Legislative Bellwethers: The Role of Committee Membership in Parliamentary Debate — Supporting Information

Jorge M. Fernandes* Max Goplerud† Miguel Won‡

October 2, 2018

A ‘Motivations’ Behind Bills in Europe

In many European parliamentary democracies, proposed legislation contains a motivation section, in which the actor introducing the piece of legislative explains the rationale, goals, and context, and reasons why it should be approved. To show that our method of classifying speeches can travel cross-nationally, we present examples of numerous European democracies where such structure exists. For each country, we present four links to parliamentary archives.

Austria

Each bill contains a motivation section entitled *Begründung*.

- Austria - Example 1
- Austria - Example 2
- Austria - Example 3
- Austria - Example 4

*Institute of Social Sciences, University of Lisbon, Av. Prof. Anibal Bettencourt, 9, Lisbon, Portugal
†Department of Government, Harvard University, 1737 Cambridge Street, Cambridge, MA, USA
‡INESC-ID, Rua Alves Redol, 9, Lisbon, Portugal

Belgium

Each bill contains a motivation section entitled *Exposé des Motifs*.

- Belgium - Example 1
- Belgium - Example 2
- Belgium - Example 3
- Belgium - Example 4

Denmark

Each bill contains a motivation section entitled *Bemærkninger til lovforslaget*.

- Denmark - Example 1
- Denmark - Example 2
- Denmark - Example 3
- Denmark - Example 4

France

Each bill contains a motivation section entitled *Exposé des Motifs*.

- France - Example 1
- France - Example 2
- France - Example 3
- France - Example 4

Germany

Each bill contains a motivation section entitled *Begründung*.

- Germany - Example 1
- Germany - Example 2

- Germany - Example 3
- Germany - Example 4

Italy

Each bill contains a motivation section entitled *Relazione Illustrativa*.

- Italy - Example 1
- Italy - Example 2
- Italy - Example 3
- Italy - Example 4

Spain

Each bill contains a motivation section entitled *Exposición de Motivos*.

- Spain - Example 1
- Spain - Example 2
- Spain - Example 3
- Spain - Example 4

B Data Preprocessing and Description

We preprocessed both bills and speeches using standard procedures by taking all bi-grams (i.e. all words and all pairs of words) and excluding low-frequency words that constituted fewer than 0.5% of all speeches. Further, we excluded highly common words that occurred in over 90% of speeches. Both were done to prevent potential over-fitting when specifying our model insofar as the inclusion of very rare words might have good prediction when training the model but perform badly when we attempt to extrapolate to speeches. This was followed by a cleaning process where all stopwords (common but uniformative words, e.g. "the" in English), digits, common Portuguese person names, and time-words (weekdays and months) were removed. Finally, we exclude all documents shorter than fifty words to ensure that short documents (e.g. short procedural speeches seeking recognition) are not driving our analysis. This preprocessing of documents is fairly standard for text analysis.[1] Before proceeding to the analysis, we provide some descriptive statistics on the texts. Table 1 summarizes the speech corpus. After our preprocessing procedure above, we have about fifty thousand speeches (three million phrases) with a vocabulary of 7317 phrases (words and bigrams).

Table 1: Speech Corpus Descriptive Statistics

English Party Name (Portuguese Abbreviation)	Number of Speeches	Number of Legislators
Left Bloc (BE)	7744	24
CDS - People's Party (CDS-PP)	7562	69
Portuguese Communist Party (PCP)	8663	37
Ecologist Party "The Greens" (PEV)	3751	6
Socialist Party (PS)	11200	313
Social Democratic Party (PSD)	10039	309

Next, Table 2 shows the descriptive statistics for the number of bills per standing committees in the Portuguese legislature.

Table 2: Bills Per Legislative Committee 2000-2015

Committee	Number of bills
Agriculture and Fisheries	207
Constitution, Civil Rights and Liberties	1011
Culture	210
National Defense	99
Economy and Public Works	713
Education and Science	578
Foreign Affairs	418
Ethics, Society and Assembly House Keeping	28
European Affairs	122
Public Administration and Budget	578
Environment, Territory and Local Government	1017
Health	407
Labor and Social Security	782
Total	6170

[1] Appendix D shows the results for our cross-validation when we vary the preprocessing procedure. The performance is effectively unchanged by different procedures, and thus we rely on the one noted above as it is fairly 'standard' and hopefully protects against over-fitting and poor performance when we extrapolate to the speeches.

C Formal Discussion of the SVM

This section briefly outlines the SVM in more detail. For notation, use $y_j \in \{-1, 1\}$ to indicate the class of our training data (bills). Our covariates x_j is a vector of counts of words, i.e. x_j has as many entires as the number of distinct bi-grams used in our analysis. We adjust these counts using 'tf-idf' (term frequency; inverse document frequency with ℓ_2 normalization) weighting to adjust for documents of different length and the fact that certain words occur much more often than others.[2]

To define the weighting more precisely, recall that we have V phrases (words and bigrams) in our vocabulary and $x_{j,v}$ indicates the number of times phrase v occurs in document j. Inverse-document frequency calculates a scaled measure of how many documents each term appears in. Define N as the total number of documents, *idf* is thus defined as:

$$idf_v = \ln \frac{N}{\sum_{j=1}^{N} I(x_{j,v} > 0)}$$

This means that rarer words will have larger idf_v, as the number of documents they occur in in smaller and thus argument of the logarithm is larger. This is designed to ensure that the rarer words are not 'swamped' by the larger words in the classification procedure. We combine the two weights using a simple multiplication, i.e.

$$tf.idf_{j,v} = tf_{j,v} \cdot idf_v$$

It is worth pausing to note the dimensionality of x_j; we have 7317 features (words and bigrams) and only 6170 documents. This would cause traditional regression frameworks to fail to generate reasonable predictions (as we have more variables than observations), and thus a turn towards machine learning approaches that enforce some regularization or penalization to solve this problem is required. After having calculated the $tf.idf$ measure, a common next step is to standardize each document's vector by its ℓ_2 norm to ensure that documents are broadly comparable even if they

[2]Appendix D shows the performance of our classifier with and without tf-idf weighting.

differ in length. This is the default option in `sklearn` and we follow it here. Formally,

$$\tilde{tf.idf}_{j,v} = \frac{tf.idf_{j,v}}{\sqrt{\sum_{v=1}^{V}[tf.idf_{j,v}]^2}}$$

Given this setup, we then attempt to find a set of weights or coefficients on each word $\mathbf{w}$ with a bias term or intercept b, where the minimum ranges over all of our bills in the training set N_j (Bishop, 2006). This is a difficult problem to solve in this form. The traditional interpretation of an SVM reformulates this in terms of the 'dual' formulation replacing the objective function a Lagrangian formulation. It turns out that in this dual function most of the Lagrange multipliers are zero; the prediction thus only relies on the observations j with non-zero Lagrange multipliers. These are the 'support vectors', and hence the name of the classification procedure. In the dual formulation, the role of C—the penalty for mis-classified items—is stated explicitly. Formally,

$$\mathbf{w}^*, b^* = \arg\max_{\mathbf{w},b} \frac{1}{||\mathbf{w}||} \min_{j\in\{1,\cdots,N_j\}} \left[y_j(\mathbf{w}^T\mathbf{x}_j + b)\right] \tag{1}$$

We then use the estimated weights $\mathbf{w}$ and bias term b to predict whether some new document, a speech s_i on the floor, is in the agriculture committee or not. This is done by evaluating $\mathbf{w}^T s_i + b$ and seeing whether it is positive or negative. If positive, s_i is classified as being from the agriculture committee, otherwise it is not. This corresponds to the SVM looking for whether something is on the 'positive' side of the decision boundary (at 0) and classifying accordingly.

The final step in our procedure is to go from a series of binary classifications, e.g. is bill j from the agriculture committee or not, to a classification of each bill into a committee jurisdiction. There are again multiple ways to do this. We rely on the common 'one-vs-the-rest' scheme by estimating one binary classification for each committee. We calculate the predicted $\mathbf{w}^T\mathbf{x}_j + b$ for each document and then assign each document to the class that has the largest value. This is roughly interpretable as saying: Assign bill j to the committee where it is the largest distance from the decision boundary.[3]

D Alternative Specifications of the SVM

As noted above, we made a number of pre-processing decisions when training our classifier. We show the F-statistic (averaged across jurisdictions) as in Table 1 using multiple different preprocessing conditions. Condition (1) is the one reported in the main text and the rest of our analysis. Conditions (2)-(6) permute various aspects of the pre-processing. Other combinations are possible, but these constitute a set of very common alternatives and thus sufficiently test the robustness of our procedure. 'Full' stopwords include various names, times, and dates alongside standard Portuguese stopwords. 'Limited' include only the standard Portuguese stopwords. The conditions are:

Table 3: Preprocessing Conditions

Number	n-grams	Min Threshold	Max Threshold	Stopwords	$tf - idf$
1	bigrams	0.5%	90%	Full	Yes
2	unigrams	0.5%	90%	Full	Yes
3	bigrams	0%	100%	Full	Yes
4	bigrams	0.5%	100%	Full	Yes
5	bigrams	0%	90%	Full	Yes
6	bigrams	0.5%	90%	Limited	Yes
7	bigrams	0.5%	90%	Full	No

When validating these conditions, we also examine the penalty C on misclassified observations. We examine $C \in [0.1, 0.5, 1, 2.5, 5]$, where $C = 1$ is a common default that we use in the results in the main text. To show the results, we consider two versions of the F-score. The first, used in the main text, averages the F-scores across jurisdictions. However, one might want to ensure that jurisdictions that are more common in 'truth' are given higher weight and thus do a weighted average of the F-scores based on the frequency of each class. These results are from the held-out training data on *bills* using all words.

We see that, across all of the measures, the differences, especially with $C = 1$ are extremely

[3]Note, that when all values are negative, the interpretation shifts slightly to be where is the bill the closest to being classified as a positive example.

Table 4: Validating SVM: Average of F-Scores

Type	C				
	0.1	0.5	1	2.5	5
1	0.795	0.817	0.809	0.812	0.814
2	0.778	0.808	0.815	0.807	0.800
3	0.746	0.817	0.821	0.825	0.822
4	0.795	0.817	0.809	0.812	0.814
5	0.746	0.817	0.821	0.826	0.822
6	0.791	0.816	0.825	0.818	0.817
7	0.774	0.804	0.823	0.807	0.798

Table 5: Validating SVM: Weighted Average of F-Scores

Type	C				
	0.1	0.5	1	2.5	5
1	0.856	0.871	0.864	0.855	0.856
2	0.846	0.864	0.860	0.854	0.854
3	0.843	0.867	0.873	0.876	0.872
4	0.856	0.871	0.864	0.855	0.857
5	0.843	0.867	0.873	0.876	0.872
6	0.850	0.868	0.868	0.858	0.858
7	0.840	0.860	0.862	0.854	0.847

slight. This gives us confidence that our results from the SVM are not being driven by the particular type of preprocessing. We rely on the set of conditions noted above as they are quite standard and we think that some of them (e.g. including the extra-stop words) will ensure stronger performance when extrapolating to speeches.

Next, we compare the cross-validation F-score of the SVM against the scores from four other specifications: Logistic Regression with a LASSO penalty,[4] Logistic Regression with a Ridge Penalty,[5] and a Random Forest classifier. The regularization penalties for the first two are chosen using cross-validation; the Random Forest consists of twenty estimators with default settings in sk-learn. We also test a simple neural network with one hidden layer of 100 neurons and a hyperbolic tangent activation; training and regularization are set to the defaults in sk-learn.

Table 6 shows the results (weighted and unweighted) for the classifiers in both the bill and speeches datasets. While many of the models have similar performance on the F-scores for bills (e.g. LASSO, Ridge, and the Neural Network are broadly similar to the SVM), the SVM does

[4]That is, multinomial logistic regression with a l_1 penalty on the coefficients.
[5]That is, multinomial logistic regression with a l_2 penalty on the coefficients.

Table 6: Comparison of Other Models

(a) Weighted F-Score

Method	Speeches	Bills
SVM	0.803	0.864
LASSO	0.752	0.850
Ridge	0.790	0.858
Random Forest	0.546	0.760
Neural Net	0.723	0.857

(b) Averaged F-Score

Method	Speeches	Bills
SVM	0.729	0.809
LASSO	0.617	0.798
Ridge	0.667	0.771
Random Forest	0.452	0.677
Neural Net	0.656	0.801

Note: 'Weighted F-Score' averages the individual class F-scores by the size of the class in the labelled data. 'Average F-Score' is the F-score without weighting by class size. 'Speeches' refers to the F-score on the hand-annotated speeches discussed in the main text; 'Bills' refers to the F-score on the 80/20 split from the cross-validation discussed above.

better by a starker margin on speeches—although Ridge is fairly close.

The random forest does markedly worse which is perhaps a product of over-fitting on the training data. The simple neural network does slightly worse than LASSO and Ridge. Overall, the fact that the SVM does better or about the same as the other methods gives us confidence that we are relying on a reasonable supervised procedure for our main analysis. As noted earlier, we do not claim that our SVM is the best or only procedure for our approach, but the results in this section give us confidence that our procedure performs well using a cross-validation approach.

E The Drivers of Classification Accuracy

To understand better how the SVM generates predictions, we extract the top 10 words for each jurisdiction (i.e. the most positive coefficient) and report them in the table below. These words are typically nouns that clearly relate to the topic at hand. They are also common words that would occur in speech versus legalistic jargon.

All of the results above are from using the entire text of the bills where the vocabulary of words (and bigrams) is defined by the words and bigrams that occur in the corpus of speeches. However, as we noted before, Portuguese legislation (as with legislation in most other parliamentary democracies) consists of two parts:[6] (1) a 'motivations' section where the bill proposer outlines the political rationale behind the legislation; (2) a technical section that is the actual statute.

[6]See Appendix A for a discussion of the cross-national prevalence of this data.

Table 7: Most Predictive Words

AGRICULTURE	CONSTITUTION	CULTURE	DEFENSE	ECONOMY	EDUCATION	FOREIGN AFFAIRS
fisheries	justice	culture	national defense	mobility	education	agreement
agricultural	PSP (police)	communication	military	tourism	teachers	convention
farmers	policemen	RTP (public broadcast)	maritime	economy	scholarships	foreign affairs
forest	GNR (police)	cultural	military	contos (previous Portuguese currency)	school	BE (political party)
Douro (region)	gender	radio	defense	tolls	school sport	minister
fish	referendum	signal	armed forces	energy	sport	affairs
agriculture ministry	judicial	heritage	arms	contracts	scholar	excellency
marketing	civil	internet	viable	connection	associations	visit
lands	order	garden	federation	customers	elementary education	emigration
production	actual	landmark classification	lines	oil	higher education	presidency minister
ETHICS	EUROPEAN AFFAIRS	BUDGET	ENVIRONMENT AND TERRITORY	HEALTH	LABOR	
family	european affairs	budget	waste	hospital	labor	
card	European (people)	financial	local government	medicine	job	
women	european	public servants	environment	health	labor market regulation	
pain	affairs	budget (public)	water	M.D. (plural)	social	
familiar	business	balance	buildings	M.D. (single)	workers	
audiences	foreign policy	taxes	rent	patients	social security	
usefulness	european council	supervision	parish	techniques	professional	
representative	republic	loans	village	co-payment	pension	
public services	external	housing	assemblies	sick people	unemployment	
parenthood	union	reconciliation	municipal	care	institutions	

We note that the motivations section can be conceptualized as an extended political 'speech' and thus the existence of this data likely helps comparability between the bill texts and the speech data—allowing us to achieve the high accuracy found above.

This section shows, however, that there is informative content in the legislation itself—and the results are not wholly driven by the motivations section. We estimate three models based on three different types of training data in the bills: (1) All Words (i.e. replicating the above analysis); (2) Motivation Only (i.e. only words that appear in the 'motivations' section are extracted from each bill); (3) Technical Only (i.e. only words that appear in the 'technical' section are extracted).[7]. Table 8 shows the F-scores for each model for the cross-validation on the bills and the validation against the human coding.[8] The results are as expected: Using all of the information (Motivations and Technical Provisions) results in an unambiguously better classifier on the speeches and for bills using the 'unweighted' F-score (i.e. simply the F-score of all documents) and for the 'weighted' F-score (i.e. weighted by the observed proportion of classes in the 'true' data).[9] It is worth noting, however, that there is a marked improvement by including the technical language rather than relying solely on the motivations section. If, however, we only had the technical language to rely on, the performance is indeed weaker, however, at least with respect to the weighted measure, it is not unacceptably bad. We would caution, therefore, in countries where no analogue to the 'motivation' information exists, that scholars hand annotate a collection of speeches to verify that their model has sufficient accuracy to be used safely to classify speeches.

A plausible interpretation of this is that there may be some rarer classes that the model does less well at prediction, but for the common types of bills and speeches, it does well. Further, it is also worth noting that the performance on speeches is somewhat lower than on bills; as the two

[7]Note that not all introduced legislation has a technical section as 'resolutions', proposals for the government to act on a particular issue, have only a motivations section. Thus, the sample size for the classifier is smaller in (3) that likely explains at least some of its partially degraded performance

[8]We show results from the models using the same combination of stopwords, tf-idf usages, etc. for the main models—see D a discussion of cross-validation on these metrics. The C is set to 1 from our validation exercise on the SVM for predicting bills. This value performs the best for the models using all words and technical words; the F-score can be slightly improved for the 'Motivations Only' model by setting $C = 0.5$ to 0.78 for weighted and 0.751 for the unweighted. It is ambiguous whether it is permissible for us to 'tune' the classifier based on the performance on the speech dataset, so we adopt a stricter test of using the model selected by cross-validation only on the bills.

[9]As we show in cross-validation in the Appendix, the 'weighted' F-score shows only fairly small variance to changes in the model for validation on other bills

documents come from different domains, this is to be expected but as we showed above and here, an F-score of 0.82 is good for our purposes.[10] Whilst we engaged in some searching over classifiers and feature engineering, it is worth noting that a future extension of this research would likely see whether more sophisticated models (e.g. bagging or boosting of the 'Motivations' and 'Feature' only classifiers might do better than a single model taken on its own).

Table 8: Comparison of Data Sources

(a) Weighted F-Score

Data Type	F-Score for Bills	F-Score for Speeches
All Words	0.867	0.824
Motivations Only	0.863	0.746
Technical Only	0.869	0.633

(b) Averaged F-Score

Data Type	F-Score for Bills	F-Score for Speeches
All Words	0.825	0.772
Motivations Only	0.814	0.704
Technical Only	0.771	0.530

Note: 'Weighted F-Score' averages the individual class F-scores by the size of the class in the labeled data. 'Average F-Score' is the F-score without weighting by class size.

F Jurisdiction Salience

As noted in the main text, the most salience portfolio for each party at each election is shown in Table 9.

Table 9: Most Salient Jurisdictions

Party Abbrev.	Election Date				
	1999	2002	2005	2009	2011
PEV	NA	Territory	Territory	Territory	Economy
BE	Constitution	Budget	Health	Health	Economy
PCP	Health	Budget	Economy	Health	Economy
CDS	Health	Budget	Budget	Economy	Health
PSD	Health	Budget	Budget	Constitution	Economy
PS	Economy	Economy	Economy	Economy	Economy

'NA' indicates that no manifesto was scored by the Comparative Manifesto Project at this election.

[10] More broadly, it is also worth pointing out that our classifier being inaccurate is likely to make our tests more conservative: Assume that some speeches that are 'truly' from committee c are mis-classified as committee c'. That means that any effect of being on the committee will be attenuated as, for example, we might conclude that a legislator did not speak at all on a topic when they did but it was mis-classified.

G Alternative Empirical Specifications

As noted in the main text, we ran our results using a less causally robust procedure with separate legislator and committee fixed effects (e.g. $\alpha_i + \gamma_c$) instead of legislator-Committee fixed effects (α_{ic}). The results are shown here. The results agree with what we presented in the main text, although the effects are noticeably larger in magnitude and statistical significance.

Table 10: Modeling the Proportion of Speeches Allocated To Each Jurisdiction

	(1)	(2)
Member	0.244***	0.163***
	(0.007)	(0.014)
Salience		-0.018***
		(0.004)
Member X Salience		0.102***
		(0.011)
Government		-0.004**
		(0.002)
Member X Government		0.034***
		(0.012)
Member X Major Party		0.077***
		(0.016)
Number of Observations	21411	21372
Number of MPs	759	759

Note: *** : $p < 0.01$; ** : $p < 0.05$; * : $p < 0.1$. Standard errors shown in parentheses. The dependent variable is the proportion of time a legislator allocates to a jurisdiction (i.e. what proportion of the total speeches that a legislator gives in legislature l are in committee c?). Standard errors are clustered on the legislator-committee. All models include the following controls (not shown): separate legislator and committee fixed effects, legislature fixed effects, and other controls noted in the main text. Note that for Model (2), the lower order effect 'Major Party' cannot be included as no individuals switch party in the period under observation.

Another way to model a binary dependent variable with fixed effects is the fixed effects logistic regression (Chamberlain's Logit). The results are shown below:

We also show that our results are robust to using the percent of words allocated to each jurisdiction. The two measures are correlated extremely highly (0.97). Table 13 presents the core regression from Table 3 and shows that the point estimates and standard errors are very similar.

To address issues with pre-trends, we collapsed the data in a different way. We calculated our two variables (proportion of speeches and any speech) month-by-month (in contrast to our legislature-by-legislature analysis in the main paper). To analyze whether there are pre-trends that would mean that assignment to the committee is not the key variable, we take each spell of

Table 11: Modeling Whether a Legislator Participates At All

	(1)	(2)
Member	0.364***	0.232***
	(0.010)	(0.021)
Salience		-0.021**
		(0.008)
Member X Salience		0.029*
		(0.015)
Government		-0.031**
		(0.013)
Member X Government		0.063***
		(0.018)
Member X Major Party		0.130***
		(0.024)
Number of Observations	21411	21372
Number of MPs	759 759	

Note: $^{***}: p < 0.01$; $^{**}: p < 0.05$; $^{*}: p < 0.1$. Standard errors in parentheses. The dependent variable is the extensive margin (i.e. does a legislator participate at all in jurisdiction c?). Standard errors are clustered on the legislator-committee pair. All models include the following controls (not shown): separate legislator and committee fixed effects, legislature fixed effects, and other controls noted in the main text. Note that for Model (2), the lower order effect 'Major Party' cannot be included as no individuals switch party in the period under observation.

Table 12: Alternative Fixed Effects Specifications (Any Participation)

	FE-OLS	FE-Logit
Member	0.166	1.149
	[0.117, 0.216]	[0.725, 1.573]
Salience	-0.002	0.019
	[-0.028, 0.025]	[-0.185, 0.223]
Member X Salience	-0.011	-0.095
	[-0.054, 0.032]	[-0.446, 0.255]
Government	-0.033	-0.274
	[-0.049, -0.017]	[-0.400, -0.148]
Member X Government	0.083	0.495
	[0.039, 0.126]	[0.113, 0.877]
Member X Major Party	0.016	0.185
	[-0.047, 0.079]	[-0.325, 0.696]
Number of Observations	21372	21372
Number of Legislators	759	759

Note: $^{***}: p < 0.01$; $^{**}: p < 0.05$; $^{*}: p < 0.1$. The first model replicates the results from Table 4. The second model ('FE-Logit') runs a fixed effect logistic regression.

committee assignment (i.e. a period of not being assigned, an assignment at time m, and then all months until the MP exits that committee) and plot the de-meaned activity for that period. Our expectation is as follows: There should be a large discontinuous jump at time '0' when the MP is first assigned to the committee and there should be no evidence of trending upwards in the months

Table 13: Alternative Dependent Variable (Words vs Speeches)

	Speeches (%)	Words (%)
Member	0.104	0.110
	[0.080, 0.128]	[0.085, 0.136]
Salience	-0.018	-0.020
	[-0.027, -0.010]	[-0.029, -0.011]
Member X Salience	0.043	0.052
	[0.020, 0.066]	[0.028, 0.077]
Government	-0.002	-0.003
	[-0.007, 0.003]	[-0.008, 0.003]
Member X Government	0.021	0.022
	[-0.004, 0.045]	[-0.004, 0.049]
Member X Major Party	0.007	0.009
	[-0.025, 0.039]	[-0.024, 0.043]
Number of Observations	21372	21372
Number of Legislators	759	759

Note: 95% confidence intervals are shown below each coefficient. Clustered-standard errors are used in both models. The first model replicates the results from Table 3. The second model uses the percent of words allocated to each jurisdiction.

immediately prior to assignment. The dots are sized by the number of observations that occur in each period. We see that there is no evidence of a pre-trend in the two years before for whether a MP participates at all. When looking at the proportion, there is similarly limited evidence of pre-trends: The one possible exception is that amongst the very limited number of speeches that occur in the weeks immediately before appointment. However, as most committee appointments are carefully managed and planned things by the party leadership, this does not undermine our conclusions as the threat to inference is MPs using the *long* period before their appointment by speaking more on a particular topic to gather expertise or impress the leadership and thereby be appointed to the committee. This concern is clearly unsupported by examining the trends in the raw data.

Figure 1: Trending in the Outcome Variable