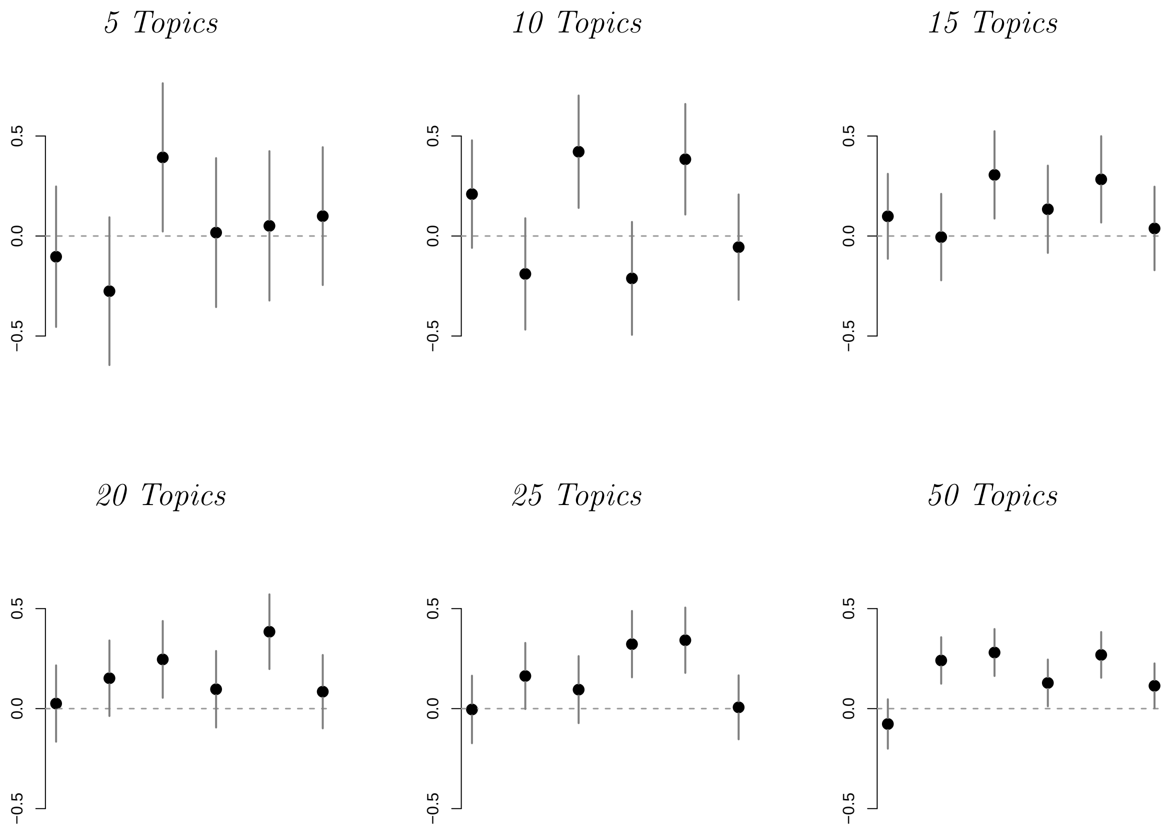

Figure 4: *Coefficient Estimates from Replications of Baird (2004) Using Topic Estimates* These plots provide the coefficient estimates (black points) and associated 95% confidence intervals (grey bars) for the effect of lags of Supreme Court issue prioritization on future Supreme Court issue attention. Each separate plot is estimated based on topic attention measures from a structural topic model with the indicated number of topics.