Understanding delegation in the European Union through machine learning*

Jason Anastasopoulos
ljanastas@princeton.edu

Anthony M. Bertelli
bertelli@nyu.edu

September 6, 2017

Abstract

The delegation of powers by legislators is essential to the functioning of modern government, and presents an interesting tradeoff in multi-level states such as the European Union (EU). More authority for member states mitigates ideological drift by the European Commission, but less authority reduces the credibility of commitments to centralized policies. Extant empirical studies of this problem have relied on labor-intensive content analysis that ultimately restricts our knowledge of how delegation responded to legislative and executive power changes in recent years. We present a machine–learning approach to replicating the content analysis of 158 laws between 1958–2000 by Franchino (2001, 2007) that will "train" classifiers to examine EU laws enacted since 2000 in a similar way. Using the trained classifier with the highest overall performance, we introduce probabilistic delegation ratios (PDR) as an alternative to the delegation ratio first introduced by Epstein and O'Halloran (1999) and also demonstrate that our trained classifier is able to automatically estimate delegation ratios in legislation as well. While our principal interest is in the European Union, the method we employ can be used to understand delegation in a variety of contexts.

*Paper presented at the Annual Meeting of the American Political Science Association, San Francisco, CA, September 1, 2017. We are extremely grateful for the assistance of Fabio Franchino and Maulik Shah in providing and preparing data for us. This is preliminary work, please do not cite without permission.

1 Introduction

Delegation of powers is a central problem in modern representative government. As a theoretical construct, delegation represents a grant of authority by a legislature, who holds policy-making power as a constitutional matter, to an agent or set of agents, whose powers are determined by the conditions identified by the legislature in enabling statutes. In the multi-level governance setting of the European Union (EU), the problem of delegation is particularly interesting because EU legislators have a choice of agents: the European Commission (EC), the principal executive body with a large bureaucratic component, and the member states of the EU, which are representative governments in their own right. Franchino (2001, 2007) offers a formal argument that captures the legislative delegation decision as essentially a tradeoff between credible commitment to a common policy and the potential for policy drift by the EC and creates a human coding framework to test its implications in 158 major pieces of European legislation from 1958–2000. This paper reports on a project that uses machine-learning techniques to reproduce Franchino's human codings and ultimately extend them to the present day. In doing so, we can better understand how important institutional changes, such that have made legislative power more equal between European Council and Parliament, impact the substantive patterns of delegation and the structure of delegating legislation. While our principal interest is in the European Union, the method we employ can be used to understand delegation in a variety of contexts.

Delegation has three elements that are crucial to any quantitative strategy to measure it. The first is the *authority*, which identifies the agent or agents as holders

of policy-making power. Second, the substantive *content* of those powers is specified. Finally, enabling legislation often features a set of *constraints*, or conditions, on the authority that is being granted. For instance, a law might state that a specific administrative agency shall make rules and regulations to protect wildlife which can only be made after consultation with the public. Moreover, as important as it is, delegating legislation is not the only kind of law enacted by governments, and while the substantive content of statutes is not difficult for a quantitative researcher to identify, the remaining features are more difficult tasks in human coding applications. This paper uses a suite of modern machine learning methods for the purpose of identifying statutes delegating authority within the set of all legislation using the framework developed by Franchino (2001, 2007). These methods, in turn, allow us to construct a new measure of delegation that can account for the extent of delegated authority in any given statute as the delegation ratio does, but also

A variety of methods for capturing delegation through proxy measures have been used on a large scale. A variety of studies use the number of words in a statute as a measure that increases as agent discretion is more restricted (Clinton et al., 2012; Huber, Shipan, and Pfahler, 2001; Huber and Shipan, 2002). Vakilifathi (n.d.) improves on this measure by distinguishing between optional and mandatory provisions by identifying contextual triggers such as the use of "shall" versus "may." While these methods allow researchers to capture a broad used because of the scale of the problem, they cannot separate authority from constraint in the way that labor-intensive content analysis can (Epstein and O'Halloran, 1999; Franchino, 2001, 2007; McCann, 2016). We offer an automated means of deploying this richer method to

uncover delegation patterns. This is especially important because the grant of authority to a member state's government or to the EC is a crucial choice that defines the (Franchino, 2007).

In this preliminary report, we describe the structure of the legislative data, the content analytic framework that we replicate, and the process that we use to select the optimal machine learning method for identifying delegation in texts. We discuss how this method produces measures both of the extent to which any given statute delegates authority as well as automatically estimates more extant metrics like delegation ratios.

2 Data

The data used to train our machine learning classifiers were taken from (Franchino, 2001, 2007). Franchino codes provisions in 158 major pieces of European legislation from 1958 to 2000 by whether these provisions delegated executive powers from the European Community to member states or whether they imposed statutory constraints. Measures of delegation and constraint in Franchino (2001, 2007) are calculated using delegation and constraint ratios as defined by Epstein and O'Halloran (1999). Of particular interest to us is whether provisions in these laws delegated powers to member states of the EU. We focus on the locus of authority, rather than constraint, because our substantive interest at this stage is in the distribution of powers to member states. This makes the initial task of training machine learning classifiers to identify provisions which delegate authority much easier.

The original dataset is constituted of 158 pieces of legislation. After excluding legislation which was entirely in languages other than English[1], we had 147 pieces of legislation in our dataset. Using a series of regular expressions, we broke down these 147 pieces of legislation into provisions which produced 7,011 total provisions. These provisions were then coded as either containing delegations to member states or not using the coding scheme provided by (Franchino, 2001). Of the 7,011 provisions coded 2,857 (40.7%) delegated executive authority to EU member states. Below we provide some summary statistics for these provisions.

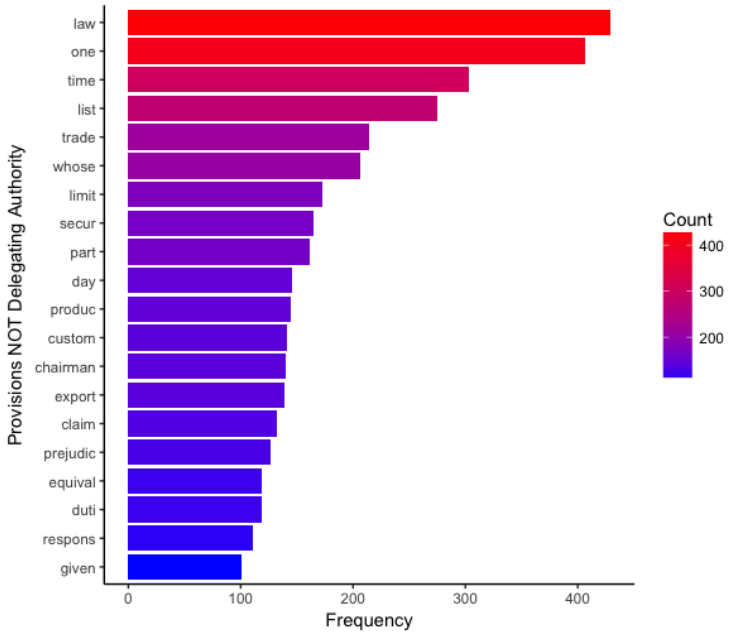

Figure 1: Word counts for pre–processed provisions not delegating authority to EC member states.

Figures 1 and 2 contain word counts for provisions which do not delegate authority to member states and delegate authority to member states, respectively. The

[1]These laws were mostly in French or German

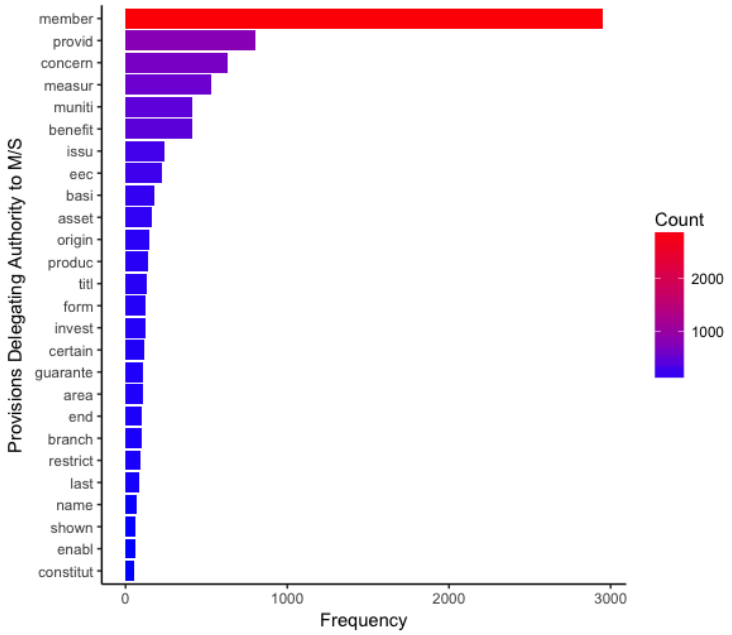

Figure 2: Word counts for pre–processed provisions delegating authority to EC member states.

terms contained within these figures were pre–processed and stemmed using a process described below. A glance at some of the language of delegation provided in Figure 2 suggests that provisions delegating authority to member states tend to mention provisions and benefits that member states will receive while provisions no delegating authority appear to be primarily related to trade and security. We study this issue in more depth using a topic model analysis of these provisions below.

3 Methods

This project involves two major stages. In the first stage, we study the nature of delegation in the EU by performing exploratory analyses of EU legislation and

provisions which delegate authority to member states. To accomplish this we use Latent Dirichlet Allocation (topic model) to extract the latent topical content of the 7,011 provisions that we have collected and explore topic proportions among provisions which are related to delegation in order to better understand which policy domains and themes member state authority tends to lie.

In the second stage, which involves multiple phases, we train a series of machine learning classifiers for the purpose of identifying delegation of authority to member states. The ultimate goal of this exercise is to construct a means of automatically computing the *delegation ratio* for any given piece of European Union legislation i (Δ_i), which is simply the ratio of the number of provisions delegating executive powers to member states divided by the total number of provisions:

$$\Delta_i = \frac{\text{\# of provisions delegating authority}}{\text{total \# of provisions}}$$

Because we deal directly with texts, provisions will be weighted according to a number of criteria discussed below.

3.1 Exploratory Analysis

Before training a series of machine learning classifiers for automated calculation of the delegation ratio, we perform an exploratory analysis of the provisions which delegate authority to EU member states. These exploratory analyses are conducted using the unsupervised machine learning technique known as the latent Dirichlet allocation which has the more common name *topic model* (Blei, Ng, and Jordan, 2003; Blei, 2012; Roberts et al., 2014). A more in depth explanation about topic

models and how they are used in these analyses can be found in the Appendix.

3.2 Estimating Delegation in EU Legislation

Phase	Stage	Action	Algorithms	Data
I	1	Training + Testing	Sparse Logistic Reg. Naive Bayes SVMs Sparse Bayesian Reg. Random Forest	Labeled EU provisions,1958–2000. (Franchino, 2001)
	2	Performance	*	Accuracy (a_m) Sensitivity (σ_1) Specificity (σ_2)
II	1	Classification	Best performing from Phase I	Unlabeled provisions from EU legislation 2001 – 2016.
	2	Delegation Ratio	$\frac{1}{P_i}\sum_{p=1}^{N_i}\delta_{pi}$	Labeled provisions
		PDR Scores	$\frac{1}{P_i}\sum_{p=1}^{N_i}P(\delta_{pi}\|X)\delta_{pi}$	from EU legislation 2001–2016

Table 1: Phases and stages of analysis for scoring delegation in EU legislation from 2001–2016

Table 1 contains a summary of each of the two phases along with each stage within the phases. Phase I uses the labeled EU provisions provided by Franchino (2001) between 1958–2000. In Phase I, a series of machine learning classifiers are trained using the labeled EU provisions for the purpose of identifying delegation to member states. For exposition purposes, we will describe how we trained the regularized logistic regression model using the EU provision data it is one of the more

intuitive models that is easiest to understand. Unfortunately, regularized logistic regression also tends to perform poorly with text data (Ng and Jordan, 2002). Detailed descriptions of the other machine learning algorithms that we used for training can be found in the Appendix.

The second stage of Phase I involves assessing the relative performance of each of the machine learning algorithms that we use. Performance of each classifier, c is assessed using three metrics, accuracy a_m, specificity σ_1 and sensitivity σ_2. If $D =$ delegating authority and $ND =$ not delegating authority, accuracy is simply the % of correctly identified provisions delegating authority to EU member states in the test set.

$$a_c = \frac{D|D + ND|ND}{P}$$

Sensitivity, or the true positive rate, is the number of provisions correctly identified as delegating authority over the number of true and false positives

$$\sigma_{1c} = \frac{D|D}{D|D + D|ND}$$

Specificity, or the true negative rate, is the ratio of the number of provisions correctly identified as not delegating authority over the total number of true and false negatives .

$$\sigma_{2c} = \frac{ND|ND}{ND|ND + ND|D}$$

We seek to build a classifier which maximizes performance across all three cate-

gories. Thus for any classifier c, an ideal classifier c^* will be the classifier that evenly maximizes some weighted combination vector $\omega : 0 < \omega < 1$ of accuracy, sensitivity and specificity:

$$c^* = \arg\max_c (a_c \ \sigma_{1c} \ \sigma_{2c})\omega \tag{1}$$

For our purposes, we assume that accuracy, specificity and sensitivity are equally important and so we set $\omega = (1/3 \ 1/3 \ 1/3)$.

After the optimal classifier is chosen, the classifier will be applied to statutes in the EU from 2000 to the present to estimate two quantities of interest: (1) delegation ratios and; (2) probabilistic delegation ratios (PDR), a new measure of delegation that we introduce. While delegation ratios only measure the proportion of provisions delegating authority in a piece of legislation, PDR scores provide information about whether the provisions delegate authority and the *probability* that each provision in the legislation delegates authority as well. The PDR score of a piece of legislation is a delegation ratio weighted by the predicted probability that the provisions in that legislation delegate authority. As a result the PDR will always be *lower* than the delegation ratio.

This provides several advantages over the delegation ratio for our purposes. First, for replication, the PDR and its uncertainty estimates provide useful information about the confidence that we have that the hand–coded delegation ratio has been reproduced by our classifiers. Second, because we intend to move beyond replication, the PDR scores will provide an important way of determining how appropriate the Franchino coding scheme is for understanding more recent legislation. This is

particularly important given institutional changes in executive and legislative power. Third, the contemporary understanding of text–as–data includes an element of uncertainty generated, for instance, by non-random biases of coders Benoit, Laver, and Mikhaylov (2009); Laver, Benoit, and Garry (2003), and comparisons between PDR scores and hand-coded delegation ratios can provide some information about this phenomenon. Overall, the PDR contains some information that delegation ratios are not able to capture.

3.2.1 Example: training a delegation classifier with sparse logistic regression

Step	Action	Description	Tools
1	Text pre–processing	Raw text is transformed into consistent terms.	Regular expressions. Tokenizer Stemmers.
2	Document–term matrix	Pre–processed text is converted into a DxT matrix where D = documents (EU provisions) and T = terms (words/phrases). TF or TF–IDF weights assigned.	Text processing packages. Eg). *tm*, *quanteda* in **R**

Table 2: Stages of Preparing EU Provision Text for Training Machine Learning Algorithms

Before we discuss how we train the delegation classifier, it is important to understand the steps that go into data preparation in order to better understand the rationale behind some of the uses of these methods. Table **??** presents the steps involved in preparing the EU provision text prior to analysis. Step 1 involves pre–processing of texts to create text that has consistent units of analysis. The raw text

from each of the EU provisions are put through a "cleaner" function that we designed which contains a combination of regular expressions and natural language processing (NLP) tools.

The regular expressions remove special characters, numbers and transform all words to lower case. Tokenizers split provisions into n–grams, which can be words (1–gram) or phrases (typically 2– or 3–grams). For these analyses, we began with a unigram model which had achieved the best current performance. Terms are also stemmed, a NLP process in which suffixes and sometimes prefixes are removed and stop words, which are typically the most common words in a language[2] are removed as they typically improve the performance of supervised and unsupervised machine learning classifiers (Ikonomakis, Kotsiantis, and Tampakas, 2005; Kotsiantis, 2007).

$$\mathcal{D} = \begin{array}{c} \\ Document_1 \\ Document_2 \\ Document_3 \\ Document_4 \\ Document_5 \\ \vdots \\ Document_S \end{array} \begin{array}{c} \begin{array}{ccccccc} Term_1 & Term_2 & Term_3 & Term_4 & Term_5 & \cdots & Term_T \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & \cdots & 0 \\ 0 & 5 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & 1 & \cdots & 0 \end{pmatrix} \end{array}$$

Figure 3: A sample document–term matrix for a corpus.

After the pre–processing steps, the provisions are then transformed from text to data via the document–term matrix or DTM. The document–term matrix is simply a matrix in which the documents comprise the rows while the terms comprise the

[2]For example "the" , "and" etc.

columns. Figure 3 contains a sample document–term matrix. Typically, the entries of the document–term matrix can be of two types: (1) term–frequencies (TF) or; (2) term–frequency inverse–document frequency (TF–IDF). Term frequency entries are simply the number or times that a term appears in a document while TF–IDF are a weighted version of the term frequencies. TF–IDF weights are the preferred method used for supervised machine learning algorithms and had originally been used for information retrieval purposes. We discuss this weighting scheme in more detail in the Appendix.

$$logit(E[D|T]) = \theta_0 + \theta_1\tau_1 + \theta_2\tau_2 + \cdots + \theta_n\tau_n + \epsilon \tag{2}$$

We now turn our attention to a description of how the provisions delegating authority are modeled using sparse logistic regression. Equation 2 is the linear form of a typical logistic regression for the *training data* that we chose. D is whether a provision delegate authority to EU members states while the variables $\tau_1, \cdots, \tau_n$ are the term vectors from the document term matrix. Training the model involves estimating the parameters, $\theta_1, \cdots, \theta_n$ and then using these parameters to estimate the probability that any given set of provisions involves delegation of authority to member states in the *training set* of data:

$$P(D = 1|T) = \frac{exp(\theta T)}{1 + exp(\theta T)} \tag{3}$$

where $P(D = 1|T)$ is the probability that a provision contains delegation of authority to member states given the terms. For the new provisions in the test set,

a provision is labeled as delegating authority to a member state when:

$$D_k = \arg\max_k P(D = k|T) \tag{4}$$

Since we only have two classes, a provision is labeled as delegating authority if $P(D = 1|T) > 0.50$ and labeled as not delegating authority otherwise.

Up to this point, we have described how we have modeled our data using the common form of logistic regression. However, since we are dealing with text data which is a high–dimensional inference problem, we use *sparse logistic regression* which is simply logistic regression with a *regularization parameter* also known as the LASSO which penalizes variables (terms in this case) which do not contribute to predicting the probability of delegation. When dealing with high–dimensional problems in which there are a large number of covariates, LASSO methods generally reduce mean squared error and classification error and in most contexts and allow for parameter estimation in high dimensional spaces in which estimation would be intractable, for example when the number of parameters estimated exceeds the number of observations, which is commonly the case in text–analysis (Tibshirani, 1996; Genkin, Lewis, and Madigan, 2007; Tibshirani, 2011; Ratkovic and Tingley, 2017).

Thus, while in ordinary logistic regression we would estimate parameter values by minimizing the following loss function:

$$\arg\min_\theta \sum_{i=1}^{n} - \left[D_i \theta^T T_i - log(1 + exp(\theta^T T_i))\right] \tag{5}$$

for *sparse logistic regression* we estimate parameters using a loss function similar

to the one above but with a ℓ_1 regularization norm[3] which penalizes or shrinks the parameter values in the document term matrix to zero by λ:

$$\arg\min_{\theta} \sum_{i=1}^{n} - \left[D_i \theta^T T_i - log(1 + exp(\theta^T T_i)) \right] + \lambda ||\theta||_1 \tag{6}$$

To sum up the steps of the process:

1. The provisions are randomly split into a test and training set (75%/25%, respectively).

2. The training data are pre–processed using the methods described in Table 2.

3. Model parameters are estimated for Equation 2 using the loss function specified in Equation 6.

4. Performance metrics such as accuracy, specificity and sensitivity are estimated for the trained model by applying the trained model to the labeled test data using the criteria specified in Equation 4.

While the initial results use only test data from the EU provisions, we plan on confirming model performance using randomly selected EU provisions over the out of sample time period of interest through an iterative process of automated classification and manual validation.

[3]There are several choices of regularization norms for the lasso but the ℓ_1 has typically been found to produce the best solutions with the lowest mean squared error in a number of supervised machine learning problems (Koh, Kim, and Boyd, 2007). Please see the Appendix for a more detailed description of regularization norms.

3.2.2 Estimation of Delegation Ratios in New Legislation

After new provisions within legislation are computed using the trained model from our final machine learning classifier, automatic computation of delegation ratios for new pieces of legislation is a trivial matter. If Δ_{it} is the delegation ratio for any piece of legislation i at time t, P_{it} is the total number of provisions in legislation i and δ_{it} is the total number of provisions that delegate executive authority to member states, then the delegation ratio is:

$$\Delta_{it} = \frac{\delta_{it}}{P_{it}} \tag{7}$$

Where δ_{it} is computed from the trained model and P_{it} can be calculated either directly from the text of the legislation or from available legislation statistics.

4 Results

Below we present the results of our exploratory analyses and machine learning classifier training. As mentioned above, the purpose of training the machine learning classifiers is to enable us to reconstruct delegation ratios and study them in EU legislation from 2000 to the present. In order to accomplish this we first train a total of 6 classifiers to identify delegation to member states in the coded provisions. Using a combined performance metric which averages accuracy, sensitivity and specificity, we choose the best performing classifier (in this case, a random forest model), reconstruct delegation ratios in the test data using this classifier, and compare these delegation ratios with the hand coded ratios from Franchino (2001). The labeled

data are the 7,011 provisions which either delegate authority to members states or do not. Reconstruction of delegation ratios in the test data is conducted by predicting delegation of authority using the random forest model and then calculating delegation ratios with these predicted labels.

4.1 Exploratory Analyses

4.2 Classifier Results

Using a 75/25 training/test split, we trained 6 machine learning classifiers to identify delegation of executive authority to member states. These classifiers include: two sparse Bayesian methods using a horseshoe prior and a lasso prior ("Bayes horseshoe" and "Bayes lasso") which have been found to yield good performance for predicting class labels in sparse signal data (Carvalho, Polson, and Scott, 2010) and text data (Genkin, Lewis, and Madigan, 2007), respectively; a sparse logistic regression classifier ("Logit lasso"); a naive Bayes classifier ("naive Bayes"); a random forest classifier ("Random forest") and a support vector machines classifier ("SVM").

Performance measures used are accuracy, sensitivity and specificity and the classifier which produced the highest average across all three categories was then used to reconstruct delegation ratios. Figure 4 contains accuracy, sensitivity and specificity measures for each of the six classifiers. It is clear from this plot that the best performing classifiers across three categories are the random forest and support vector machines classifier, each yielding accuracy, sensitivity and specificity performance above 70% with the random forest containing the highest levels of accuracy across

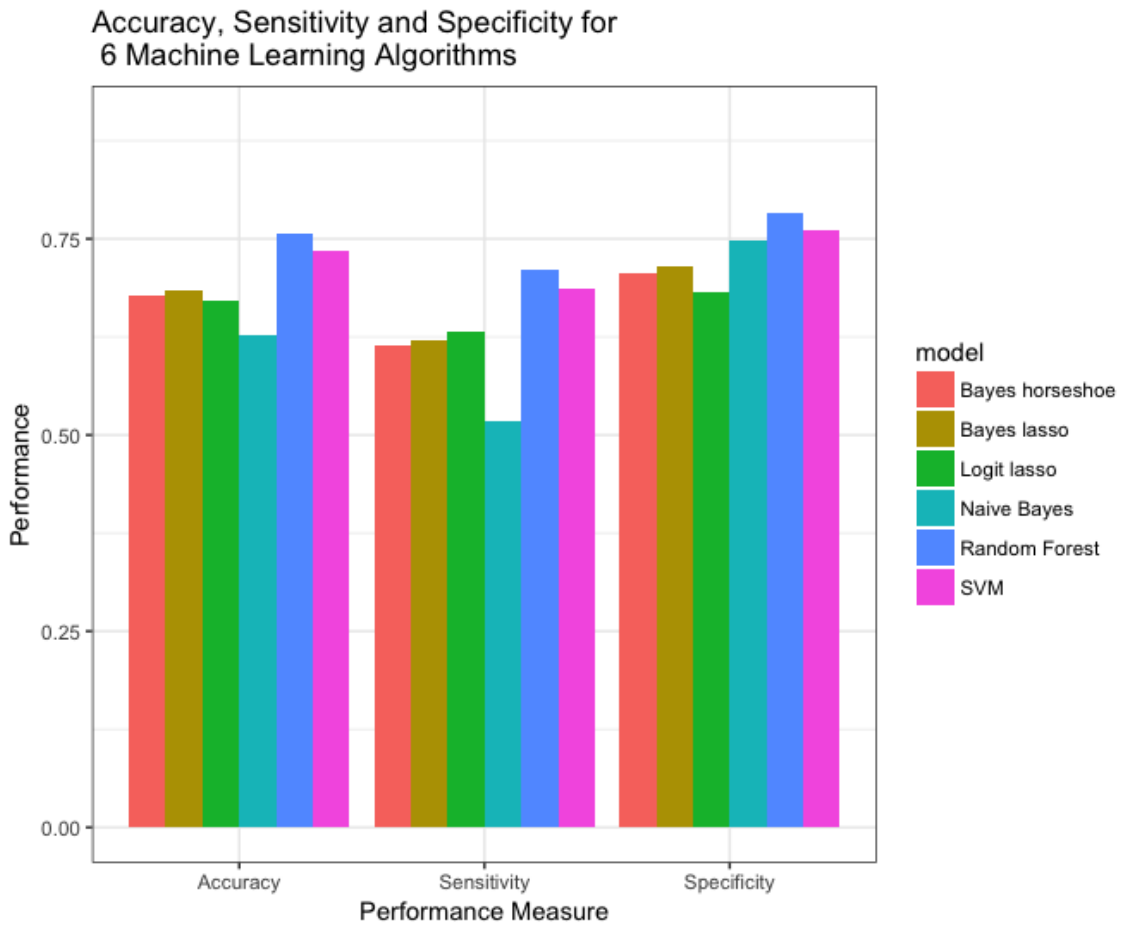

Figure 4: Test data performance in terms of accuracy, sensitivity and specificity for all algorithms trained.

all three categories.

As such, we chose the random forest classifier as our ideal means of identifying delegation in provisions and estimating delegation ratios in new legislation. Figure 5 contains average performance across all three performance metrics. This plot makes the best and worst performing classifiers clearer. The random forest model has the best overall performance across these three categories, averaging 74% while the naive

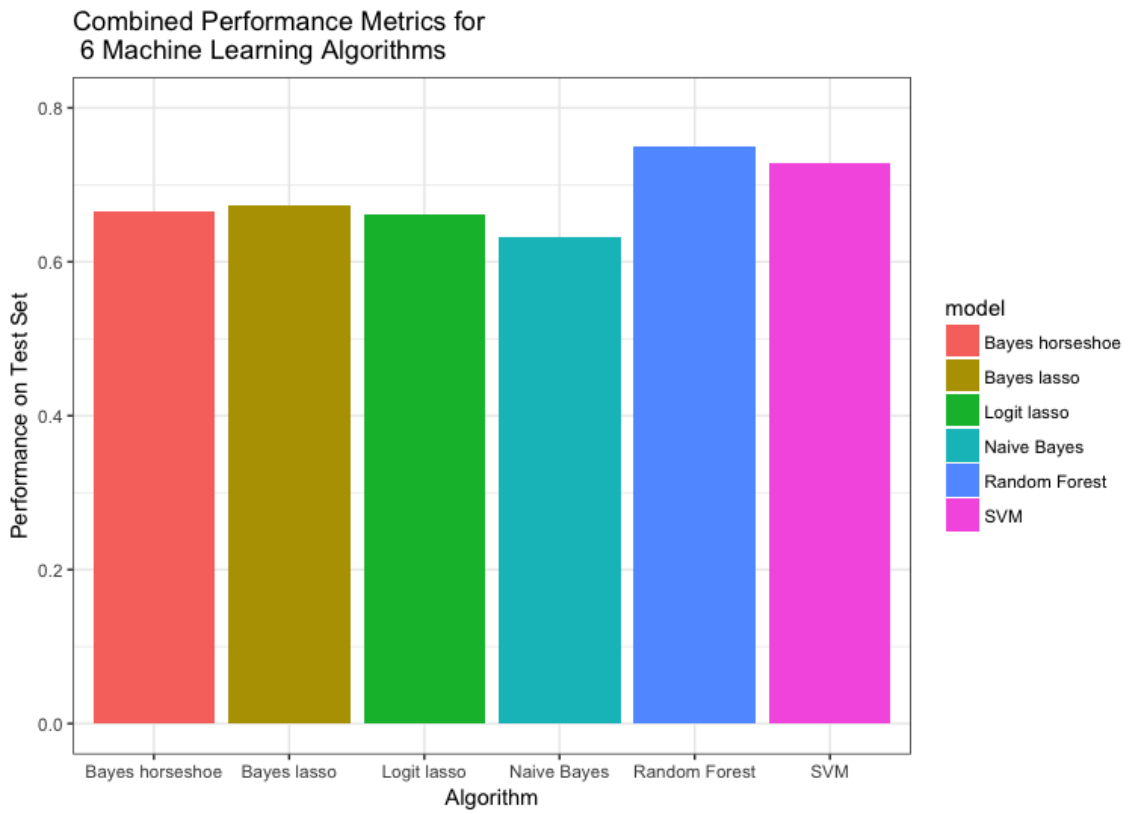

Figure 5: Aggregate test data performance for all algorithms trained. Aggregate performance is computed as $\frac{1}{3}$(accuracy + sensitivity + specificity).

Bayes model has the worst performance averaging 63%. Since the random forest classifier has yielded the best performance in this context, we have decided to use the random forest classifier as our "ideal" model for the purpose of reconstructing delegation ratios.

Figure 6 contains information about word importance for the random forest classifier. The mean decrease in the Gini statistic measures the extent to which each term (variable) is important in the classification of delegation of authority. Higher values suggest more importance in word classification. From Figure 6 we can see

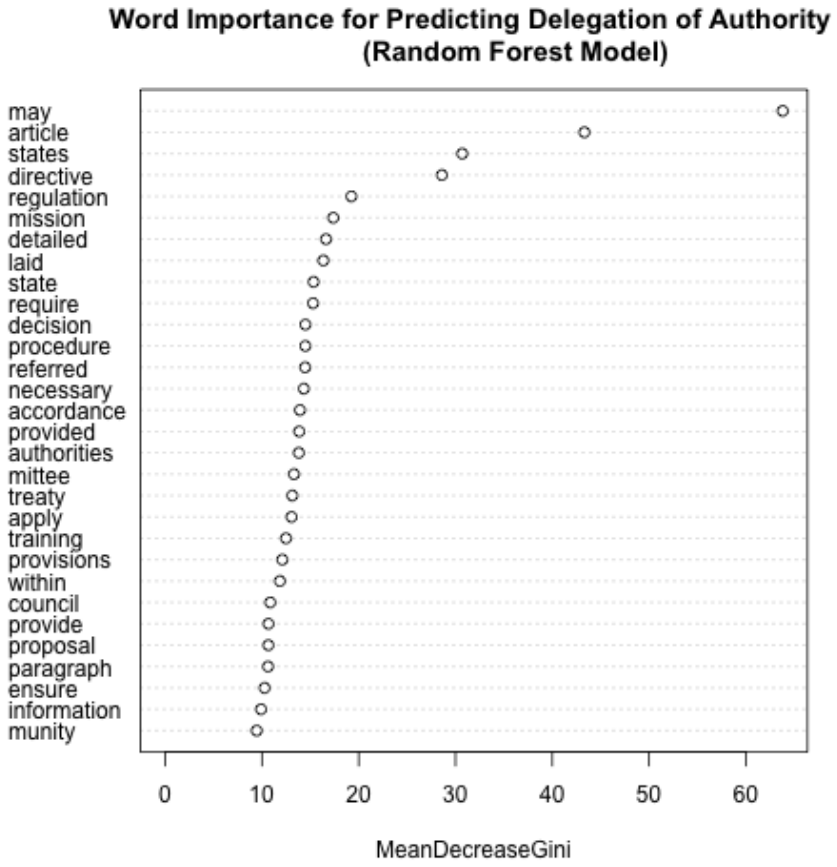

Figure 6: Average decrease in the Gini coefficient across all trees estimated using the random forest classifier. Higher values imply that the term had greater importance in terms of its ability to classify provisions delegating authority across trees.

that the terms which were most important in predicting delegation of authority in provisions were "may", "direct", the stemmed version of "apply" (ie apply, applied, etc), "provid" ("provide", "provided" etc) and "regul" ("regulate", "regulated" etc.) For example, Article 14 Section 2(d) of EEC Council Regulation 11 concerning inspection of goods transported over EU member state boundaries delegates authority to member states in the case of procedures for refusing inspections:

> *If any undertaking refuses inspection as provided for in this Regulation, the Member State concerned shall give the authorised representatives of the Commission such support and assistance as may be necessary for the purpose of carrying out their inspections as instructed. Member States shall introduce the necessary measures for this purpose before 1 July 1961, after consulting the Commission.*

This provision contains several of the terms which were most important in predicting delegation of authority by the random forest classifier such as "may", "authorised" and "regulation." This word importance plot tells us a great deal of substantive information about the language of delegation in the European council. Specifically, it shows that delegation of authority in the EU is most strongly tied to the word "may", a term which implies the granting of permission to do something. The use of the word "may" is interesting because it implies that the nature of the relationship between the EU and member states is that member states are subordinate units to which the EU grants certain powers. This is different from the relationship between Congress and bureaucracies in the United States in which delegation of authority from Congress to various federal agencies typically involves *entrusting* a federal agency with certain abilities and powers rather than *permitting* them to retain powers that Congress has. It is also interesting that this legislation relies on optional rather than mandatory phrasing, suggesting that the mandatory ("shall") provisions are not as important as optional ("may") provisions in this context (Vakilifathi, n.d.).

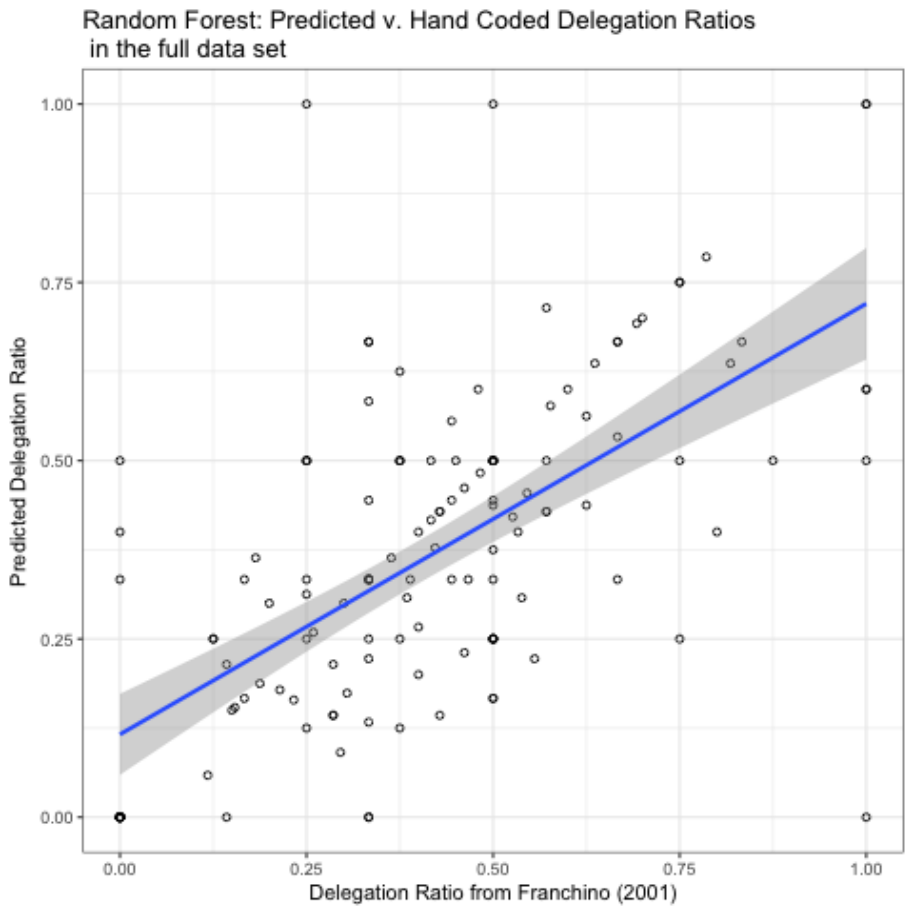

Figure 7: Ground truth delegation ratios (Δ_i) versus predicted delegation ratios from the random forest classifier ($\widehat{\Delta_i}$) using provisions in the test data only. $r = 0.64$

4.3 Automated Prediction of Delegation Ratios

Finally, we use the best performing classifier to predict delegation ratios in legislation as defined by Epstein and O'Halloran (1999); Franchino (2001, 2007). Recall that the ground truth delegation ratio for regulation i is Δ_i:

$$\Delta_i = \sum_p \delta_{ip} / \sum_p P_{ip} \tag{8}$$

Where $\sum_p \delta_{ip}$ are the number of provisions delegating authority in regulation i and $\sum_p P_{ip}$ are the total number of provisions in regulation i.

The predicted delegation ratio in the test set for regulation i is then $\widehat{\Delta_i}$:

$$\widehat{\Delta_i} = \sum_p \widehat{\delta_{ip}} / \sum_p P_{ip} \tag{9}$$

where $\sum_p \widehat{\delta_{ip}}$ are the total number of provisions delegating authority predicted by the random forest model. Figure 7 is a plot of ground truth delegation ratios (Δ_i) from Franchino (2001) computed for laws in the test set (x–axis) versus predicted delegation ratios ($\widehat{\Delta_i}$) produced by the random forest classifier for the same laws (y–axis). The correlation between the ground truth delegation ratios and the predicted delegation ratios is $r = 0.64$ suggesting that the predicted delegation ratios do very well in recovering the hand coded delegation ratios in Franchino (2001).

4.4 Measuring Delegation with PDR Scores

In addition to measuring delegation of authority, the random forest classifier that we trained allows us to create a measure of delegation for each provision within legislation which allows us to make statements about the probability that is delegates authority. Specifically, for any statute p in law i, the random forest classifier allows us to predict the probability that that a provision delegates authority given a document-term matrix $P(D_{ip}|X_{ip})$. The average over these probabilities for any given bill i is that bill's PDR, or probabilistic delegation ratio:

$$\frac{1}{P_i}\sum_{p=1}^{N_i} P(D_{ip}|X_{ip})\delta_{ip} \tag{10}$$

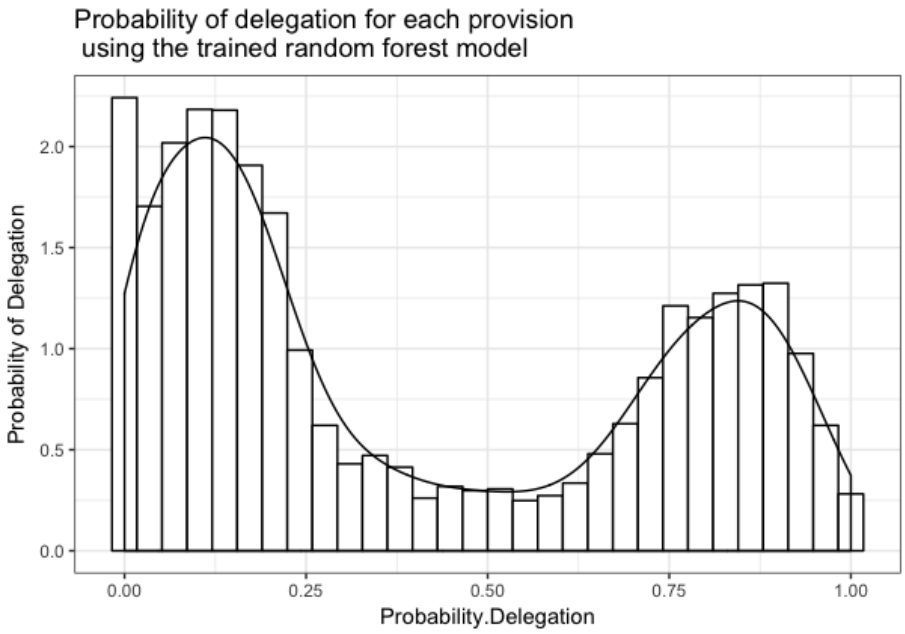

Figure 8: Distribution of probability of delegation in provisions from all 147 EU bills estimated by the random forest classifier

Figure 8 is the distribution of the estimated probability of delegation for each bill in the 147 EU laws used for these analyses. Since these probabilities are estimated with the random forest classifier, the bimodal pattern suggests that the classifier is able to accurately distinguish between provisions which delegate authority and those that do not.

One advantage that PDR scores have over delegation ratios is that they include information about the likelihood that each provision delegates authority. For example, a provision that is classified as delegating authority may have a probability of

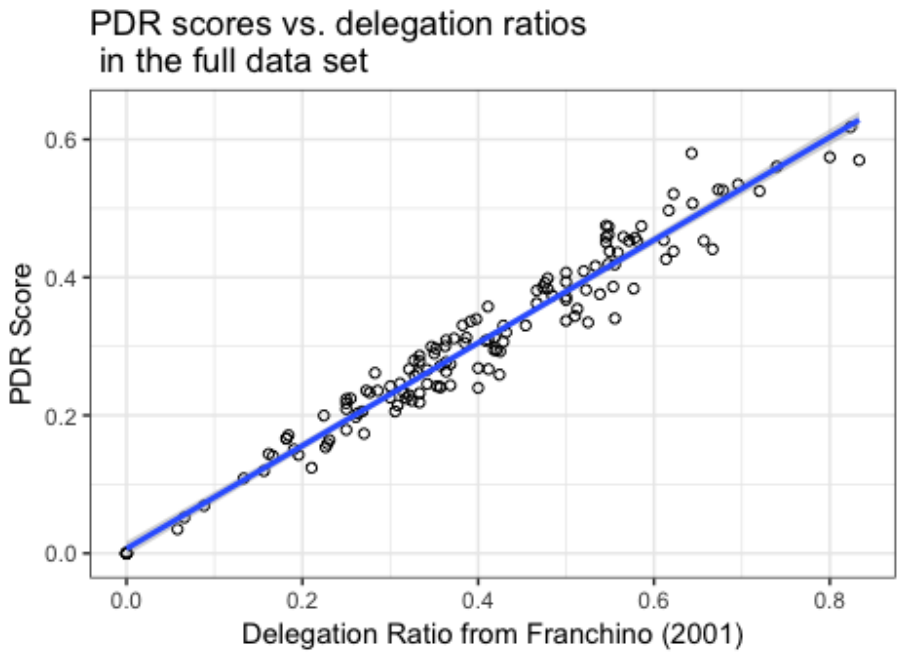

Figure 9: Estimated PDR scores v. ground truth delegation ratios in all 147 EU bills. $r = 0.95$

delegation equal to 0.99 or 0.51. We are clearly far more certain that the former provision delegates authority than we are about the latter position, yet the "classical" delegation ratio does not not allow us to incorporate this information. This becomes especially important when provisions are classified as delegating authority with higher levels of uncertainty.

For example, take a bill that that has 10 provisions, 3 of which are coded as delegating authority. The delegation ratio for this legislation is 3/10 or 0.33. But imagine that the languages of these provisions is not strongly tied to delegation and the probability of delegation for each provision in the bill is equal to 0.6. The PDR for this bill would thus be 0.18, which suggests that this bill provides weak

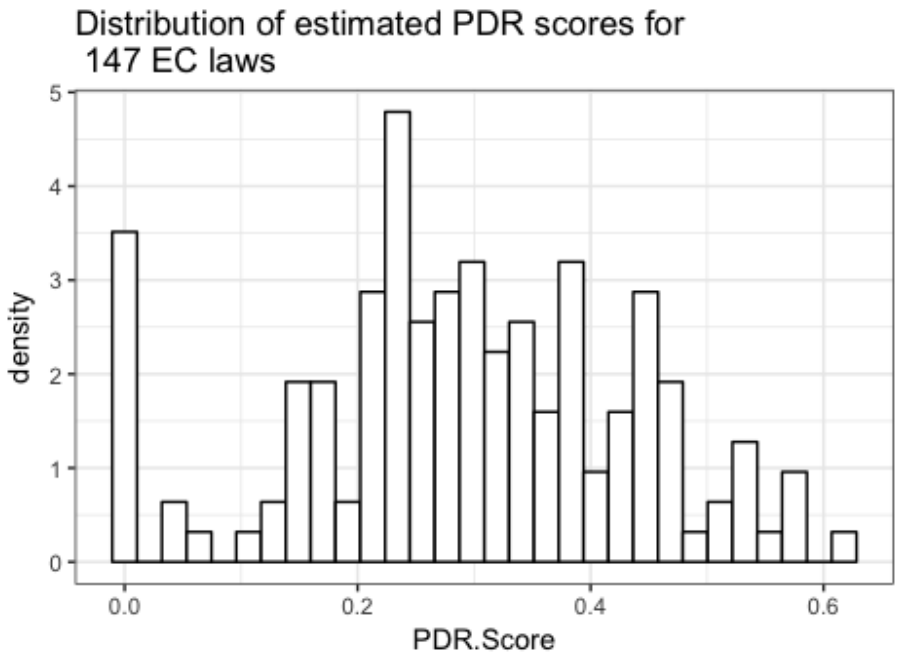

Figure 10: Distribution of estimated PDR scores in all 147 EU laws

information about delegation of authority, despite its relatively high delegation ratio. On the other hand, if the same bill has language strongly tied to delegation and each provision delegating authority has a probability of delegation equal to 0.95, then the PDR for this bill would be 0.285 which is much closer to its delegation ratio and suggests that we have much stronger evidence that this bill delegates authority.

Figure 9 is a plot of ground truth delegation ratios in each of the 147 EU laws versus estimated PDR scores. The high correlation suggests that they are measuring similar constructs, but as mentioned above the PDR scores contain more information about how sure we are the the provisions delegate authority.

Figure 10 is the distribution of estimated PDR scores in all 147 EU laws explored in these analyses.

5 Conclusion and Next Steps

We have provided an initial overview of a machine-learning approach to understanding delegation in the EU. Our approach, at this stage in the project, has examined one element of delegation, the locus of authority, which is particularly important in our context as it distinguishes between centralized authority and that left to member states. Replicating the leading coding framework in the literature (Franchino, 2001, 2004, 2007) provides some initially intriguing insights, such as the widespread use of optional language when reserving authority for member states.

The locus of authority in the EU context has typically been captured via delegation ratios, which capture the concentration of member state delegations within a particular law. Our automated content analysis framework admits a probabilistic alternative to this measure that has both face validity and advantages in understanding both non-random error in coding and substantive delegation. The next steps in our project will exploit these advantages.

References

Benoit, Kenneth, Michael Laver, and Slava Mikhaylov. 2009. "Treating words as data with error: Uncertainty in text statements of policy positions." *American Journal of Political Science* 53 (2): 495–513.

Blei, David M. 2012. "Probabilistic topic models." *Communications of the ACM* 55 (4): 77–84.

Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. "Latent dirichlet allocation." *Journal of machine Learning research* 3 (Jan): 993–1022.

Carvalho, Carlos M, Nicholas G Polson, and James G Scott. 2010. "The horseshoe estimator for sparse signals." *Biometrika* 97 (2): 465–480.

Clinton, Joshua D, Anthony Bertelli, Christian R Grose, David E Lewis, and David C Nixon. 2012. "Separated powers in the United States: The ideology of agencies, presidents, and congress." *American Journal of Political Science* 56 (2): 341–354.

Epstein, David, and Sharyn O'Halloran. 1999. *Delegating powers: A transaction cost politics approach to policy making under separate powers*. Cambridge University Press.

Franchino, Fabio. 2001. "Delegation and constraints in the national execution of the EC policies: a longitudinal and qualitative analysis." *West European Politics* 24 (4): 169–192.

Franchino, Fabio. 2004. "Delegating powers in the European Community." *British Journal of Political Science* 34 (2): 269–293.

Franchino, Fabio. 2007. *The Powers of the Union: Delegation in the EU*. Cambridge University Press.

Genkin, Alexander, David D Lewis, and David Madigan. 2007. "Large-scale Bayesian logistic regression for text categorization." *Technometrics* 49 (3): 291–304.

Hinton, Geoffrey E, and Ruslan R Salakhutdinov. 2009. "Replicated softmax: an undirected topic model." In *Advances in neural information processing systems*. pp. 1607–1614.

Hoffman, Matthew, Francis R Bach, and David M Blei. 2010. "Online learning for latent dirichlet allocation." In *advances in neural information processing systems*. pp. 856–864.

Huber, John D, and Charles R Shipan. 2002. *Deliberate discretion?: The institutional foundations of bureaucratic autonomy*. Cambridge University Press.

Huber, John D, Charles R Shipan, and Madelaine Pfahler. 2001. "Legislatures and statutory control of bureaucracy." *American Journal of Political Science*: 330–345.

Ikonomakis, M, S Kotsiantis, and V Tampakas. 2005. "Text classification using machine learning techniques." *WSEAS transactions on computers* 4 (8): 966–974.

Koh, Kwangmoo, Seung-Jean Kim, and Stephen Boyd. 2007. "An interior-point method for large-scale l1-regularized logistic regression." *Journal of Machine learning research* 8 (Jul): 1519–1555.

Kotsiantis, SB. 2007. "Supervised Machine Learning: A Review of Classification Techniques." *Informatica* 31: 249–268.

Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting policy positions from political texts using words as data." *American Political Science Review* 97 (2): 311–331.

McCann, Pamela J Clouser. 2016. *The Federal Design Dilemma*. Cambridge University Press.

Ng, Andrew Y, and Michael I Jordan. 2002. "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes." In *Advances in neural information processing systems*. pp. 841–848.

Ratkovic, Marc, and Dustin Tingley. 2017. "Sparse estimation and uncertainty with application to subgroup analysis." *Political Analysis* 25 (1): 1–40.

Roberts, Margaret E, Brandon M Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58 (4): 1064–1082.

Tibshirani, Robert. 1996. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)*: 267–288.

Tibshirani, Robert. 2011. "Regression shrinkage and selection via the lasso: a retrospective." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73 (3): 273–282.

Vakilifathi, Mona. N.d. "Constraining Bureaucrats Today Knowing You'll Be Gone Tomorrow: The Effect of Legislative Term Limits on Statutory Discretion." *Policy Studies Journal*.

Appendix

5.1 Overview of Topic Models

Below we describe how the topic model structures texts. As with all text analysis problems which we discussed above, the fundamental unit of data used in topic models are *terms* as represented in the document–term matrix. Terms are treated as items from a *vocabulary*, indexed by a set of numbers $\{1, ..., V\}$. The vocabulary are all of the terms in a given *corpus* or collection of documents as discussed above.

A *document* is a bag of N terms. We describe a document as a "bag of terms" rather than a series or sequence of terms in a particular order because the topic model does not take the order of terms or words into account. These N terms can be represented by a vector $\mathbf{w} = (w_1, w_2, ..., w_N)$. A *corpus*, as above, is a collection of M documents which can be represented by $D = \{\mathbf{w_1}, \mathbf{w_2}, ..., \mathbf{w_M}\}$. The topic model treats each document within a corpus as a mixture of a fixed number of k latent topics which is represented by a distribution over words.

5.2 Modeling Provisions in EU Legislation

The essential first step toward modeling any set of texts using the LDA is division of these texts into *corpora* and *documents*. For our purposes, we define:

- **Document** – A provision a within a piece of EU legislation d, is represented as a sequence of N words $\mathbf{w} = (w_1, w_2, ..., w_N)$.

- **Corpus** - The collection of all 7,011 provisions within the labeled EU legislation

collected between 1958–2000 by (Franchino, 2001).

The LDA is a generative probabilistic model of the corpus of EU provisions treated as a random mixture over K latent topics while each topic as a distribution over the words. But how can we know how many topics a corpus contains? Because the LDA does not automatically select the number of topics that a corpus is comprised of, the researcher must decide on the basis of a number of factors how many topics they believe a corpus is divided into. One popular method used to choose the "optimal" number of topics is to estimate a topic model using $K = \{2, \cdots, n\}$ topics, measure the perplexity of each model and choose the model for which the marginal perplexity stops decreasing (Hinton and Salakhutdinov, 2009). Perplexity is an information theoretic metric used to measure how well probability models predict a sample which we describe in further detail below.

Lower values of perplexity imply models that better fit the data. While perplexity often provides a good means of guiding researchers, many argue that it should only be used as a guide rather than the sole means of choosing the appropriate number of topics (Hoffman, Bach, and Blei, 2010). In most cases, theoretical guidance given the problem at hand and the extent to which the model can be readily interpreted by a human should also be considered in addition to perplexity.

As mentioned above, the first cut of this data involved arbitrarily setting $K = 10$. In words, this implies that the set of provisions that we explore can be categorized into a total of 10 latent thematic elements and each of the provisions is comprised of a mixture of these thematic elements. For example, we might discover that "global warming" and "EU financing for member states" are two topics among the 10 that

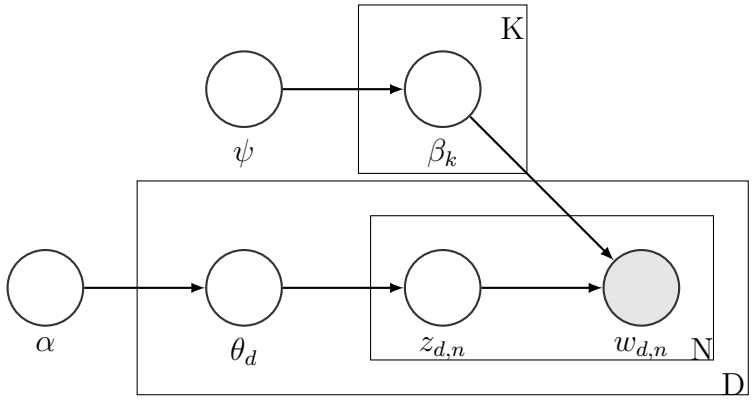

Figure 11: Graphical Representation of Latent Dirichlet Allocation Applied to EU Provisions Using Plate Notation

were estimated. If a provision is about financing research for global warming, this provision might appear in the topic model as containing roughly 50% of content related to the "global warming" and 50% of the content related to "EU financing for member states."

Figure 11 is a graphical model of the LDA as applied to our corpus of provisions using plate notation to denote replicates of the provisions A and the words within each provision N. Each of the nodes represents a random variable. The only observed variable is the collection of words which comprise the set of all provisions. All other variables are unobserved latent variables which are estimated by the LDA. The graphical model above assumes that $w_{d,n}$, each word in each document (provision) in the corpus (all provisions in 147 English–language laws) is generated from both a distribution over latent topics and a distribution over words.

We define:

1. $\beta_k \sim Dir(\psi)$, where $k \in \{1, ..., 10\}$ - the distribution over words that defines

each of the $K = 10$ latent topics assumed to encompass the provisions.

2. $\theta_{ad} \sim Dir(\alpha)$, where $d \in \{1, ..., 7011\}$ and $a \in \{1, ..., A\}$ - the distribution over topics for each provision.

3. $z_{ad,n}$ - topic assignment of the nth word in the ath provision for the dth law.

4. $w_{ad,n}$ - the nth word of the ath provision in the dth law.

The probability distributions of topic proportions for each provision $p(\theta_d|\alpha)$ and of each topic in all provisions $p(\beta_k|\psi)$ are Dirichlet with hyperparameters α and ψ respectively. Thus topic proportions in a news outlet d has the distribution:

$$p(\theta_{ad}|\alpha) = \frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma\big(\sum_{i=1}^{K} \alpha_i\big)} \prod_{i=1}^{30} \theta_{adi}^{\alpha_i - 1}$$

And each topic k across all articles has the distribution over words:

$$p(\beta_k|\psi) = \frac{\prod_{i=1}^{N} \Gamma(\psi_i)}{\Gamma\big(\sum_{i=1}^{N} \psi_i\big)} \prod_{i=1}^{N} \beta_{ki}^{\psi_i - 1}$$

The remaining distributions that we need in order to specify the model including topic assignment conditional on topic distribution $p(z_{ad,n}|\theta_{ad})$ and word conditional on topic assignment $p(w_{ad,n}|z_{ad,n}, \beta_k)$ are multinomial with:

$$z_{ad,n} \sim Multinom(\theta_{ad}) \tag{11}$$

$$w_{ad,n} \sim Multinom(\beta_k) \tag{12}$$

Putting all this together, we arrive at the fully specified model over all provisions:

$$p(\theta, \mathbf{z}, \mathbf{w}, \beta|\psi, \alpha) = \prod_{k=1}^{30} p(\beta_k|\psi) \prod_{a=1}^{A} \left(p(\theta_d|\alpha) \prod_{n=1}^{N} p(z_{d,n}|\theta_d) p(w_{d,n}|z_{d,n}, \beta_k) \right) \tag{13}$$

Estimating θ_d, which we use to explore provisions which delegate executive authority to EU member states and all other relevant hidden parameters requires posterior inference using the variational expectation-maximization algorithm (VEM) algorithm (Blei, Ng, and Jordan, 2003) which is implemented in **R** packages such as *topicmodels* and *lda*.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
1	regul	measur	institut	account	nation	product	decis	pani	mission	employ
2	direct	group	petent	amount	certif	price	author	undertak	accord	legisl
3	mission	may	benefit	valu	activ	market	mission	person	procedur	insur
4	provis	condit	author	tax	requir	good	undertak	law	laid	person
5	period	particular	resid	may	train	muniti	inform	capit	council	worker

Table 3: Estimated topics for EU provisions between 1958–2000 using a $K = 10$ topic model. Variational inference was used to estimate model parameters.