Strategic Issue Selection and Ideological Polarization: Evidence from the Congressional Record Data

Imil Nurutdinov

This draft: August 2, 2013

In this paper, I examine how selection of divisive issues in the Congress relates to the ideology of candidates and electoral incentives. I develop an extension of the existing models of automated content analysis implementing a general discrete choice model. This model accounts for observed individual-level and time-specific characteristics, as well as unobserved choice attributes. Using the Congressional Record data I find that during general elections, divergence on the issues like immigration, domestic security, and the Iraq War is not related to ideology. However, there is evidence that more extreme candidates are more likely to run their campaigns on issues like stem cell research and alternative energy than moderate ones.

Acknowledgements

I am grateful to Ruben Enikolopov, Maria Petrova, Kirill Kalinin, Konstantin Sonin, Andrei Bremzen, Grigory Kosenok, Shlomo Weber, Evgeny Yakovlev, Sergey Chernov, Samuel Greene, Nikolay Khomutov, Maxim Ananyev, Ilia Krasikov, Kirill Borusyak, and the political economy seminar participants at the New Economic School for helpful discussions and valuable comments.

Contents

Introduction

When transitioning from the 109th to 110th Congress, the median representative, who was too conservative before elections, became too liberal, compared to the median U.S. voter (Bafumi and Herron, 2010). This evidence suggests that the median voter was somewhat misrepresented both before and after elections and violates the Downs (1957) median voter theorem.[1] The main objective of this paper is to examine whether the prevalence of certain issues in public discussion, such as immigration and foreign policy, on which two major parties diverge, is associated with the ideology of politicians and proximity of elections.

Specifically, I consider a static model in which candidates from two parties select their issues conditional on their characteristics and the current state of the world—primary election, general election, or none. But they not only choose what issues their campaigns are about—they also announce their positions on those issues. Depending on the distribution of voters' preferences in each state of the world, it might be optimal to choose different issues and take different positions across states of the world.[2] I do not, however, model how exactly the policies are announced.[3] Rather, I focus on those issues about which the researcher knows *ex ante* that candidates from the competing parties take non-median positions. That is, we cannot say whether two candidates agree on the fiscal policy but we know for sure that they divide on religious values.[4] The list of examples of divisive issues might

[1] Meanwhile, according to Bafumi and Herron (2010), the Senate appears to be a more moderate institution.

[2] I allow candidates to have policy preferences.

[3] This is partially due to analytical complexity. To the best of my knowledge, Egorov (2012) is the only recent paper that considers a game-theoretic model of political competition with endogenous selection of issues, in which candidates have to choose a few issues from the pool.

[4] As it is the case in modern U.S. politics(Glaeser et al., 2005).

include not only moral issues but also foreign policy, immigration, environmental issues, etc. My goal is to analyze how selection of divisive issues depends on the candidates preferences, i.e., ideology, and incentives to win primary and general elections.

But how can one empirically examine which issues matter in day-to-day discussions? Even though a media coverage of political events is intense, for a thorough answer, this question requires comprehensive text analysis of speech transcripts and press releases. During recent years, different approaches to the automated analysis of political texts have been developed (e.g., Hopkins and King, 2010; Quinn et al., 2010). An increasing number of these models are based on Latent Dirichlet allocation (LDA; Blei et al., 2003) which has the flexible hierarchical nature.[5] The pitfall of these methods is that they are primarily intended to fit the data, neglecting important information about how political texts were generated, so that only limited inference is possible.

To address this difficulty, I extend the existing LDA-based approaches endogenizing legislators' topic choices. In particular, I adopt a random utility model suggested in McFadden and Train (2000) in the context of strategic issue selection, so that one can estimate how observed demographic or time-specific characteristics affect the decision to speak about a certain topic. Chamber, party, ideology, electoral cycle, district-level information—all these variables not only better explain the variation in the data but, more importantly, enter as components of individuals' utilities. Using my endogenous attention allocation model (EAA), social scientists can answer the questions of the following sort: What is the impact of the party affiliation or proximity of elections on how legislators behave in the

[5]For a comprehensive overview of computer-assisted models, see Grimmer and Stewart (2013).

Congress? Though here I follow the "reduced-form" approach and estimate the utility parameters for the cross-section data only, in principle it can be generalized to the dynamic decision-making and estimation of other quantities of substantive interest, such as "inertia" in issue selection.

Conceptually, a work most closely related to mine is Roberts et al. (2013). Their approach, the "structural topic model" (STM), also allows to take into account a number of covariates associated with each document and estimate the "average treatment effect".[6] However, in settings similar to mine, EAA has at least two desirable properties compared to STM. First, I explicitly build EAA on a set of economic considerations and provide an interpretation of the estimates, as it is common in the economics and marketing literature. Second, when deriving these estimates, I assume that legislators observe attributes of issues, while the researcher does not. Yet, as shown in section 1.4, STM may be a more appropriate tool than EAA in other applications.

My findings suggest that during general elections, divergence on some issues, e.g., immigration, domestic security, and the Iraq War, is not associated with the ideology of candidates. At the same time, along other dimensions of the conflict, such as stem cell research and alternative energy, there is evidence that more extreme politicians engage in debates on those issues more frequently.

The political economy literature considers several environments in which platform divergence may be welfare-enhancing. For instance, Bernhardt et al. (2009) show that if candidates do not know voters' preferences perfectly, divergence of platforms is socially desirable. The model in Weelden (2013a) suggests that po-

[6] It is worth noting that the word "structural" here implies no structural relationship between a theoretical model and a particular dataset.

larization between citizen-candidates who cannot bind a certain policy decreases the amount of rent-seeking in equilibrium. Weelden (2013b) shows that divergent platforms are socially optimal if the electorate is polarized itself and the utility function is not too concave. Also, Weelden (2013b) generalizes to the multidimensional case and finds that it is beneficial to have greater polarization on issues on which voters are similarly polarized. Kamada and Kojima (2013) introduce convex utility, which appears to be more suitable for moral and religious issues, and demonstrate that in such a case an equilibrium with divergent platforms is welfare-maximizing. If there are several dimensions, Kamada and Kojima (2013) predict that candidates converge on "concave" issues and diverge on the "convex" ones.[7]

The contribution of the paper is twofold. First, to the best of my knowledge this is the first paper that studies the problem of polarization in the U.S. Congress using text inputs not for descriptive or exploratory purposes, but for making inferences about underlying incentives of political actors. Second, I nest the variant of LDA within the discrete choice framework. Doing so, I derive the estimates for observed individual- and time-level characteristics, accounting for a potentially large number of unobserved topic attributes. My model, therefore, captures the unobserved heterogeneity in preferences not only across individuals—a common feature of the applications of the mixed logit model—but across alternatives as well. Furthermore, with minor changes my EAA algorithm can be applied not

[7] Also, there are a number of alternative explanations of polarization. Glaeser et al. (2005) focus on one-stage elections in which voters can abstain and politicians can send targeted messages to the "core" electorate. Hummel (2010) and Hummel (2013) study two-stage elections with primaries and analyze to what extent candidates can run their campaigns on moderate platforms during general elections. For a general discussion on political polarization in the Congress, see, e.g., Fiorina (1999) and Abramowitz and Saunders (2008).

only to topic models but also to the roll-call voting and other instances in which choice alternatives—bills, services, products, etc.—cannot be represented in the conventional attributes space.[8] This innovation expands the area of applicability of discrete choice methods.

The rest of the paper is organized as follows. Section 1 introduces the statistical model of issue selection. Section 2 describes the data. In sections 3 and 4, I discuss the results, implications, and possible directions of future research. Section 5 concludes.

1. The Statistical Model of Issue Selection

Below I propose the endogenous attention allocation model (EAA). I do it in several steps. First, I describe the random utility model taking the set of alternatives as given. Second, I discuss how the choice alternatives, i.e., topics, are defined in this context. Third, I briefly speculate on estimation. Finally, I discuss how EAA relates to the existing topic models.

[8] There are two main reasons for that. In some cases, the researcher may face a considerable difficulty when trying to describe concepts of the complex nature. This is especially true for the decision to speak about a certain topic, because it is ambiguous what attributes characterize a set of topics well enough. Additionally, choice attributes might be subject to a substantial measurement error. Possible examples of this sort include such attributes as "attractiveness" or "political correctness".

1.1. Random Utility

Assume that legislator i's choice in time period t, c_{it}, is associated with the following utility maximization problem,

$$\max_{k \in \{1,\dots,K\}} g(\mathbf{x}_{it}, \boldsymbol{\nu}_k), \tag{1.1}$$

where k indexes topics, $\mathbf{x}_{it}$ is a vector of H individual- and time-specific attributes values observed both by the legislator and econometrician[9], $\boldsymbol{\nu}_k$ is a vector of P topic attributes observed by the legislator only. I assume that at a time every legislator speaks only about one topic.

Specify utility as follows:

$$U(c_{it} = k) = \mathbf{x}'_{it}\boldsymbol{\beta}^o + \boldsymbol{\nu}'_k\boldsymbol{\beta}^u + \varepsilon_{itk}, \tag{1.2}$$

where each ε_{itk} is iid from a Gumbel distribution[10] and superscripts "o" and "u" read as "observed" and "unobserved", respectively. The non-stochastic part of 1.2 can be viewed as a linear approximation of true utility. Then, assume for a while that $\boldsymbol{\nu}_k$ is observed by the researcher. Notice that if the utility function is linear in its parameters, $\boldsymbol{\beta}^o$ and $\boldsymbol{\beta}^u$ can be viewed as the marginal utilities, since $\dfrac{\partial U}{\partial \mathbf{x}'_{it}} = \boldsymbol{\beta}^o$ and $\dfrac{\partial U}{\partial \boldsymbol{\nu}'_k} = \boldsymbol{\beta}^u$. Legislator i decides to speak about topic k in time period t if and only if

$$U(c_{it} = k) > U(c_{it} = l)\ \forall l \neq k.$$

[9] The constant enters $\mathbf{x}_{it}$.

[10] Recall that the difference between two Gumbel random variables has a logistic distribution.

The implied multinomial probabilities are as follows (McFadden, 1973):

$$\begin{aligned}
\Pr\left[c_{it} = k \mid \mathbf{x}_{it}, \boldsymbol{\nu}_k\right] &= \Pr\left[U(c_{it} = k) > U(c_{it} = l)\ \forall l \neq k \mid \mathbf{x}_{it}, \boldsymbol{\nu}_k\right] \\
&= \frac{\exp\left((\mathbf{x}_{it} + \boldsymbol{\nu}_k)'\boldsymbol{\beta}\right)}{\sum_{l=1}^{K} \exp\left((\mathbf{x}_{it} + \boldsymbol{\nu}_l)'\boldsymbol{\beta}\right)} \\
&= \sigma\left((\mathbf{x}_{it} + \boldsymbol{\nu}_k)'\boldsymbol{\beta}\right),
\end{aligned} \tag{1.3}$$

where $\boldsymbol{\beta}$ is the $H + P$ vector.

It is well-known that the multinomial logit specification 1.3 suffers from several limitations.[11] Here I focus on the two. First is the independence from irrelevant alternatives (IIA) property—that is, substitution over topics is proportional. Second is ignoring the random taste variation—multinomial logit cannot explain differences in tastes that are unrelated to $\mathbf{x}_{it}$ and $\boldsymbol{\nu}_k$. In other words, multinomial logit fails to explain why legislators with similar observed characteristics $\mathbf{x}_{it}$ tend to speak about different topics, even if topic attributes $\boldsymbol{\nu}_k$ are similar as well.

Mixed multinomial logit (MML) has attracted a considerable attention in the economics and marketing literature as a potential solution (Allenby and Rossi, 1999; McFadden and Train, 2000). Specifically, MML relaxes the assumption that utility parameters $\boldsymbol{\beta}^o$ and $\boldsymbol{\beta}^u$ are constant in the whole population. Instead of 1.3 let us consider the following model,

$$U(c_{it} = k) = (\mathbf{x}_{it} + \boldsymbol{\nu}_k)'\boldsymbol{\beta}_i + \varepsilon_{itk},$$

where $\boldsymbol{\beta}_i$ is the vector of legislator i's utility parameters and has a normal distri-

[11] See, e.g., Train (2003) for a discussion of limitations of multinomial logit and different ways to deal with them.

bution $F(\boldsymbol{\beta}_i)$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. The advantage of this model is that it allows for arbitrary correlations between topic choices for the same legislator. This means, for instance, that if two topics, k and l, have similar attributes, $\boldsymbol{\nu}_k$ and $\boldsymbol{\nu}_l$, legislator i's decision to speak about one topic is positively correlated with that to speak about another. Additionally, the variation in $\boldsymbol{\beta}_i$ across legislators can explain why legislators similar in their observed individual characteristics $\mathbf{x}_{it}$ can make different topic choices.

Generally the researcher makes inferences about population-level parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, not $\boldsymbol{\beta}_i$.[12] In order to derive the expected choice probabilities, one needs to apply formula 1.3 and integrate out $\boldsymbol{\beta}_i$,

$$\Pr\left[c_{it} = k \mid \mathbf{x}_{it}, \boldsymbol{\nu}_k, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right] = \int \sigma\left((\mathbf{x}_{it} + \boldsymbol{\nu}_k)'\boldsymbol{\beta}_i\right) \mathrm{d}F(\boldsymbol{\beta}_i). \tag{1.4}$$

Note that due to the presence of the additive idiosyncratic term, ε_{itk}, this model predicts non-zero probability for *every* topic, whenever legislator i makes her choice. However, as Athey and Imbens (2007) argue, the gain of including this term in the random coefficients framework is the increased robustness to measurement error in the dependent variable, c_{it}. This property is particularly important in my framework, because as I point out in subsection 1.3, I treat c_{it} as latent and hence the misprediction of c_{it} can be interpreted as measurement error. Therefore, if one allows non-zero choice probabilities for every topic, the cost of including ε_{itk} is relatively small.

[12] In some models, the researcher is interested in the estimation of individual-level parameters though. E.g., see Allenby and Rossi (1999).

Now I return to the initial problem with unobserved topic attributes $\boldsymbol{\nu}_k$,

$$U(c_{it} = k) = \mathbf{x}'_{it}\boldsymbol{\beta}^o_i + \boldsymbol{\nu}'_k\boldsymbol{\beta}^u_i + \varepsilon_{itk}. \tag{1.5}$$

Provided that the utility function has form 1.5, two serious problems arise. First is a lack of identification.[13] Specifically, the variation in $\boldsymbol{\beta}_i$ alone might not explain why legislator i ever chooses different topics. When the econometrician does not observe *any* topic attributes, all that can explain legislator i's behavior over time, that is, across different choice events, are time-varying characteristics. Yet if these time-varying characteristics are discrete and, *a fortiori*, dummy, the observed part of utility, $\mathbf{x}'_{it}\boldsymbol{\beta}^o_i$, might be constant across different choice events.

Second is the omitted variable bias caused by the interaction between $\mathbf{x}_{it}$ and $\boldsymbol{\nu}_k$. Imagine, for instance, that there is the unobserved "salience" variable, which is a topic attribute, and we examine the coefficient in the "general election" dummy, which is time-specific. Clearly, legislator i's utility of speaking about a salient issue during general election is not the same as at the rest of the time. Consequently, if one tries to estimate model 1.5, she will result in the biased estimate of the general election variable.

To what extent the inability to control for attributes of the alternatives can be viewed as a dead issue? In many economics and marketing applications, the researcher is interested in the estimation of the brand preferences and thus she has the information about the alternatives (products). However, as long as inference about parameters $\boldsymbol{\beta}^o_i$ is important even if it is impossible to estimate $\boldsymbol{\beta}^u_i$, discrete

[13] Athey and Imbens (2007) study a similar issue and derive conditions for the "rationalizability" of the data given a particular specification of utility. However, in their setting they have *some* choice attributes observable by the researcher, while in mine there are *none*.

choice models such as 1.5 may be of interest to the researcher. Below I discuss two major classes of applications in which no attributes of the alternatives can be accounted for.

The first motivation is that in some environments, such as the bill voting or issue selection, the researcher must describe complex choice alternatives in the finite-dimensional attributes space and this task might be of substantial difficulty. At the same time, the assumption that individuals make their choices rationally implies that they can rank alternatives given the information about them.

Second, the measures of attributes of the alternatives might be based on self-reported responses to subjective questions.[14] Assume that the researcher has survey data on attitudes towards two topics, foreign policy and abortion, and each legislator has been asked whether they are politically correct—one binary outcome per topic. The point is that everyone might have her own perception of political correctness and thus this attribute enter utility 1.5 in different ways. The variation in $\boldsymbol{\beta}_i^o$ across individuals, as specified above, does not address this issue because the problem is in $\mathbf{x}_{it}$. Subjectivity of responses induces a measurement error that biases the estimates of $\boldsymbol{\beta}_i^o$, as long as the perception of political correctness correlates with individual-level characteristics, e.g., the party affiliation. Therefore, such attributes of the alternatives are rarely used.

So, the range of possible applications described by utility 1.5 is wide. As I will show, one still can estimate preferences of observed characteristics even in the absence of the choice-specific information $\boldsymbol{\nu}_k$. To that end, I introduce an additional level of heterogeneity to capture the impact of unobserved parameters on the observed ones.

[14]For a rigorous discussion, I refer to Bertrand and Mullainathan (2001).

The idea is to condition on the part of $\mathbf{x}_{it}$ that interacts with $\boldsymbol{\nu}_k$. Assume that $\mathbb{E}[\mathbf{x}_{it} \mid \boldsymbol{\nu}_k] = \mathbf{x}_{it}'\mathbf{\Pi}_{ik}\boldsymbol{\nu}_k$, where $\mathbf{\Pi}_{ik}$ is the $H \times P$ matrix whose element (h,p) denotes the marginal utility of the interaction term between h-th observed individual or time-related characteristic and p-th unobserved attribute of choice k. Then, rewrite utility in the form

$$U(c_{it} = k) = \mathbf{x}_{it}'\boldsymbol{\beta}_{ik} + \varepsilon_{itk}, \tag{1.6}$$

with

$$\begin{aligned} \boldsymbol{\beta}_{ik} &= \boldsymbol{\beta}_i + \mathbf{\Pi}_{ik}\boldsymbol{\nu}_k, \\ \boldsymbol{\beta}_i &\sim \text{Normal}(\boldsymbol{\mu}, \mathbf{\Sigma}). \end{aligned}$$

Mixing distribution of $\boldsymbol{\beta}_i$ is the same as in conventional MML. Note that the constant that enters $\mathbf{x}_{it}$ is also treated as random.[15] The conditional choice probabilities implied by utility 1.6 vary across topics as long as $\boldsymbol{\beta}_{ik}$ do,

$$\Pr\left[c_{it} = k \mid \mathbf{x}_{it}\right] = \sigma\left(\mathbf{x}_{it}'\boldsymbol{\beta}_{ik}\right). \tag{1.7}$$

Further, let $\boldsymbol{\eta}_{ik}$ be the H-vector equal to $\mathbf{\Pi}_{ik}\boldsymbol{\nu}_k$. One can think of $\boldsymbol{\eta}_{ik}$ as capturing the variation in tastes across both legislators and topics. Imagine for a while that we "fix" the variation across choices and only focus on the variation

[15] In a manner, that constant can be viewed as the only "observed" topic-specific variable. To see that, rewrite 1.6 as $U(c_{it} = k) = \alpha_{ik} + \tilde{\mathbf{x}}_{it}'\boldsymbol{\beta}_{ik} + \varepsilon_{itk}$, where $\tilde{\mathbf{x}}_{it}$ is the $H-1$ vector obtained from $\mathbf{x}_{it}$ separating 1.

across individuals. That is, we consider $\boldsymbol{\eta}_{ik} \mid \widetilde{\boldsymbol{\zeta}}_k$, such that

$$\boldsymbol{\eta}_{ik} \mid \widetilde{\boldsymbol{\zeta}}_k \sim \text{Normal}(\widetilde{\boldsymbol{\zeta}}_k, \widetilde{\boldsymbol{\Omega}}_k).$$

Denote with $\boldsymbol{\Omega}_k$ the covariance of $\boldsymbol{\beta}_i + \boldsymbol{\eta}_{ik} \mid \widetilde{\boldsymbol{\zeta}}_k$. Then

$$\boldsymbol{\beta}_{ik} \mid \boldsymbol{\zeta}_k \sim \text{Normal}(\boldsymbol{\zeta}_k, \boldsymbol{\Omega}_k),$$

where vector $\boldsymbol{\zeta}_k = \boldsymbol{\mu} + \widetilde{\boldsymbol{\zeta}}_k$ reflects the preferences of a representative legislator towards topic k and matrix $\boldsymbol{\Omega}_k$ captures correlations between the preferences towards that topic. The conditional probability to speak about topic k is

$$\Pr\left[c_{it} = k \mid \mathbf{x}_{it}, \boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_K, \boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_K\right] = \int \sigma\left((\mathbf{x}_{it} + \boldsymbol{\nu}_k)' \boldsymbol{\beta}_{ik}\right) \,\mathrm{d}F(\boldsymbol{\beta}_i), \quad (1.8)$$

where $F(\boldsymbol{\beta}_i)$ is the joint cumulative distribution $F(\boldsymbol{\beta}_{i1}, \ldots, \boldsymbol{\beta}_{iK} \mid \boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_K)$.

It remains to specify heterogeneity across topics. I do so as follows:

$$\boldsymbol{\zeta}_k \sim \text{Normal}(\boldsymbol{\zeta}_0, \boldsymbol{\Omega}_0).$$

Covariance matrix $\boldsymbol{\Omega}_0$ captures arbitrary correlations between topic choices in the population, while $\boldsymbol{\zeta}_0$ is the vector of utility parameters of a representative legislator on average—across all available topics. Unlike $\boldsymbol{\zeta}_k$ and $\boldsymbol{\Omega}_k$, these parameters are not topic-specific and thus it would be incorrect to refer to them as preferences. To obtain the unconditional probability[16] of choosing alternative k, one needs to

[16] I.e., unconditional on $\boldsymbol{\zeta}_k$ and $\boldsymbol{\Omega}_k$.

integrate out $\boldsymbol{\zeta}_k$ in 1.8,

$$\begin{aligned} \Pr\left[c_{it} = k \mid \mathbf{x}_{it}, \boldsymbol{\zeta}_0, \boldsymbol{\Omega}_0\right] &= \iint \sigma\left((\mathbf{x}_{it} + \boldsymbol{\nu}_k)'\boldsymbol{\beta}_{ik}\right) \mathrm{d}F(\boldsymbol{\beta}_i)\, \mathrm{d}F(\boldsymbol{\zeta}) \\ &= \pi_{itk}, \end{aligned} \tag{1.9}$$

where $F(\boldsymbol{\zeta})$ is the joint cumulative distribution $F(\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_K)$.

My primary goal is to estimate topic-level parameters $\boldsymbol{\zeta}_k$ and $\boldsymbol{\Omega}_k$, not $\boldsymbol{\zeta}_0$ and $\boldsymbol{\Omega}_0$, because a representative legislator's preferences over topics, not her utility parameters on average, are of substantial interest in this paper. As shown in section 1.3, one can approach this task directly without the need to estimate higher-level parameters $\boldsymbol{\zeta}_0$ and $\boldsymbol{\Omega}_0$.

To sum up, heterogeneity in utility parameters $\boldsymbol{\beta}_i$ across individuals has been introduced to deal with the limitations of multinomial logit, such as the IIA property and ignoring the random taste variation. Heterogeneity in utility parameters $\boldsymbol{\beta}_{ik}$, or preferences, across topics, in turn, has been introduced to explain legislators' behavior in the absence of the topic-specific information.

Finally, I would like to review the core assumption I used when deriving the choice probabilities.

Assumption 1.1 (Single-Issue Debate)**.** Every legislator devotes her speech to exactly one topic at a time.

In principle, this assumption is not stark because legislators face time constraints when delivering their speeches in the Congress. However, the problems arise when it comes to aggregation—even within a day legislators are likely to engage in discussions on a variety of issues. Unfortunately, I have to use day as a unit in my empirical application and occasionally issues indeed overlap. In section 3 we will see what problem that causes.

1.2. Topics

In this subsection, I describe the structure of topic choices available to legislators every time period.

Let V be the size of the vocabulary and $w = 1, \ldots, V$ index words in the vocabulary. Each speech document $\mathbf{y}_{it} = (y_{itw})$ is the V-vector of non-negative integers (word frequencies). Let n_{it} be the total number of words in document $\mathbf{y}_{it}$. Denote with $\boldsymbol{\theta}_k = (\theta_{kw}) \in \Delta^{V-1}$ the vector of multinomial probabilities, where Δ^N is the N-dimensional simplex. Its typical element, θ_{kw}, determines the likelihood of word w being used in topic k. One can think of $\boldsymbol{\theta}_k$ as a representative speech document for topic k, with more important words carrying larger weights.

Therefore, in my discrete choice model agents essentially "choose" among the limited number of points on the simplex. Exogeneity of the number of topics, K, is a considerable drawback of the topic modeling. There is no consensus in the literature on how K is determined. The main criterion—how well the model fits the data—allows for a high degree of discretion. In my empirical test, I conclude that 44 topics summarize information in the Congressional Record quite accurately.[17]

It is important to discuss assumptions I have implicitly imposed on the model and observed data.

Assumption 1.2 (Interchangeability)**.** Substantial information about a given topic is only reflected in how often each word occurs in documents related to that topic, while the relative location of these words within documents does not

[17] Indeed, $K \approx 40$ can be viewed as a rule of thumb for such kind of political texts (Grimmer, 2010; Quinn et al., 2010).

matter.

Surely, this is a great simplification of the natural language, yet no sophisticated model can emulate it arbitrarily closely, while approaches based on property 1.2 performed well in various applied settings (e.g., Blei and Lafferty, 2006).

Assumption 1.3 (Stationarity)**.** The pool of available topics remains unchanged over time in the sense that within each topic word frequencies are constant.

This property might not hold even for a short period of time, nor for texts of general interest nor for political ones. As I show in section 3, some of the discovered topics are about the U.S. subprime mortgage crisis that began in 2008. Surely, my model predicts a nearly zero share[18] of that topic for the preceding periods. There is a number of papers (Blei and Lafferty, 2007; Quinn et al., 2010, among others) that address the issue of topic evolution from different perspectives. Yet I try to keep things as simple as possible and ignore the topic dynamics.

1.3. Identification and Estimation

Taking every $\boldsymbol{\theta}_k$ as fixed, I introduce topic assignments,

$$z_{itk} = \begin{cases} 1, & \text{if } \mathbf{y}_{it} \text{ belongs to topic } k \\ 0, & \text{otherwise} \end{cases}$$

As noted above, I assume that individuals choose a single alternative at a time, so that $\mathbf{z}_{it} = (z_{itk})$ is the indicator vector. If actual choices were observed, $z_{itk} = 1$ would mean that in time period t legislator i devoted her speech to topic k.

[18] As I noted above, *exactly* zero shares are impossible due to the presence of the error term, ε_{itk}.

However, the researcher never observes these choices merely because the topics themselves are inferred from the data. How can one break this vicious circle? The point is that the estimation procedure is iterative. At the initial stage, instead of treating choice outcomes $\mathbf{z}_{it}$ as given sequences of zeros and ones, I assume that they are distributed randomly. Specifically, if $\boldsymbol{\pi}_{it} = (\pi_{itk})$ is a vector of legislator i's choice probabilities, as defined in 1.9, then

$$\mathbf{z}_{it} \,|\, \boldsymbol{\pi}_{it} \sim \text{Multinomial}\,(1, \boldsymbol{\pi}_{it}\,)\,. \tag{1.10}$$

That is, we draw each choice outcome from distribution 1.10 and then treat it as a "real" observation when estimating other latent parameters, including $\boldsymbol{\theta}_k$, at the subsequent stages.[19]

It remains to derive the distribution of the *observed* data, $\mathbf{y}_{it}$. If topic k solves legislator i's utility maximization problem at time t,

$$\mathbf{y}_{it} \,|\, (z_{itk} = 1)\,, n_{it}, \boldsymbol{\theta}_k \sim \text{Multinomial}\,(n_{it}, \boldsymbol{\theta}_k\,)\,. \tag{1.11}$$

Notice that relationship 1.11 incorporates all the available information about latent variables $\boldsymbol{\theta}_k$ and $\mathbf{z}_{it}$, as well as higher-level preferences variables $\boldsymbol{\zeta}_k$ and $\boldsymbol{\Omega}_k$. To see that, write down the likelihood,

$$L(\boldsymbol{\zeta}, \boldsymbol{\Omega}, \boldsymbol{\theta}, \mathbf{z} \,|\, \mathbf{X}, \mathbf{Y}\,) = \prod_{i=1}^{I} \prod_{t=1}^{T_i} \prod_{k=1}^{K} \left[\iint \sigma\left(\mathbf{x}_{it}'\boldsymbol{\beta}_{ik}\right)\, \mathrm{d}F(\boldsymbol{\beta}_i)\, \mathrm{d}F(\boldsymbol{\zeta}) \prod_{w=1}^{V} \theta_{kw}^{y_{itw}} \right]^{z_{itk}} . \tag{1.12}$$

[19] Alternatively, one could say that though the econometrician does not observe choice outcome $\mathbf{z}_{it}$ *ex ante*, she learns it *ex post*.

where $\mathbf{X}$ and $\mathbf{Y}$ denote the collections of $\mathbf{x}_{it}$ and $\mathbf{y}_{it}$, respectively.

Train (2003) considers two approaches to the estimation when the likelihood is as complex as 1.12. One is based on the maximum simulated likelihood estimator (MSLE). However, MSLE is intractable, since the number of parameters to be estimated is too large and each integral must be simulated. Furthermore, unlike "conventional" discrete choice models, mine takes choice decisions as latent, so that $\mathbf{z}_{it}$ cannot be identified simultaneously with other parameters.

Therefore, I adopt the alternative, Bayesian, approach to the estimation of $\boldsymbol{\zeta}_k$, $\boldsymbol{\Omega}_k$, $\boldsymbol{\theta}_k$, and $\mathbf{z}_{it}$, that does not require the maximization of 1.12, nor does it require the simulation of integrals. From a Bayesian perspective, all the parameters of interest are treated as random, and this is very convenient in my framework with latent variables. In this setting, $\boldsymbol{\zeta}_k$ and $\boldsymbol{\Omega}_k$ are viewed as hyperparameters of distribution $p(\boldsymbol{\beta}_{ik} \,|\, \boldsymbol{\zeta}_k, \boldsymbol{\Omega}_k)$, common for each individual i. Though in theory one could introduce hyperpriors for $\boldsymbol{\zeta}_k$ and $\boldsymbol{\Omega}_k$.[20], in my context it suffices to estimate these parameters directly, rather than estimate their posterior means[21]

My specification is completed once I assume uninformative symmetric priors,

$$\boldsymbol{\theta}_k \sim \text{Dirichlet}\left(\boldsymbol{\alpha}\right), \tag{1.13}$$

where α is the V-vector with typical element $\alpha_w = 1/V$.

I then estimate the joint posterior distribution of $\boldsymbol{\beta}_{ik}$, $\boldsymbol{\theta}_k$, and $\mathbf{z}_{it}$ and obtain the point estimates for $\boldsymbol{\zeta}_k$ and $\boldsymbol{\Omega}_k$ given $\mathbf{X}$ and $\mathbf{Y}$. I describe the estimation procedure and analysis of convergence in more detail in appendix A.[22]

[20] A hyperprior for $\boldsymbol{\zeta}_k$ would be a Normal($\boldsymbol{\zeta}_0, \boldsymbol{\Omega}_0$) distribution.

[21] In the statistics literature, this method is called empirical Bayes (Robbins, 1956).

[22] The package for the Python/Cython programming languages that implements the EAA algorithm is available upon request.

1.4. Comparison with the Existing Topic Models

Up to how the probabilities of speaking about certain topics are determined, EAA is similar to the "expressed agenda model" (EAM) proposed in Grimmer (2010). What I denoted with π_{itk} in 1.9 in Grimmer (2010) is referred to as legislator i's "expressed agenda". Essentially, EAM can be viewed as a particular case of EAA with the only covariate, a constant. As this constant is fixed across legislators and choice events, EAM accounts for no individual-level and time-specific information and thus cannot predict how the variation in these characteristics affects the choice probabilities. EAM might be useful for analysis of topics, yet inference is limited.

The "structural topic model" (STM) described in Roberts et al. (2013) is more similar to mine in the sense that it can use more available information about how a particular document was generated. Not only this innovation allows to fit the data better—the inclusion of additional variables makes it possible to estimate the quantities of substantive interest, such as the average treatment effects. There are, however, four major differences between EAA and STM, and I argue that in the specific context of this paper EAA is a more appropriate framework.

First and foremost, it is important to emphasize that STM in Roberts et al. (2013) was introduced for different purposes, namely for survey analysis. In EAA, in turn, the derivation of estimates of interest is based on the concept of utility maximization, which allows to model strategic behavior of legislators and provide an economic interpretation of the parameters.

Second, unlike Roberts et al. (2013), I focus on *single-membership* models, with each document belonging to exactly one topic. As pointed out earlier (see assumption 1.1), if speech transcripts are organized properly—each document corresponds to one speech—then single membership is not a severe limitation. Importantly, both economic and econometric modeling is easier and more intuitive once one allows legislators to deliver their speeches on one topic at a time. On the contrary, careful consideration is required to justify the use of *mixed-membership* models, with each document being a mixture of topics, from the economic perspective.[23]

Third, when specifying the form of utility in 1.6, I use the random coefficients framework. In contrast, in its definition of the choice probabilities[24], STM exploits the fixed coefficients framework. As I have shown, random coefficients allow to get rid of concerns related to the IIA property and random taste variation. One implication is that EAA accounts for the correlation between topic choices by the same legislator over time. But the main advantage is that capturing heterogeneity both across legislators and topics, EAA can explain legislators' behavior even when the topic-specific information is entirely unobserved for the researcher and, under some assumptions, yields consistent estimates of the parameters. STM treats the coefficients as fixed for the whole population, which can lead to bias.[25]

Fourth, STM allows for the variation in the word use within the same topic across different groups, while EAA does not. In my model, the choice set is fixed, and topics are completely defined by their (unobserved) attributes. In practice

[23] One possible motivation comes as a solution of the data aggregation problem. In particular, the researcher can consider each speaker-event observation as a separate "market" (in the industrial organization tradition; see Berry et al., 1995), with event being a day, a week, etc. Alternatively, mixed membership could accommodate the multiple discrete choice problem in the spirit of Hendel (1999), when legislators are allowed to speak about more than one topic at a time. Existing mixed-membership models, however, cannot address either possibility, and I leave them beyond the scope of this paper.

[24] Strictly speaking, STM does not use the notion of discrete choice. Yet since the topic shares, or *prevalence*, in STM and the choice probabilities in EAA are determined similarly, for consistency I refer to the topic shares as to the choice probabilities.

[25] Though an additional level of heterogeneity (across topics) might not be necessary to explain the variation in survey responses, the inability to control for the unobserved topic-specific variables nevertheless would cause bias.

it means that, for example, Republicans and Democrats use the same vocabulary when speaking about foreign policy. As studies show (e.g., Monroe et al., 2008), the difference in how both parties debate on the same issue is indeed significant. The discrete choice approach cannot capture this feature of political debates, since it represents legislators as the "demand side", while the variation in the word use occurs on the "supply side".

To summarize, both STM and EAA have their pros and cons. Though STM is somewhat more robust, tends to fit the data better, and applicable to a wider range of problems (such as survey experiments), EAA is more analytically rigorous and appropriate for well-defined economic problems.

2. The Data

2.1. Speech Transcripts

My dataset includes transcripts of speeches from the Congressional Record, 2007–08 (110th Congress). I use the data provided by the independent website, GovTrack.us.[26] Though it does not perfectly cover the period of 2007–08[27], each document in this collection has a semantic markup, so that speaker-to-speech mappings are easily generated. Using the GovTrack version of the Congressional Record, I group each single observation, i.e., document, by legislator and time (day). There are in total 530 speakers, both senators and representatives, and 15,418 speaker-

[26]Unfortunately, the bulk collection of documents—"as is"—delivered by the Library of Congress via the THOMAS system does not meet my needs. Specifically, it is not possible to identify speakers within each document and thus to create the speaker-to-speech mapping.

[27]The structure of the Congressional Record is not adopted for the machine processing, hence parsing, that is, extracting valuable information, is very difficult. However, one can assume that all the missings are random.

to-speech observations.

I then apply several preprocessing procedures that are standard in the automated text analysis tasks (e.g., Grimmer and Stewart, 2013). First, I discard the word ordering and punctuation and convert every word in documents to lowercase. Second, in each document I only keep nouns, adjectives, verbs, and adverbs.[28] Third, each word in my dataset is replaced by its stem (Porter, 1980), so that words with similar meanings are treated as identical, e.g., different words countries and country turn into a single stem countri. After these steps, we result in the "bag-of-words" representation of our corpora[29], in which every document is a vector of stem counts.

Finally, I remove all extreme words[30], that is, those that occur in more than 94% and less than 0.5% documents in corpora. The rationale for that is as follows. On the one hand, it allows us to reduce the dimensionality of the problem. On the other, as LDA and, in turn, EAA tend to identify topics by exploiting the variation in the word use, the prevalence of extreme words leads either to weak identification—resulting clusters are too similar—or to identification of marginal topics.

Ultimately, my vocabulary consists of 2,961 words and the whole Congressional Record is represented by the $15,418 \times 2,961$ matrix $\mathbf{Y}$, whose typical element y_{dw} is the number of word w occurrences in document d.

[28]The pattern.en package (Smedt and Daelemans, 2012) significantly simplifies that task.
[29]In linguistics, *corpora* stands for the collection of texts.
[30]From now on, I will refer to stems as words.

2.2. Covariates

As noted earlier, in the EAA model each document is associated with a set of covariates. In my empirical test, I use the following specification,

1. Chamber dummy, with 1 for the Senate.
2. Party dummy, with 1 for the Republican Party.
3. Ideology, from 0 to 1, where 0 is moderate and 1 is extreme.
4. Primary election dummy.
5. General election dummy.

I use the absolute values of the DW-NOMINATE scores (Poole and Rosenthal, 1997) as a proxy for ideology.[31] These scores are based on the congressional roll-calls. Though not being perfect, they have become a common measure of ideology in the political science literature.

Note that we need to map a continuous electoral process into the dummy variable outcome. Though this task could be solved differently, I consider it as a problem of choosing an appropriate bandwidth—a point in time when one state of the world (primary/general/none) switches to another. For empirical purposes, I put

- the primary election dummy equal 1 if there remain less than 60 days prior to primary election and
- the general election dummy equal 1 if the primary election is held and there are less than 180 days prior to general election.[32]

[31] The original DW-NOMINATE scores are on the $[-1, 1]$ scale, where -1 and 1 stand for extreme liberal and extreme conservative, respectively. So, the absolute values of these scores measure the deviation from the moderate position.

[32] I account for the state differences, runoffs, and periodicity of the Senate elections.

In principle, one could also include the state- and district-level information in order to control for the differences in electoral competition across states and districts, e.g., "safety" for a certain party. However, as I only focus on incumbents, i.e., those who already serve in the Congress, omitting such variables should not significantly affect my estimates. This is because once the legislator is in office, she is likely to be of high quality. Hirano and Snyder (2012) present evidence that selection of the good-type politicians takes place even in "safe" congressional districts due to the intra-party competition in primaries. Therefore, when studying incentives of politicians seeking re-election, it is natural to assume that their decisions are influenced by the competition per se, not by specific environments of constituents they represent.

3. Results

3.1. General Background

Application of the EAA algorithm to the Congressional Record data generates the following topic classification (see tables B.1 and B.2). Each topic is represented by most frequent words, with frequencies defined as estimates $\widehat{\boldsymbol{\theta}}_k$. Prior to assigning a label to each topic, I review 150 most frequent topic-specific words and randomly draw up to 100 original speech transcripts that belong to that topic. Around eleven topics, labeled by the "Misc.", have no clear economic interpretation in the sense that they tend to absorb multiple separate issues. For instance, topic 11 includes elements of issues "Agriculture" and "Energy". Though the presence of the "Misc." topics hampers the analysis, it turns out that this problem cannot be solved by simply increasing the number of topics. Recall assumption 1.1 which states that just one topic is discussed at a time—due to the data aggregation problem, some documents in principle cannot belong to a single topic, as long as one requires that topic to be meaningful.

The remaining topics cover a variety of issues discussed in the Congress, from the immigration policy to transportation. Note that there is at least one topic about the subprime mortgage crisis, what violates assumption 1.3 about the stationarity of topics. Approximately ten topics focus on the Iraq War. Though that number is relatively large, these topics highlight different aspects of the conflict, such as funding, the veterans health care, terrorist activity, etc.

In what follows, I analyze the divisive issues only, that is, those on which both parties almost surely take positions far from the median voter's. I have not modeled which issues are divisive and which are not, though. Once the researcher obtains a classification of speech transcripts, she can conclude which topics describe divisive issues and infer how, on average, observed individual- and time-specific characteristics affect legislators' decisions to speak about thee issues.

3.2. Divisive Issues

Among all the discovered topics I pick those that describe the following issues: energy and environment, immigration, stem cell research, war on terrorism and domestic security, and the Iraq War. Below I discuss legislators' preferences towards these issues. Each discussion is accompanied by an excerpt from a speech transcript related to the corresponding topic.

In tables B.3 to B.13 the first row corresponds to the estimate $\widehat{\boldsymbol{\zeta}}_k$ and the correlation matrix is derived from the estimate $\widehat{\boldsymbol{\Sigma}}_k$. Note that I only report on those topics that are represented by at least 100 documents.

3.2.1. Energy and Environment

> Unfortunately, it seems as though every time we bring up the issue of more domestic supply, our friends on the other side of the aisle, who control the floor and control the agenda by virtue of their being in the majority, have simply said: No. No. Unfortunately, no new energy continues to mean higher prices for the American consumer. [...] We provide a comprehensive solution by saying yes to domestic oil supply, using what God has given us in this country in a way that will allow us to be less dependent on imported oil from the Middle East. (Sen. Cornyn, 07/22/2008)

The issues of energy and environment are represented by topics 31, 32, and 34 (tables B.3 to B.5). In this subsection, I analyze topic 32 in more detail.

First let me examine the *Primary* dummy. Though its sign in table B.4 is positive, it is near zero. This implies that a representative legislator does not engage in the debate on renewable energy more frequently during primary elections than usual.[33] However, the coefficient in the *General* dummy is large and negative. This means that a representative legislator tends to speak about alternative energy less frequently during general elections than usual.

Now I examine the *Ideology* variable. The negative sign in table B.4 means that on average, independent of electoral cycle, more extreme legislators engage in the debate on alternative energy less frequently. As the correlation between *Ideology* and *Primary* is negative, more extreme legislators also tend to speak about alternative energy less during primary elections. However, the opposite is true during general elections, since the correlation between *Ideology* and *General*

[33] Recall that the absence of electoral competition is a reference level.

is positive, though relatively small. We can say that, on average, during general elections more extreme candidates substitute speaking about environmental issues for speaking about other issues.

So, more extreme politicians prefer engaging in the debate about alternative energy during general elections regardless of the fact that in the whole population, that issue is not "popular" during that period. Therefore, if the issue of energy and environment is indeed divisive, then these are extreme candidates who take extreme positions more frequently than moderate candidates. This finding suggests that during general elections, the policy divergence between parties along the dimension of alternative energy can be associated with political activity of more extreme candidates.

3.2.2. Immigration

> [The AgJOBS Act] will also enhance national security and reduce illegal immigration. It will reduce the chaotic, illegal, and all-too-deadly flows of immigrants at our borders by providing safe and legal avenues for farm workers and their families. Future temporary workers will be carefully screened to meet security concerns. Enforcement resources will be more effectively focused on the highest risks. By bringing undocumented farm workers out of the shadows and requiring them to pass through security checks, it will enable officials to concentrate more effectively on terrorists and criminals.
>
> (Sen. Kennedy, 01/10/2007)

In my analysis, the issue of immigration is covered by topic 2 only (table B.6).

Notice first that both *Primary* and *General* have large negative signs. This means that a representative legislator tends to speak about immigration during elections less frequently. At the same time, the *Ideology* variable is large and positive. That is, more extreme politicians discuss this topic more frequently, independently of electoral cycle. However, during both primary and general elections extreme candidates tend not to participate in the debate on immigration—this is because both correlations between *Ideology* and the electoral dummies are negative.

Therefore, if the immigration policy is a divisive issue, then the divergence between Democrats and Republicans during elections, if any, cannot be linked to the ideology of candidates.

3.2.3. Stem Cell Research

> First, it [stem cell research] has an incredible propensity to morph into tumors. Secondly, if embryonic stem cells are ever successful and transplanted into humans, embryonic stem cells carry an enormous proclivity for rejection. And third, embryonic stem cell research requires the killing of human embryos. If it ever worked, the limited supply of so-called spare embryos, and that's a very offensive word, let me just say. Those children who have been adopted from cryogenic tanks—snowflake babies—are a witness against this idea of saying somehow there's a spare embryo. But just take that for what it is. If it ever worked, there would be a near insatiable demand for freshly killed human embryos. [...] That is a brave new world. This is the tip of the iceberg today, and hopefully we will not go that way. We must do ethical stem cell research instead. (Rep. Smith, 06/07/2007)

This issue is captured by topic 36 (see table B.7).

Due to the negative sign of *Primary*, on average, this moral issue seems to attract less attention during primary elections than in the absence of the competition. On the contrary, during general elections this issue becomes relatively more popular. Though the sign of *Ideology* is large and positive, more extreme legislators substitute this issue for others only during general elections—this is why the correlation between *Ideology* and *Primary* is small and negative and the correlation between

Ideology and *General* is positive.

Hence this finding suggests that at least partially the policy divergence during general elections may be associated with the inflow of more extreme candidates.

3.2.4. War on Terror / Domestic Security

> [The American people] want us to fight terrorism, and we intend to fight terrorism; but we intend to do it with a greater focus on those who attacked us on 9/11, with a greater focus on homeland security, on making sure that we are keeping nuclear weapons out of the hands of terrorists. Perhaps the greatest threat we face, which went by the boards because of this administration's preoccupation with fighting the wrong war in Iraq which has diverted us from really focusing on the concentrated effort we need from law enforcement, from intelligence, from military, from diplomacy, from the soft power that America, has been extending our cultural ideals and principles out into the world to show people that we are not merely going to bully people with weapons, but we are also going to stand on our ideals and principles. (Rep. Hodes, 04/19/2007)

Two topics, 18 and 24, represent this issue (B.8 and B.9).

Consider, e.g., topic 18. Interestingly, both *Primary* and *General* are positive. The correlation between *Ideology* and *Primary* is positive, what implies that during primaries more extreme candidates tend to speak about domestic security more frequently. The correlation between *Ideology* and *General* is negligibly small.

This evidence suggests the following. If the issue of domestic security is divisive, two parties diverge in the policies during general elections not because of the prevalence of extreme candidates but due to some other reason. Most likely, this is the electoral competition that shifts the policies from the median. However, without introducing voters' preferences it is impossible to say whether elections have a "distortionary" effect on the social welfare. As pointed out in the introduction, in some cases the divergence might be welfare-maximizing.

3.2.5. Iraq War

> We have not been attacked here at home for almost 7 years—a direct result of the strategy of getting on the offense and pushing back against those who would attack us here at home. [...] Precipitous withdrawal we know would lead to a new haven for terrorists with the opportunity to attack us here at home. [...] The American people have this on their minds, obviously. They also have on their minds the economy, health care, and other matters. They are interested in their future. I think the American people are not interested in having additional attacks on the homeland in the future.
>
> (Sen. McConnell, 04/09/2008)

Four topics, 3, 30, 39, and 42 relate to this issue (tables B.10 through B.13).

I will consider topic 30 as an example. The small correlation between *Ideology* and *General* implies that, on average, during general elections more extreme politicians do not prefer the debate on the Iraq War over the debate on other issues. As for the policy divergence, the implications for that issue are the same as in the previous subsection.

4. Discussion

To sum up, there are different patterns for different divisive issues. If one focuses on general elections, divergence on the issues like immigration, domestic security, and the Iraq War seems to be unrelated to the effect of ideology. On the other hand, more extreme candidates are more likely to run their campaigns on issues like stem cell research and alternative energy than other issues, including moderate ones (on which there may be convergence). As discussed earlier, welfare implications crucially depend on the distribution of voters' preferences.

Interestingly, all the topics that we consider as examples indicate that, on average, those divisive issues are relatively more "popular" among Republicans than Democrats.[34] The same applies to the members of the Senate.

Evidence presented in the previous section, however, is not conclusive for a number of reasons. The issues highlighted below comprise the direction of future research.

First, the collection of divisive issues considered in the paper is somewhat arbitrary and there need to be more empirical evidence to support the point that both parties indeed diverge on these issues.

Second, as already noted, my assumptions regarding the distribution of topics over time do not always hold. Also, the data aggregation problem generates a number of topics that cannot be identified and analyzed.

Third, I assumed no dynamics in the decision-making, that is, in each time period legislators choose topics independently from their past choices. Yet it is unlikely that they behave in such a way. If in time period t legislator i chooses topic k conditionally on the past, then the error terms ε_{itk} are not independently distributed and, subsequently, the estimates are no longer consistent.

Fourth, some issues are represented by more than one topic and the estimates obtained for topics representing the same issue are sometimes different. Therefore, prior to making conclusions about platform divergence, one needs to differentiate between any two topics defining the same issue.

In principle, one could report only on those topics whose preference estimates are statistically significant. However, I can say nothing about the statistical sig-

[34]See the positive signs in the *Party* variable in tables B.4, B.6, B.7, B.8, and B.11.

nificance of the parameter estimates, $\widehat{\boldsymbol{\zeta}}_k$ and $\widehat{\boldsymbol{\Omega}}_k$, as their asymptotic distribution is unknown. Unfortunately, the computation of the bootstrapped standard errors is not a tractable solution because even a single iteration of the EAA algorithm is time-consuming.

5. Conclusion

In this paper I have introduced the endogenous attention allocation model in order to analyze how selection of divisive issues depends on the ideology of candidates and electoral incentives. For a given collection of texts, this model simultaneously discovers what topics are discussed in the day-to-day debate and estimates how observed individual- and time-specific characteristics affect the probability to speak about these topics. Additionally, my model accounts for heterogeneity in preferences towards different topics in the population, what solves the problem of unobserved topic attributes.

I have applied this model to the Congressional Record data. My findings indicate that during general elections, divergence on the issues like immigration, domestic security, and the Iraq War is not associated with the ideology of candidates. At the same time, there is evidence that more extreme politicians engage in the debate on such issues as stem cell research and alternative energy more frequently than moderate ones.

References

Abramowitz, A. I. and K. L. Saunders (2008). Is Polarization a Myth? *The Journal of Politics 70*(2), 542–555.

Allenby, G. M. and P. E. Rossi (1999). Marketing Models of Consumer Heterogeneity. *Journal of Econometrics 89*, 57–78.

Athey, S. and G. W. Imbens (2007). Discrete Choice Models with Multiple Unobserved Choice Characteristics. *International Economic Review 48*(4), 1159–1191.

Bafumi, J. and M. C. Herron (2010). Leapfrog Representation and Extremism: A Study of American Voters and Their Members in Congress. *American Political Science Review 104*(3), 519–542.

Bernhardt, D., J. Duggan, and F. Squintani (2009). The Case for Responsible Parties. *American Political Science Review 103*(4), 570–587.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile Prices in Market Equilibrium. *Econometrica 63*(4), 841–890.

Bertrand, M. and S. Mullainathan (2001). Do People Mean What They Say? Implications for Subjective Survey Data. *The American Economic Review 91*(2), 67–72.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* New York: Springer.

Blei, D. M. and J. D. Lafferty (2006). Dynamic Topic Models. In *Proceedings of the 23rd International Conference on Machine Learning*, pp. 113–120.

Blei, D. M. and J. D. Lafferty (2007). A Correlated Topic Model of Science. *The Annals of Applied Statistics*, 17–35.

Blei, D. M., A. Ng, and M. Jordan (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research 3*, 993–1022.

Braun, M. and J. McAuliffe (2010). Variational Inference for Large-Scale Models of Discrete Choice. *Journal of the American Statistical Association 105*(489), 324–335.

Downs, A. (1957). *An Economic Theory of Democracy.* New York: Harper Press.

Egorov, G. (2012). Single-Issue Campaigns and Multidimensional Politics.

Fiorina, M. P. (1999). Whatever Happened to the Median Voter? Stanford University mimeo.

Glaeser, E. L., G. A. M. Ponzetto, and J. M. Shapiro (2005). Strategic Extremism: Why Republicans and Democrats Divide on Religious Values. *The Quarterly Journal of Economics 120*(4), 1283–1330.

Grimmer, J. (2010). A Bayesian Hierarchial Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases. *Political Analysis* (1), 1–35.

Grimmer, J. and B. Stewart (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 1–31.

Hendel, I. (1999). Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns. *Review of Economic Studies 66*, 423–446.

Hirano, S. and J. M. Snyder (2012). Primary Elections and the Quality of Elected Officials.

Hopkins, D. J. and G. King (2010). A Method of Automated Nonparametric Content Analysis for Social Science. *American Journal of Political Science 54*(1), 229–247.

Hummel, P. (2010). Flip-flopping from Primaries to General Elections. *Journal of Public Economics 94*, 1020–1027.

Hummel, P. (2013). Candidate Strategies in Primaries and General Elections with Candidates of Heterogeneous Quality. *Games and Economic Behavior 78*, 85–102.

Jordan, M. I., Z. Ghahramani, T. S. Jaakkola, and L. K. Saul (1998). An Introduction to Variational Methods for Graphical Models. *Machine Learning 37*(2), 183–233.

Kamada, Y. and F. Kojima (2013). Voter Preferences, Polarization, and Electoral Policies.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*. New York: Academic Press.

McFadden, D. and K. Train (2000). Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics 15*(5), 447–470.

Monroe, B. L., M. P. Colaresi, and K. M. Quinn (2008). Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict. *Political Analysis 16*, 372–403.

Poole, K. T. and H. Rosenthal (1997). *Congress: A Political-Economic History of Roll Call Voting*. Oxford University Press.

Porter, M. F. (1980). An Algorithm for Suffix Stripping. *Program: Electronic Library and Information Systems 14*(3), 130–137.

Quinn, K. M., B. L. Monroe, M. Colaresi, M. H. Crespin, and D. R. Radev (2010). How to Analyze Political Attention with Minimal Assumptions and Costs. *American Journal of Political Science 54*(1), 209–228.

Robbins, H. (1956). An Empirical Bayes Approach to Statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, pp. 157–163.

Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. Gadarian, B. Albertson, and D. Rand (2013). Topic Models for Open Ended Survey Responses with Applications to Experiments.

Smedt, T. D. and W. Daelemans (2012). Pattern for Python. *Journal of Machine Learning Research*, 2031–2035.

Snyder, J. M. and T. Groseclose (2000). Estimating Party Influence in the Congressional Roll-Call Voting. *American Journal of Political Science 44*(2), 193–211.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Wang, C. and D. M. Blei (2013). Variational Inference in Nonconjugate Models. *Journal of Machine Learning Research*.

Weelden, R. V. (2013a). Candidates, Credibility, and Re-Election Incentives. *Review of Economic Studies*, 1–30.

Weelden, R. V. (2013b). Moderate Voters, Polarized Parties.

Appendices

A. Variational Inference

Now I will describe the estimation algorithm. My goal is to uncover the posterior

$$p(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z} \,|\, \mathbf{X}, \mathbf{Y}\,) \propto \prod_{i=1}^{I}\prod_{k=1}^{K} p(\boldsymbol{\beta}_{ik}) \times \prod_{k=1}^{K} p(\boldsymbol{\theta}_k) \times \prod_{i=1}^{I}\prod_{t=1}^{T_i}\prod_{k=1}^{K}\left[\sigma\left(\mathbf{x}'_{it}\boldsymbol{\beta}_{ik}\right)\prod_{w=1}^{V}\theta_{kw}^{y_{itw}}\right]^{z_{itk}}, \tag{A.1}$$

that is, we wish to obtain posterior distributions of the unknown parameters. Alike many sophisticated models, mine does not allow exact inference, since one cannot integrate out the unknowns in A.1. The Markov Chain Monte Carlo methods (MCMC), widely used in the literature, do not apply to my case, as they require too much draws, provided the dimensionality of the problem (Braun and McAuliffe, 2010). Hence I use an alternative technique, variational approximations developed in Jordan et al. (1998), in order to obtain the estimates of posterior distributions of $\boldsymbol{\beta}, \boldsymbol{\theta}$, and $\mathbf{z}$.

Notice that although I did not include parameters $\boldsymbol{\zeta}$ and $\boldsymbol{\Omega}$ in A.1 explicitly, the posterior is computed for the fixed values of $\boldsymbol{\zeta}$ and $\boldsymbol{\Omega}$. Below I will show to obtain their empirical Bayes estimates.

A.1. Minimizing the KL Divergence

I seek an approximating distribution $q(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z})$ which minimizes the *Kullback-Leibler (KL) divergence*[35],

$$\begin{aligned}&\min_q \mathrm{KL}(q(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z}) \parallel p(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z} \mid \mathbf{X}, \mathbf{Y})) \\ &= -\iiint q(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z}) \log \frac{p(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z} \mid \mathbf{X}, \mathbf{Y})}{q(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z})} \, \mathrm{d}\boldsymbol{\beta} \, \mathrm{d}\boldsymbol{\theta} \, \mathrm{d}\mathbf{z}.\end{aligned} \tag{A.2}$$

However, solving the problem A.2 directly is challenging. Instead, I maximize the *evidence lower bound* (ELBO) which is an equivalent problem (Bishop, 2006). To obtain it, notice first that the marginal log-probability of the observed variables, $\mathbf{X}$ and $\mathbf{Y}$[36], can be expressed as follows:

$$\log p(\mathbf{X}, \mathbf{Y}) = \log \iiint p(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z}, \mathbf{X}, \mathbf{Y}) \, \mathrm{d}\boldsymbol{\beta} \, \mathrm{d}\boldsymbol{\theta} \, \mathrm{d}\mathbf{z}.$$

It can be shown (Bishop, 2006) that

$$\log p(\mathbf{X}, \mathbf{Y}) = \mathcal{L}(q) + \mathrm{KL}(q \parallel p), \tag{A.3}$$

where

$$\mathcal{L}(q) = \iiint q(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z}) \log \frac{p(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z}, \mathbf{X}, \mathbf{Y})}{q(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z})} \, \mathrm{d}\boldsymbol{\beta} \, \mathrm{d}\boldsymbol{\theta} \, \mathrm{d}\mathbf{z} \tag{A.4}$$

is the evidence lower bound. As we can see from equation A.3, $\log p(\mathbf{X}, \mathbf{Y})$ is fixed for any q, while the $\mathrm{KL}(q \parallel p)$ term in the right-hand side is always nonnega-

[35] Derivations in the section A.1 are standard in the computer science and applied literature (e.g., Blei et al., 2003; Blei and Lafferty, 2007; Grimmer, 2010).

[36] $p(\mathbf{X}, \mathbf{Y})$ is also called "evidence".

tive. Therefore, as $\mathrm{KL}(q \parallel p)$ decreases, $\mathcal{L}(q)$ must increase. This implies that maximizing the expression in A.4 is equivalent to solving problem A.2. The only assumption I impose on the approximate distribution is that it belongs to the class of factorized distributions,

$$q\left(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z}\right) = \prod_{i=1}^{I} \prod_{k=1}^{K} q(\boldsymbol{\beta}_{ik}) \prod_{k=1}^{K} q\left(\boldsymbol{\theta}_{k}\right) \prod_{i=1}^{I} \prod_{t=1}^{T_i} q\left(\mathbf{z}_{it}\right). \tag{A.5}$$

I describe the iterative optimization in sections A.2 to A.4. In order to assess the convergence of my algorithm, at each step we check if the relative change of $\mathcal{L}(q)$ is less than $\varepsilon_0 > 0$. Unfortunately, as in Blei and Lafferty (2007) and Braun and McAuliffe (2010), the objective for maximizing the evidence lower bound does not have a closed form. In section A.3, I derive the surrogate evidence lower bound function.

A.2. Updating Steps in the Variational E-Step

The challenge is that $p(\boldsymbol{\beta}_{ik})$, a multivariate normal distribution, is not conjugate to $p\left(\mathbf{z}_{it} \,|\, \boldsymbol{\beta}\right)$, which is discrete. This complicates updating steps in variational inference.

Assume that $q(\boldsymbol{\beta}_{ik})$, $q(\boldsymbol{\theta}_k)$, and $q(\mathbf{z}_{it})$ are initialized randomly. Bishop (2006) derives the following optimal conditions,

$$\begin{aligned}
\log q\left(\boldsymbol{\beta}_{ik}\right) &\propto \mathbb{E}_{\boldsymbol{\theta},\mathbf{z}}\left[\log p(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z}, \mathbf{X}, \mathbf{Y})\right], \\
\log q\left(\boldsymbol{\theta}_{k}\right) &\propto \mathbb{E}_{\boldsymbol{\beta},\mathbf{z}}\left[\log p(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z}, \mathbf{X}, \mathbf{Y})\right], \\
\log q\left(\mathbf{z}_{it}\right) &\propto \mathbb{E}_{\boldsymbol{\beta},\boldsymbol{\theta}}\left[\log p(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z}, \mathbf{X}, \mathbf{Y})\right],
\end{aligned} \tag{A.6}$$

where expectations are taken with respect to corresponding variational distributions. Further, it is easy to see that

$$p\left(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z} \,|\, \mathbf{X}, \mathbf{Y}\right) = \exp\left(\log p\left(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z} \,|\, \mathbf{X}, \mathbf{Y}\right)\right) \propto \exp\left(\log p(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z}, \mathbf{X}, \mathbf{Y})\right). \quad \text{(A.7)}$$

Thus in A.6 we use the log marginal posterior,

$$\begin{aligned} \log p\left(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{z} \,|\, \mathbf{X}, \mathbf{Y}\right) &= const + \sum_{i=1}^{I}\sum_{k=1}^{K} \log p(\boldsymbol{\beta}_{ik}) + \sum_{k=1}^{K} \log p(\boldsymbol{\theta}_k) \\ &+ \sum_{i=1}^{I}\sum_{t=1}^{T_i}\sum_{k=1}^{K} z_{itk} \left[\log \sigma\left(\mathbf{x}_{it}'\boldsymbol{\beta}_{ik}\right) + \sum_{w=1}^{V} y_{itw} \log \theta_{kw}\right]. \end{aligned} \quad \text{(A.8)}$$

A.2.1. Update for $q\left(\boldsymbol{\theta}_k\right)$

From A.6 and A.8 we obtain,

$$\log q\left(\boldsymbol{\theta}_k\right) = const + \sum_{w=1}^{V} \log \theta_{kw} \left[\alpha_{kw} - 1 + \sum_{i=1}^{I}\sum_{t=1}^{T_i} \mathbb{E}_{\mathbf{z}}[z_{itk}] y_{itw}\right]. \quad \text{(A.9)}$$

Here the constant term absorbs the components that do not depend on $\boldsymbol{\theta}_k$ directly. Exponentiating the both sides of A.9 we conclude that $q\left(\boldsymbol{\theta}_k\right)$ is a Dirichlet $\left(\boldsymbol{\phi}_k\right)$ distribution with a typical parameter,

$$\phi_{kw} = \alpha_{kw} + \sum_{i=1}^{I}\sum_{t=1}^{T_i} \mathbb{E}_{\mathbf{z}}[z_{itk}] y_{ijw}. \quad \text{(A.10)}$$

I will compute $\mathbb{E}_{\mathbf{z}}[z_{itk}]$ after deriving the form of $q\left(\mathbf{z}_{it}\right)$.

A.2.2. Update for $q\left(\boldsymbol{\beta}_{ik}\right)$

This one has no closed-form solution. I use Laplace variational approximation developed in Wang and Blei (2013). It is flexible and proved to perform better than *ad hoc* solutions.

Equations A.6 and A.8 imply,

$$\log q(\boldsymbol{\beta}_{ik}) = const + \underbrace{\log p(\boldsymbol{\beta}_{ik}) + \sum_{t=1}^{T_i} \mathbb{E}_{\mathbf{z}}[z_{itk}]\mathbb{E}_{-\boldsymbol{\beta}_k}\left[\log \sigma\left(\mathbf{x}'_{it}\boldsymbol{\beta}_{ik}\right)\right]}_{f(\boldsymbol{\beta}_{ik})}, \tag{A.11}$$

where the $-\boldsymbol{\beta}_k$ subscript means that the expectation is taken with respect to all the variables $\boldsymbol{\beta}_{il}, l = 1, \ldots, K$, except $\boldsymbol{\beta}_{ik}$.[37] Recall that the log-density of $\boldsymbol{\beta}_{ik}$ equals,

$$\log p(\boldsymbol{\beta}_{ik}) = -\frac{1}{2}\left[H \log 2\pi + \log|\boldsymbol{\Omega}_k| + (\boldsymbol{\beta}_{ik} - \boldsymbol{\zeta}_k)'\boldsymbol{\Omega}_k^{-1}(\boldsymbol{\beta}_{ik} - \boldsymbol{\zeta}_k)\right].$$

Here I treat the higher-level parameters, $\boldsymbol{\zeta}_k$ and $\boldsymbol{\Omega}_k$, as fixed. Then, consider the second-order Taylor expansion of $\log q(\boldsymbol{\beta}_{ik})$ around the maximum *a posteriori* estimate (MAP), $\widehat{\boldsymbol{\mu}}_{ik}$,

$$\log q(\boldsymbol{\beta}_{ik}) \approx f(\widehat{\boldsymbol{\mu}}_{ik}) + \frac{1}{2}(\boldsymbol{\beta}_{ik} - \widehat{\boldsymbol{\mu}}_{ik})'\nabla^2 f(\widehat{\boldsymbol{\mu}}_{ik})(\boldsymbol{\beta}_{ik} - \widehat{\boldsymbol{\mu}}_{ik}). \tag{A.12}$$

Note that in A.12 $\nabla f(\widehat{\boldsymbol{\mu}}_{ik})(\boldsymbol{\beta}_{ik} - \widehat{\boldsymbol{\mu}}_{ik})$ is zero due to the first-order conditions.

[37] More precisely, I should have written the subscript as $-\boldsymbol{\beta}_{ik}$, but the expression $\log \sigma\left(\mathbf{x}'_{it}\boldsymbol{\beta}_{ik}\right)$ is independent of $\boldsymbol{\beta}_{i'k}$ for all $i' \neq i$.

Exponentiating the both sides gives us,

$$q(\boldsymbol{\beta}_{ik}) \approx \text{Normal}\left(\boldsymbol{\mu}_{ik}, \boldsymbol{\Sigma}_{ik}\right), \tag{A.13}$$

where

$$\begin{aligned}
\boldsymbol{\mu}_{ik} &= \widehat{\boldsymbol{\mu}}_{ik}, \\
\boldsymbol{\Sigma}_{ik} &= -\left(\nabla^2 f(\widehat{\boldsymbol{\mu}}_{ik})\right)^{-1}, \\
\widehat{\boldsymbol{\mu}}_{ik} &= \arg\max_{\boldsymbol{\beta}_{ik}} f(\boldsymbol{\beta}_{ik}).
\end{aligned} \tag{A.14}$$

Approximate calculation of $\mathbb{E}_{-\boldsymbol{\beta}_k}\left[\log\sigma\left(\mathbf{x}_{it}'\boldsymbol{\beta}_{ik}\right)\right]$ is based on the application of Jensen's inequality.[38] First, rewrite the expression as

$$\mathbb{E}_{-\boldsymbol{\beta}_k}\left[\log\sigma\left(\mathbf{x}_{it}'\boldsymbol{\beta}_{ik}\right)\right] = \mathbf{x}_{it}'\boldsymbol{\mu}_{ik} - \mathbb{E}_{-\boldsymbol{\beta}_k}\left[\log\left(\sum_{l=1}^{K}\exp\left(\mathbf{x}_{it}'\boldsymbol{\beta}_{il}\right)\right)\right].$$

Then we result in the following upper bound for the expected log-sum,

$$\begin{aligned}
&\mathbb{E}_{-\boldsymbol{\beta}_k}\left[\log\left(\sum_{l=1}^{K}\exp\left(\mathbf{x}_{it}'\boldsymbol{\beta}_{il}\right)\right)\right] \\
&\leq \log\left(\exp\left(\mathbf{x}_{it}'\boldsymbol{\beta}_{ik}\right) + \sum_{\substack{l=1 \\ l\neq k}}^{K}\exp\left(\mathbf{x}_{it}'\boldsymbol{\mu}_{il} + \frac{1}{2}\mathbf{x}_{it}'\boldsymbol{\Sigma}_{il}\mathbf{x}_{it}\right)\right).
\end{aligned} \tag{A.15}$$

Here I used the property of a log-normal distribution. Indeed, if $\boldsymbol{\beta}_{ik}$ is approximately normal with mean $\boldsymbol{\mu}_{ik}$ and variance $\boldsymbol{\Sigma}_{ik}$, then $\exp(\boldsymbol{\beta}_{ik})$ is log-normal and

[38] Braun and McAuliffe (2010) use an alternative approximation called the multivariate delta method of moments. Despite the analytical gain, computation becomes even more complicated, unless one restricts each $\boldsymbol{\Sigma}_{ik}$ to be diagonal, which is not appropriate in my case. Also, the use of Jensen's inequality ensures that the evidence lower bound will be achieved, conditional on convergence of the algorithm.

its mean is $\boldsymbol{\mu}_{ik} + \frac{1}{2}\boldsymbol{\Sigma}_{ik}$.

I will compute the remaining unknown term that enters f, $\mathbb{E}_{\mathbf{z}}[z_{itk}]$, after deriving the form of $q(\mathbf{z}_{it})$.

It should be emphasized that I did not assume a specific form of the variational distribution of $\boldsymbol{\beta}_{ik}$. Rather, it was obtained from the Taylor expansion. Note that variational parameters $\boldsymbol{\mu}_{ik}$ and $\boldsymbol{\Sigma}_{ik}$ do not have an economic interpretation and do not relate to the discussion of the random utility model in section 1.1—they only arise in the minimization of the KL divergence.

A.2.3. Update for $q(\mathbf{z}_{it})$

Notice that

$$\log q(\mathbf{z}_{it}) = const + \sum_{k=1}^{K} z_{itk} \left[\mathbb{E}_{\boldsymbol{\beta}}\left[\log \sigma\left(\mathbf{x}_{it}'\boldsymbol{\beta}_{ik}\right)\right] + \sum_{w=1}^{V} y_{itw} \mathbb{E}_{\boldsymbol{\theta}}\left[\log \theta_{kw}\right] \right]. \quad \text{(A.16)}$$

We conclude that $q(\mathbf{z}_{it})$ is a Multinomial$(1, \boldsymbol{\xi}_{it})$ distribution with a typical parameter,

$$\xi_{itk} = \frac{1}{S} \exp\left\{ \mathbb{E}_{\boldsymbol{\beta}}\left[\log \sigma\left(\mathbf{x}_{it}'\boldsymbol{\beta}_{ik}\right)\right] + \sum_{w=1}^{V} y_{itw} \mathbb{E}_{\boldsymbol{\theta}}\left[\log \theta_{kw}\right] \right\}, \quad \text{(A.17)}$$

where S is a normalizing constant. As Blei et al. (2003) show,

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\log \theta_{kw}\right] = \Psi(\phi_{kw}) - \Psi\left(\sum_{w'=1}^{V} \phi_{kw'}\right), \quad \text{(A.18)}$$

where $\Psi(\phi_{kw})$ is the first derivative of the log gamma function called the digamma function. Also, as we already know,

$$\mathbb{E}_{\boldsymbol{\beta}}\left[\log\left(\sum_{l=1}^{K}\exp\left(\mathbf{x}_{it}'\boldsymbol{\beta}_{il}\right)\right)\right] \leq \log\left(\sum_{l=1}^{K}\exp\left(\mathbf{x}_{it}'\boldsymbol{\mu}_{il}+\frac{1}{2}\mathbf{x}_{it}'\boldsymbol{\Sigma}_{il}\mathbf{x}_{it}\right)\right). \qquad \text{(A.19)}$$

A.2.4. Completing $q\left(\boldsymbol{\theta}_k\right)$ and $q(\boldsymbol{\beta}_{ik})$

Now that we know the parametric form of $q\left(\mathbf{z}_{ij}\right)$, we can complete the update step for $q\left(\boldsymbol{\theta}_k\right)$ and $q(\boldsymbol{\beta}_{ik})$.

As A.10 and A.17 imply,

$$\phi_{kw} = \alpha_{kw} + \sum_{i=1}^{I}\sum_{t=1}^{T_i}\xi_{itk}y_{itw}. \qquad \text{(A.20)}$$

Finally, in A.14 we substitute ξ_{itk} for $\mathbb{E}_{\mathbf{z}}[z_{itk}]$.

A.3. Approximate Variational Objective

First rewrite the expression for the true evidence lower bound A.4,

$$\mathcal{L}(q) = \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{\beta},\mathbf{z}}\left[\log p(\boldsymbol{\beta},\boldsymbol{\theta},\mathbf{z},\mathbf{X},\mathbf{Y})\right] - \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{\beta},\mathbf{z}}\left[\log q(\boldsymbol{\beta},\boldsymbol{\theta},\mathbf{z})\right]. \qquad \text{(A.21)}$$

As the expectation is essentially unconditional, henceforth I drop the parameter subscript. Notice that the full joint probability $p(\boldsymbol{\beta},\boldsymbol{\theta},\mathbf{z},\mathbf{X},\mathbf{Y})$ can be split into the product of marginal distributions,

$$p(\boldsymbol{\beta},\boldsymbol{\theta},\mathbf{z},\mathbf{X},\mathbf{Y}) = p\left(\mathbf{Y}\,|\,\boldsymbol{\theta},\mathbf{z}\right)\times p\left(\mathbf{z}\,|\,\boldsymbol{\beta}\right)\times p(\boldsymbol{\beta})\times p(\boldsymbol{\theta}). \qquad \text{(A.22)}$$

As the derivations in section A.2 are based on approximations, one cannot maximize A.21 directly. Instead, I use the surrogate objective function, $\widehat{\mathcal{L}}(q) \approx \mathcal{L}(q)$, where

$$
\begin{aligned}
\widehat{\mathcal{L}}(q) \;=\;& \underbrace{\sum_{i=1}^{I}\sum_{t=1}^{T_i}\sum_{k=1}^{K}\xi_{itk}\left[\sum_{w=1}^{V} y_{itw}\left(\Psi(\phi_{kw}) - \Psi\left(\sum_{w'=1}^{V}\phi_{kw'}\right)\right)\right]}_{\mathbb{E}[\log p(\mathbf{Y}\,|\,\boldsymbol{\theta},\mathbf{z})]} \\
+\;& \underbrace{\sum_{i=1}^{I}\sum_{t=1}^{T_i}\sum_{k=1}^{K}\xi_{itk}\left[\mathbf{x}_{it}'\boldsymbol{\mu}_{ik} - \log\left(\sum_{l=1}^{K}\exp\left(\mathbf{x}_{it}'\boldsymbol{\mu}_{il} + \frac{1}{2}\mathbf{x}_{it}'\boldsymbol{\Sigma}_{il}\mathbf{x}_{it}\right)\right)\right]}_{\mathbb{E}[\log p(\mathbf{z}\,|\,\boldsymbol{\beta})]} \\
+\;& \underbrace{\sum_{k=1}^{K}\left(-\frac{1}{2}\right)\left[HI\log 2\pi + I\log|\boldsymbol{\Omega}_k| + \mathrm{Trace}\left\{\boldsymbol{\Omega}_k^{-1}\sum_{i=1}^{I}\left[\boldsymbol{\Sigma}_{ik} + (\boldsymbol{\mu}_{ik}-\boldsymbol{\zeta}_k)(\boldsymbol{\mu}_{ik}-\boldsymbol{\zeta}_k)'\right]\right\}\right]}_{\mathbb{E}[\log p(\boldsymbol{\beta})]} \\
+\;& \underbrace{\sum_{k=1}^{K}\left\{\log\Gamma\left(\sum_{w=1}^{V}\alpha_{kw}\right) + \sum_{w=1}^{V}\left[-\log\Gamma(\alpha_{kw}) + (\alpha_{kw}-1)\left(\Psi(\phi_{kw}) - \Psi\left(\sum_{w'=1}^{V}\phi_{kw'}\right)\right)\right]\right\}}_{\mathbb{E}[\log p(\boldsymbol{\theta})]} \\
-\;& \underbrace{\sum_{i=1}^{I}\sum_{t=1}^{T_i}\sum_{k=1}^{K}\xi_{itk}\log\xi_{itk}}_{\mathbb{E}[\log q(\mathbf{z})]} \\
-\;& \underbrace{\sum_{k=1}^{K}\left(-\frac{1}{2}\right)\left[HI\log 2\pi + \sum_{i=1}^{I}\log|\boldsymbol{\Sigma}_{ik}| + I\right]}_{\mathbb{E}[\log q(\boldsymbol{\beta})]} \\
-\;& \underbrace{\sum_{k=1}^{K}\left\{\log\Gamma\left(\sum_{w=1}^{V}\phi_{kw}\right) + \sum_{w=1}^{V}\left[-\log\Gamma(\phi_{kw}) + (\phi_{kw}-1)\left(\Psi(\phi_{kw}) - \Psi\left(\sum_{w'=1}^{V}\phi_{kw'}\right)\right)\right]\right\}}_{\mathbb{E}[\log q(\boldsymbol{\theta})]}.
\end{aligned}
\tag{A.23}
$$

A.4. Variational M-Step

So far I considered the parameters $\boldsymbol{\zeta}_k$ and $\boldsymbol{\Omega}_k$ fixed. Now that I have derived the surrogate objective function, $\widehat{\mathcal{L}}(q)$, I can maximize it with respect to these parameters—this would be the ultimate updating step in my procedure.[39]

Notice that both $\boldsymbol{\zeta}_k$ and $\boldsymbol{\Omega}_k$ enter only the third term in the right-hand side of equation A.23. From the first-order conditions we obtain the following estimators,

$$\begin{aligned} \widehat{\boldsymbol{\zeta}}_k &= \frac{1}{I}\sum_{i=1}^{I} \boldsymbol{\mu}_{ik}, \\ \widehat{\boldsymbol{\Omega}}_k &= \frac{1}{I}\sum_{i=1}^{I} \boldsymbol{\Sigma}_{ik} + \frac{1}{I}(\boldsymbol{\beta}_{ik} - \boldsymbol{\zeta}_k)(\boldsymbol{\beta}_{ik} - \boldsymbol{\zeta}_k)' \end{aligned} \tag{A.24}$$

A.5. Analysis of Convergence

Pseudocode A.1 summarizes the estimation procedure. E step repeats until the relative change in the joint norm of $\boldsymbol{\mu}_{ik}$ does not drop below ε, though alternative convergence criteria are possible.

Note that the maximization of ELBO is not a convex problem and thus there is no guarantee that my numerical solution is a global optimum.

[39] This part is similar to that in Braun and McAuliffe (2010).

Algorithm A.1 Variational inference for EAA

initialize $q(\boldsymbol{\beta}_{ik})$, $q(\boldsymbol{\theta}_k)$, $q(\mathbf{z}_{it})$ randomly
while relative improvement in $\widehat{\mathcal{L}}(q) > \varepsilon_0$ **do**
  E step:
  repeat
    for i, t, k, w **do**
      set $\boldsymbol{\mu}_{ik} = \arg\max_{\boldsymbol{\beta}_{ik}} f(\boldsymbol{\beta}_{ik})$
      set $\boldsymbol{\Sigma}_{ik} = -\left(\nabla^2 f(\boldsymbol{\mu}_{ik})\right)^{-1}$
      set $\phi_{kw} = \alpha_{kw} + \sum_{i=1}^{I}\sum_{t=1}^{T_i} \xi_{itk} y_{ijw}$
      set $\xi_{itk} = \frac{1}{S}\exp\left\{\mathbb{E}_{\boldsymbol{\beta}}\left[\log\sigma\left(\mathbf{x}'_{it}\boldsymbol{\beta}_k\right)\right] + \sum_{w=1}^{V} y_{itw}\mathbb{E}_{\boldsymbol{\theta}}\left[\log\theta_{kw}\right]\right\}$
  until not converged
  M step:
  for k **do**
    set $\widehat{\boldsymbol{\zeta}}_k = \frac{1}{I}\sum_{i=1}^{I}\boldsymbol{\mu}_{ik}$
    set $\widehat{\boldsymbol{\Omega}}_k = \frac{1}{I}\sum_{i=1}^{I}\boldsymbol{\Sigma}_{ik} + \frac{1}{I}(\boldsymbol{\beta}_{ik} - \boldsymbol{\zeta}_k)(\boldsymbol{\beta}_{ik} - \boldsymbol{\zeta}_k)'$

The numerical optimization routine is based on the Newton-CG method. Supplying $\nabla f(\boldsymbol{\mu}_{ik})$ and $\nabla^2 f(\boldsymbol{\mu}_{ik})$ in analytical form, we achieve very good performance.

Finally, it turns out that the algorithm converges better if one initializes the variational parameters, $\boldsymbol{\mu}_{ik}$, $\boldsymbol{\Sigma}_{ik}$, $\boldsymbol{\phi}_k$, $\boldsymbol{\xi}_{it}$, at the beginning of each E step.

B. Estimated Parameters

B.1. Discovered Topics

Topic ID	Topic label	Representative stems
1	"Energy"	*energi,oil,price,gas,provid,cost,fuel,tax,natur,suppli,drill,coal*
2	"Immigration"	*person,border,drug,come,work,illeg,enforc,patrol,depart,texa,job,republican*
3	"Iraq War"	*iraq,war,nation,life,troop,soldier,armi,militari,sergeant,forc,sacrific,marin*
4	"Misc. (Native American issues / health)"	*health,fund,indian,care,cell,servic,tribe,land,reserv,public,right,protect*
5	"Misc."	*yield,energi,work,child,percent,debat,fund,incom,health,cost,increas,credit*
6	"Science and research"	*program,science,educ,school,research,fund,teacher,grant,technolog,institut,engin,innov*
7	"Education funding"	*student,loan,colleg,educ,cost,tax,famili,cut,public,grant,school,budget*
8	"Intelligence agencies"	*protect,intellig,fisa,surveil,discrimin,immun,foreign,civil,secur,target,limit,liberti*
9	"Subprime mortgage crisis"	*home,famili,economi,mortgag,job,foreclosur,loan,percent,unemploy,rate,stimulus,crisi*
10	"Spending on healthcare"	*drug,tax,price,safeti,fund,pay,program,spend,percent,child,fda,health*
11	"Misc. (agriculture / energy)"	*energi,farm,fuel,farmer,market,ethanol,price,produc,agricultur,food,crop,busi*
12	"Health insurance"	*drug,child,health,care,safeti,wage,insur,minimum,fund,spend,medic,patient*
13	"Government spending"	*tax,support,secur,program,fund,iraq,budget,percent,money,job,public,polici*
14	"Natural disasters and catastrophes"	*insur,farm,flood,home,disast,flood,develop,hurrican,lousiana,storm,katrina,risk*
15	"Misc. (energy / crisis)"	*energi,mortgag,oil,loan,market,price,gas,consum,economi,crisi,borrow,financi*
16	"Economic policy"	*tax,spend,increas,budget,percent,rate,cut,taxpay,econom,incom,inflat,reform*
17	"Economic policy"	*republican,energi,tax,packag,reserv,oil,debat,spend,polici,fund,cut,financ*
18	"War on terror / domestic security"	*secur,war,iraq,protect,gun,border,forc,attack,gang,drug,crime,terror*
19	"Immigration"	*alien,employ,secur,immigr,labor,status,violat,homeland,border,visa,enforc,penalti*
20	"Student loans"	*student,colleg,educ,tax,loan,school,provid,famili,cost,support,help,incom*
21	"Misc. (Iraq War / energy)"	*iraq,fund,energi,forc,oil,water,militari,border,polici,war,troop,resourc,*
22	"Misc."	*energi,water,farm,health,fund,project,system,protect,area,river,job,land*

Table B.1: Topics discovered by the EAA algorithm

Topic ID	Topic label	Representative stems
23	"(Secondary) education"	*school,district,grant,facil,public,educ,child,fund,student,system,teacher,sport*
24	"War on terror / domestic security"	*secur,depart,homeland,protect,iraq,attack,safeti,terrorist,enforc,guard,border,forc*
25	"Misc. (energy / public spending)"	*tax,budget,oil,support,educ,medicar,energi,secur,technolog,care,busi,job*
26	"Economic policy"	*tax,budget,spend,pay,debt,economi,benefit,middl,deficit,fiscal,minimum,growth*
27	"Misc."	*health,budget,cover,incom,immigr,farm,system,privat,job,polici,illegal,carbon*
28	"Health care and education funding"	*health,care,program,percent,tax,servic,educ,school,student,pay,medicar,colleg*
29	"Iraq War"	*iraq,war,defens,forc,troop,secur,command,deploy,weapon,serv,armi,threat*
30	"Iraq War"	*iraq,war,troop,budget,money,tax,spend,famili,increas,percent,fund,pay*
31	"Energy"	*oil,energi,price,gas,fuel,drill,barrel,suppli,develop,renew,ethanol,technolog*
32	"Energy"	*energi,oil,price,gas,economi,fuel,consum,effici,clean,altern,carbon,global*
33	"Misc."	*farm,fund,agricultur,lobbi,servic,health,educ,employe,scienc,energi,disclosur,research*
34	"Energy"	*energi,oil,fuel,gas,futur,renew,suppli,ethanol,compani,coal,develop,invest*
35	"Iraq War"	*iraq,republican,war,democrat,debat,nation,veteran,fund,ryan,militari,famili,forc*
36	"Stem cell research"	*cell,research,stem,life,human,health,diseas,scienc,woman,medic,organ,cancer*
37	"Iraq War"	*iraq,troop,war,nation,veteran,support,continu,militari,govern,unit,forc,polit*
38	"Public spending"	*energi,famili,serv,life,oil,school,communiti,public,program,educ,care,busi*
39	"Iraq War"	*iraq,troop,strategi,war,militari,polit,baghdad,fight,qaida,withdraw,terrorist,sunni*
40	"Iraq War"	*iraq,troop,war,protect,terrorist,foreign,secur,plan,order,target,attack,fisa*
41	"Transportation"	*airlin,transport,servic,energi,safeti,coast,public,passeng,fuel,aviat,rail,traffic*
42	"Iraq War"	*iraq,war,militari,forc,fight,mission,soldier,strategi,terrorist,afghanistan,qaida,withdraw*
43	"Health care"	*health,life,care,hiv,woman,aid,organ,prevent,african,diseas,protect,treatment*
44	"Veterans of Iraq War"	*iraq,veteran,care,troop,war,soldier,home,militari,support,health,fund,medic*

Table B.2: Topics discovered by the EAA algorithm (cont.)

B.2. Preferences towards Divisive Issues

	Const	Chamber	Party	Ideology	Primary	General
mean	6.4518	2.4009	2.5973	0.9233	0.5978	-1.4306
correlation	1	0.5063	0.4515	0.0885	0.1753	-0.5295
		1	0.501	-0.231	0.3292	-0.1172
			1	-0.0225	0.1658	0.0878
				1	-0.2283	0.0103
					1	0.0082
						1

Table B.3: Preferences towards topic "Energy" (#31)

	Const	Chamber	Party	Ideology	Primary	General
mean	2.1049	1.5775	1.3517	-0.6926	0.0883	-1.5517
correlation	1	0.378	0.29	-0.4589	-0.0039	-0.4712
		1	0.1756	-0.5533	0.082	-0.608
			1	-0.0045	0.241	0.1326
				1	-0.4167	0.1749
					1	0.4865
						1

Table B.4: Preferences towards topic "Energy" (#32)

	Const	Chamber	Party	Ideology	Primary	General
mean	0.3967	-0.03	1.2474	0.4217	-0.7076	0.5404
correlation	1	-0.1357	-0.3303	-0.1219	0.3333	-0.3092
		1	-0.4537	-0.5969	-0.2254	0.1593
			1	0.3448	-0.2476	0.2571
				1	-0.3925	0.34
					1	-0.717
						1

Table B.5: Preferences towards topic "Energy" (#34)

	Const	Chamber	Party	Ideology	Primary	General
mean	7.7431	4.5162	1.7232	3.2607	-2.491	-2.8801
correlation	1	0.2309	-0.0139	0.2052	-0.2405	-0.2746
		1	-0.0787	-0.1374	0.038	-0.0168
			1	0.0534	-0.3162	-0.239
				1	-0.3106	-0.4546
					1	0.3952
						1

Table B.6: Preferences towards topic "Immigration" (#2)

	Const	Chamber	Party	Ideology	Primary	General
mean	1.8421	2.6965	0.3106	2.7655	-0.3662	0.5739
correlation	1	0.3484	-0.1335	-0.057	0.1852	0.0827
		1	-0.29	0.5039	-0.2494	-0.0361
			1	0.1026	-0.0797	0.2405
				1	-0.3511	0.2825
					1	0.1516
						1

Table B.7: Preferences towards topic "Stem cell research" (#36)

	Const	Chamber	Party	Ideology	Primary	General
mean	10.4444	9.5594	8.4396	7.3493	4.7879	7.8631
correlation	1	0.1179	0.2157	0.0401	-0.159	0.2089
		1	0.1683	0.2597	0.2275	0.1951
			1	-0.461	0.1598	0.4065
				1	0.3156	0.0311
					1	0.3138
						1

Table B.8: Preferences towards topic "War on terror / domestic security" (#18)

	Const	Chamber	Party	Ideology	Primary	General
mean	3.8661	-0.8081	3.3815	0.9283	1.3093	1.7395
correlation	1	-0.2658	0.1691	-0.0519	0.114	0.2062
		1	-0.4884	-0.0221	-0.3776	-0.2159
			1	0.0156	0.6853	0.072
				1	-0.2394	0.4262
					1	-0.1093
						1

Table B.9: Preferences towards topic “War on terror / domestic security” (#24)

	Const	Chamber	Party	Ideology	Primary	General
mean	12.837	9.5479	1.3258	8.8828	-4.7808	-0.5871
correlation	1	0.4439	-0.1209	0.3144	-0.4204	-0.2051
		1	0.0524	0.3529	-0.3059	-0.0242
			1	0.0065	0.1016	0.2039
				1	-0.0819	0.0367
					1	-0.065
						1

Table B.10: Preferences towards topic “Iraq War” (#3)

	Const	Chamber	Party	Ideology	Primary	General
mean	5.7049	4.2104	3.8983	1.3216	-0.2326	0.4736
correlation	1	0.7106	0.4242	0.1017	-0.2005	0.1881
		1	0.0906	0.2897	-0.3939	0.2455
			1	-0.2344	0.3048	0.0046
				1	0.1357	-0.0419
					1	-0.3582
						1

Table B.11: Preferences towards topic “Iraq War” (#30)

	Const	Chamber	Party	Ideology	Primary	General
mean	4.4392	-1.4275	8.0046	-1.1957	0.0435	-2.0038
correlation	1	-0.4592	0.406	-0.3385	-0.1968	-0.1455
		1	-0.4592	0.3424	0.4998	0.1326
			1	-0.5557	-0.1693	-0.4829
				1	0.3654	0.1536
					1	0.2133
						1

Table B.12: Preferences towards topic "Iraq War" (#39)

	Const	Chamber	Party	Ideology	Primary	General
mean	7.6442	2.6035	0.8919	0.2315	-1.8855	0.4732
correlation	1	0.5084	-0.0889	-0.1673	-0.4239	0.2091
		1	-0.2653	-0.147	-0.372	0.2075
			1	-0.3701	0.4808	-0.5568
				1	-0.2077	0.1682
					1	-0.2302
						1

Table B.13: Preferences towards topic "Iraq War" (#42)