Supplemental Information for

Setting the Committee Agenda: Measuring Speaker Influence in Congressional Hearings

Joshua Y. Lerner Robert Shaffer

November 9, 2020

Contents

1 SITS Model Specification and Robustness

1.1 Preprocessing

Before fitting the models described in-text, we conducted a series of standard preprocessing steps. For each Congress in our dataset, we converted all characters to their lower-case representations and removed punctuation, stopwords, and words that occurred in greater than 95% of statements or fewer than 200 statements in that Congress. Since the Congress with the fewest statements contained approximately 4.3 million statements, the lower bound in our prepocessing setup is approximately equivalent to removing tokens that occurred in fewer than 0.005% of statements. This cutoff is conservative relative to those investigated elsewhere in the literature (see, e.g., Denny & Spirling 2018), and is designed to ease the computational burden imposed by the models we fit without affecting our substantive results.

1.2 Parameter settings

To measure our key dependent variable we rely on the parameteric Speaker Identity for Topic Segmentation (SITS) model (Nguyen *et al.* 2012, 2014).[1] Specifically, for each Congress in our dataset, we fit independent parametric SITS models, with the topic parameter set to 100. All other parameters are left at their default settings. Specifically, we used Dirichlet topic- and word-concentration parameters of $\alpha = \beta = 0.1$, and a symmetric Beta topic shift parameter of $\gamma = 0.25$. During estimation, we use 5000 total iterations with a 2500 iteration burn-in period.

[1]Implemented at https://github.com/vietansegan/sits.

2 Influence descriptives

In this section, we include additional descriptive statistics for our measure of influence. We include a table summarizing the most influential members of the Senate, which parallels the top ten House members from Table 1 in the paper. We then compare our measure of speaker influence to a simple volume-of-speech measure to show the absence of a real relationship between the two.

2.1 Senate top ten influence scores

Table 1 shows the parallel set of top members for the United States Senate to Table 1 from the main paper for the House. Again, this table shows the top scoring members of the Senate averaged over all sessions— the 108^{th} through the 115^{th}. We limit our set to those members who spoke in multiple hearings in multiple Congresses.

Table 1: Highest Career Scoring Members of the Senate

Rank	Member of Congress	Score
1	Daniel Inouye	0.0699
2	Jeff Bingaman	0.0589
3	Herbert Kohl	0.0523
4	Michael DeWine	0.0469
5	Robert Byrd	0.0466
6	Conrad Burns	0.0449
7	Tim Johnson	0.0444
8	Mitch McConnell	0.0427
9	John D. Rockefeller	0.0417
10	Daniel Akaka	0.0407

Averages for members over the Congresses they served in. Limited to people who participated in hearings in multiple Congresses.

Similar to Table 1 in the main paper, we find an overwhelmingly older and male set of Senators in the top ten. The only non-White Senators are Senators Inouye and Akaka, who are both Asian/Pacific Islander. Like Congressman Clyburn in the House, Inouye served in the same chamber of Congress for decades, and occupied a variety of leadership roles in the Democratic Party. As a Medal of Honor recipient and the chair of five separate Senate committees, Inouye was a respected war veteran and a consummate insider, and offers a clear example of the kind of socially and politically influential member that we would expect to be an effective committee participant. The other members all fit the same pattern, and represent "elder statesmen" and/or members of the party leadership.

2.2 Speaker frequency

Figure 1: Comparing Mean Influence with Speech Volume

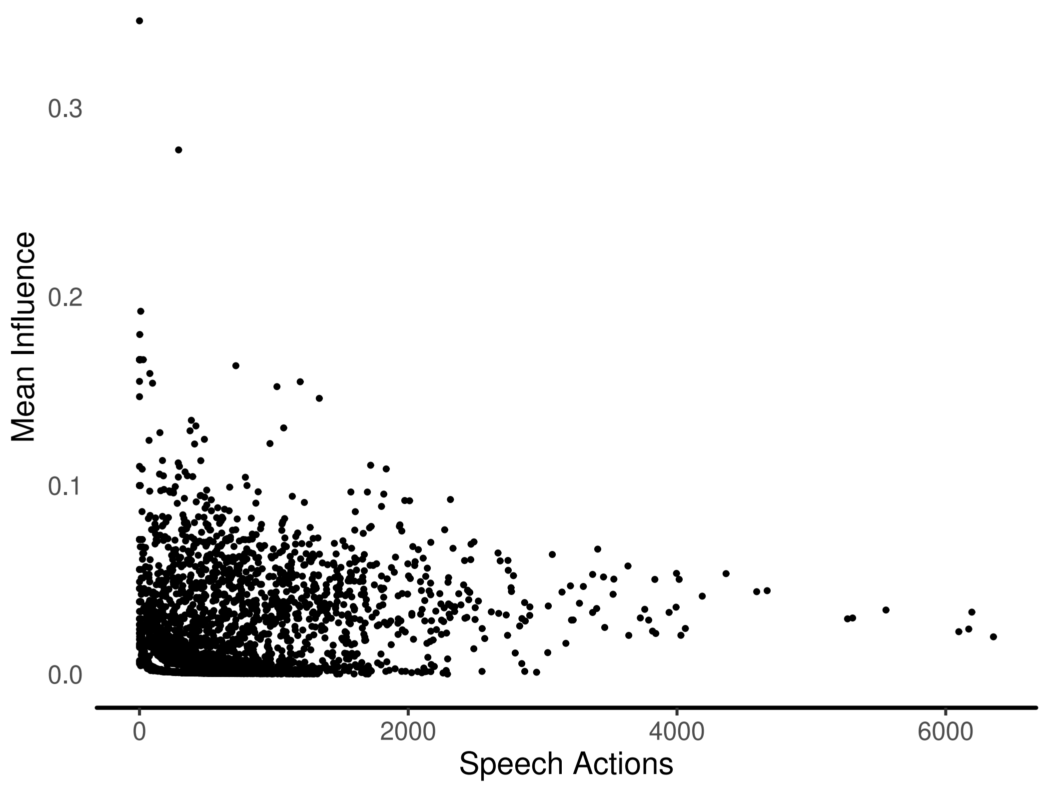


Figure 1 shows a direct comparison between the number of speech actions per member per session and their respective "effectiveness" scores. Potentially, our measure might simply capture speakers who speak more frequently, rather than those who speak more effectively. To check this possibility, we counted the number of speaking turns taken by each speaker in each Congress. Then, we correlated this count with our speaker effectiveness scores. The results of this comparison again support our expectations. Overall, the correlation between speaking turns and speaker effectiveness is 0.16, which is substantively small (though significantly different from 0, $p < 0.01$). This finding is intuitive: to be an effective speaker surely requires some non-zero quantity of speech, but higher quantity of speech should not be strongly related to speaker effectiveness.

3 Model specification

3.1 Speaker effectiveness model

As described in §5.1, we use a hierarchical Bayesian beta regression to model speaker effectiveness. We specifically rely on the `brms` library's interface to the Stan programming language (Carpenter *et al.* 2016), with the likelihood specified in the library's `Beta()` function. For priors, we use weakly informative prior distributions for all variables. These priors are intended to be uninformative when substantial data are present while preventing parameters from attaining implausible values when data are more limited. Specifically, we use $t(3, 0, 10)$ prior for all intercept coefficients, a $t(3, 0, 2.5)$ prior for all other regression coefficients, and a $Cauchy(0, 5)$ prior on all standard deviation parameters (Ghosh *et al.* 2018). During estimation, we use four chains with 1,500 warmup iterations and 1,500 post-warmup iterations in each chain, and random initializations for all parameter values. We additionally set $adapt\_delta$ value in the sampler to 0.91 to avoid divergent transitions. Visual plots suggested good mixing across chains in all models, with $\hat{R} \leq 1.01$ for all parameters and $n_{eff} \geq 1000$ for all parameters.[2]

3.2 Legislative productivity model

For the models of legislative productivity, (Table 4 in the main paper and all of section 4.2 and on in this Appendix), we use a hierarchical Bayesian negative binomial regression. We use the `rstanarm` package to interface with Stan, using the `stan.glmer()` function. Again, we use weakly informative prior distributions for all variables, specifically $N(0, 10)$ prior for all intercept coefficients, a $N(0, 2.5)$ prior for all other regression coefficients, and an $Exp(1)$ coefficient for the dispersion parameter. Similarly to the effectiveness model, we use four chains with 1,500 warmup iterations and 1,500 post-warmup iterations in each chain, and random initializations for all parameter values. The $adapt\_delta$ value in the sampler is set at 0.95. Diagnostics suggest similarly good mixing across chains in all models, with no convergence issues.

[2]With $\hat{R}$ a diagnostic quantifying the consistency of an ensemble of Markov chains, and n_{eff} a rough effective sample size calculation (Gelman *et al.* 2014).

4 Robustness checks

4.1 Topic parameter variation

Throughout this paper, we rely on the parametric SITS model variant. As we note in-text, our reasons for selecting parametric SITS are practical in nature. Due to the size of our corpus, we found that estimating a non-parametric SITS model was infeasible. Experimental results presented by Nguyen *et al.* (2012) suggest that the choice between parametric and non-parametric SITS has a limited impact on downstream model performance. However, since the parametric SITS model requires users to specify the number of topics used in the model, results generated using the parametric SITS model may be sensitive to the topic parameter.

To investigate this possibility, we re-ran the per-Congress model shown in Figure 4 with 25, 50, 100, and 150 topics. All other parameters remained the same as those described in Section 1 of this Appendix. As shown in Figure 2, all posterior means retain their signs under each topic value save for the "African American" indicator coefficient in the 150-topic model. The size of some coefficient estimates variables – including the "Committee Chair", "Senate", and "Majority" indicators – are sensitive to the number of topics included in the model, which suggests that the effect sizes we present should be interpreted with caution. However, the basic pattern of relationships that we describe is robust to selection of the topic parameter.

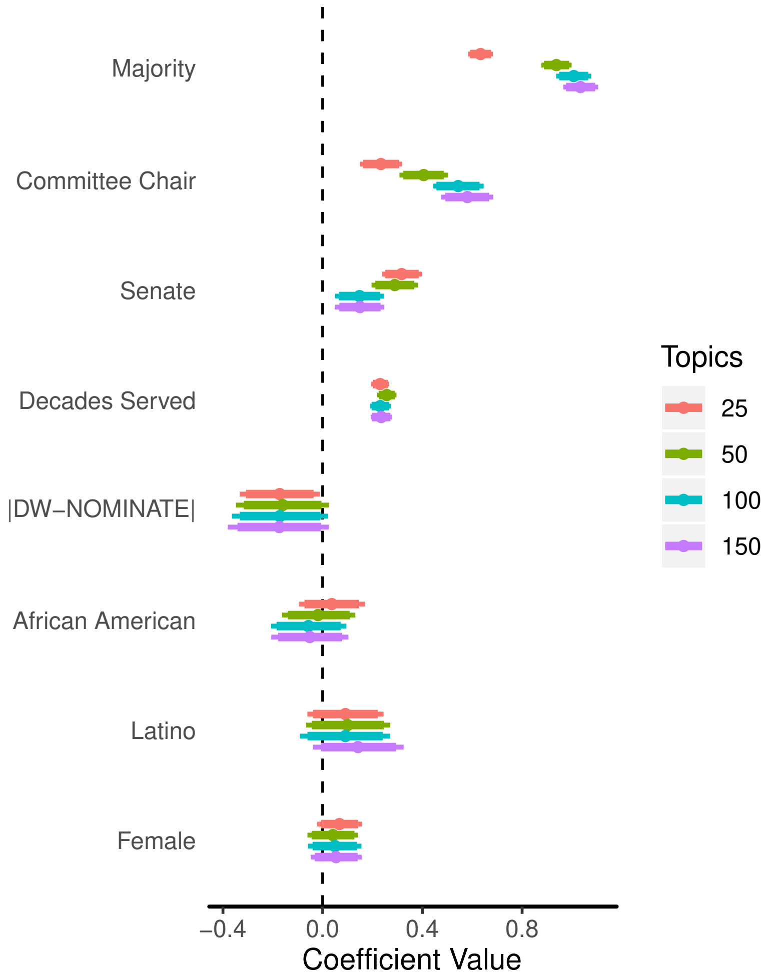
Table 2: Predicting Number of Bills Passed Through Committee, Topic Robustness

	25	**50**	**100**	**150**
Speaker Influence	0.528	0.974	1.846	1.870
	(0.216)	(0.336)	(0.639)	(0.805)
N	3560	3560	3560	3560
Full Covariates	Yes	Yes	Yes	Yes

Dependent variable is total bills sponsored by each MC in each Congress that advanced beyond the committee stage. Bayesian hierarchical negative binomial model, with intercept partially pooled by member committee assignment and Congress. Columns indicate number of topics used to estimate the mean influence score.

We also tested the robustness of our models of legislative productivity, presented in Table 4 of our main paper. Again, using the same 4 different topic distributions discussed above, we reran Model 3 from Table 4, which is the model with full covariates. All other specifications from the model remain the same. Since we are only interested in the coefficient of our measure on advancing legislation through the committee process, we only report those values in Table 2. While the value of the coefficient and the standard errors change between topic specifications, the effect is consistently positive and roughly proportional to the standard errors. All other covariates retain a similar orientation.

Figure 2: Robustness to topic parameter



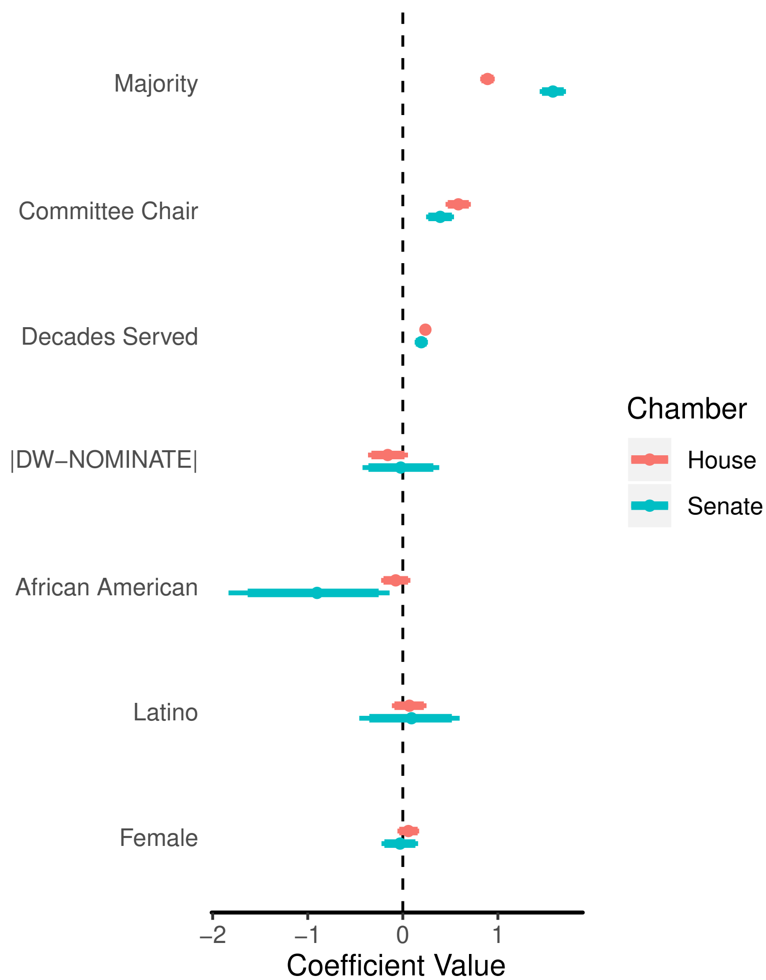
Dependent variable is the estimated influence score for each speaker in each Congress. Bayesian hierarchical beta regression model. Dots indicate posterior mean values. Thick lines indicate 90% credible intervals, and thin lines indicate 95% credible intervals. Positive estimates indicate that an increase in a given coefficient is associated with an increase in estimated speaker influence. Intercept is partially pooled by individual and by Congress, but suppressed for readability.

4.2 Chamber-specific models

Since House and Senate committees have different rules and norms for participation, the coefficient estimates from our in-text models might plausibly vary by chamber. To test this possibility, we estimated separate versions of the speaker influence model we describe in Figure 4 of our manuscript. We exclude chamber as a predictor from this model, but all other model specification details, priors, and parameter settings remain the same as those described in Appendix 1 and 5.1.

The results of this comparison are shown in Figure 3. Since the Senate model contains substantially fewer observations than the House model, the credible intervals on this model's parameter estimates are substantially larger. In addition, unlike in the House or pooled models, African American senators are estimated to be substantially less effective than their non-Latino counterparts, though since we observe a total of six African American senators this result should be interpreted with caution. The majority coefficient is also substantially larger in the Senate than the House, while the committee chair coefficient displays the opposite pattern. These differences likely result from rules differences in House and Senate committees, and represents a plausible direction for future research. However, as in the topic parameter comparison in §4.1, the basic pattern of relationships we describe is robust to chamber-level differences.

Figure 3: House/Senate model parameter estimates



Dependent variable is the estimated influence score for each speaker in each Congress. Bayesian hierarchical beta regression model. Dots indicate posterior mean values. Thick lines indicate 90% credible intervals, and thin lines indicate 95% credible intervals. Positive estimates indicate that an increase in a given coefficient is associated with an increase in estimated speaker influence. Intercept is partially pooled by individual and by Congress, but suppressed for readability.

4.3 Alternative Model Specifications

4.3.1 Committee REs vs individual REs

In this section, we replicate Figure 4 and Table 4 from our main paper, but change the structure of the models we use. Instead of including partially pooled random intercepts for each speaker, we instead include indicator variables for each speaker's committee assignment, as identified by Stewart III & Woon (2017)'s Congressional Committees dataset. Since the speaker-level random effects we include should only capture time-invariant unmodeled speaker-level characteristics, this approach might miss important variation created by changing committee assignments or other shifts in speaker status. As we discuss in-text, committee assignments are stable for most members and committee-level missingness patterns make conditioning on committee assignments potentially unreliable, which leads us to favor the speaker-level random effect approach we use in-text. However, we include versions of our main in-text models with committee assignment indicator variables as a robustness check.

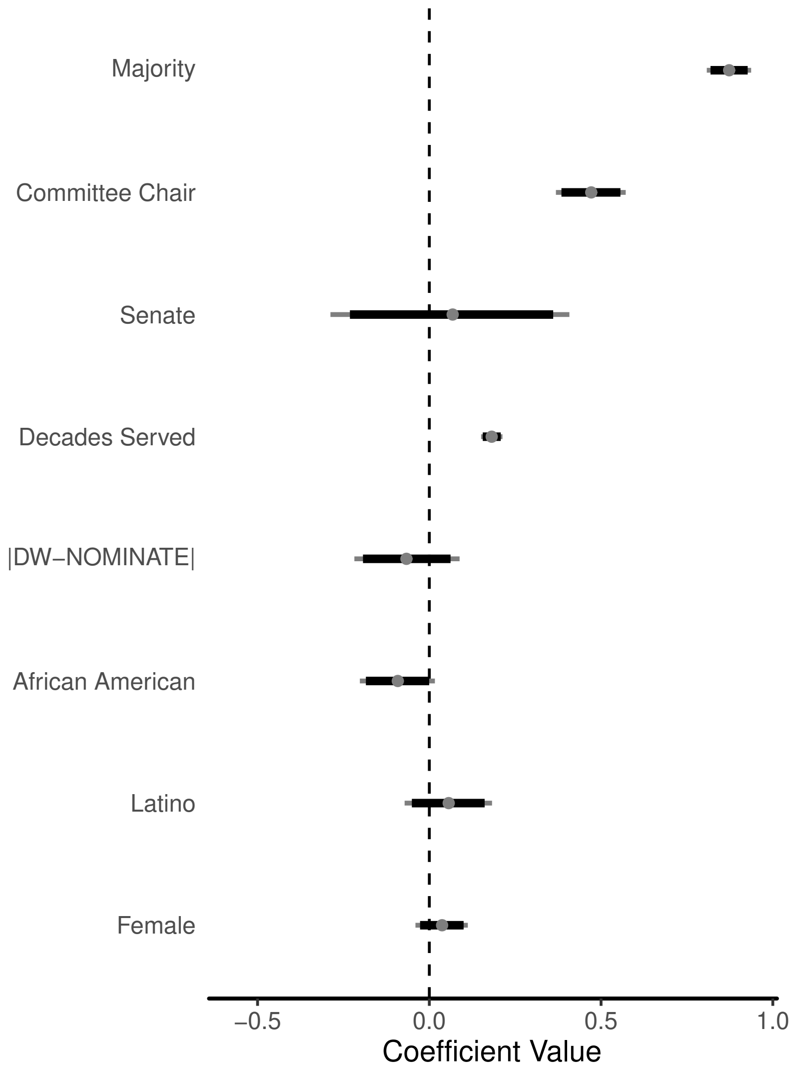
In Figure 4, we give the results for the descriptive model that correspond to Figure 4 in our paper. All coefficient results are identical in sign and similar in substantive interpretation. The one clear exception is the Senate coefficient, which now crosses zero. Since the committee status indicators are collinear with the chamber indicator outside of joint committees, these indicators likely absorb the variance previously absorbed by the chamber indicator. In addition, the absolute DW-NOMINATE coefficient in this model is slightly smaller in absolute value, which causes the 95% credible interval on this variable to cross zero. Since this coefficient was not precisely estimated in our original model, this finding reinforces the tenuous nature of our conclusions regarding the relationship between speaker ideology and effective committee participation.

Similarly, in Table 3, we replicate the legislative effectiveness model described in Table 4 of our paper with committee assignment indicator variables. Again, we observe similar coefficient estimates to those main paper, though with slightly smaller standard errors due to the smaller number of committee indicator variables compared to speaker random effects. The only significant difference between these models and the models presented in the paper can be found in Model 3, where the coefficient on |DW-Nominate D1| score tripled in size and is now negative and strongly significant.

4.3.2 Commemorative bills excluded

In Table 4, we replicate Table 4 from our main paper, but exclude non-commemorative bills, as designated by Volden & Wiseman (2014). The findings remain essentially unchanged for all three versions of the model, with a slightly higher coefficient on our measure of speaker influence in 2 out of the 3 models presented here.

Figure 4: Committee indicator variable model estimates



Dependent variable is the estimated influence score for each speaker in each Congress. Bayesian hierarchical beta regression, with intercept partially pooled by Congress. Grey dots indicate posterior mean values. Thick lines indicate 90% credible intervals, and thin lines indicate 95% credible intervals. Positive estimates indicate that an increase in a given coefficient is associated with an increase in estimated speaker influence. Intercept and committee assignment indicators are suppressed for readability.

Table 3: Predicting significant bills passed through committee with committee random intercepts

	Model 1	**Model 2**	**Model 3**
Speaker Influence	14.868	2.736	1.672
	(0.591)	(0.620)	(0.631)
Majority		0.844	0.926
		(0.038)	(0.040)
Committee Chair		1.006	0.850
		(0.055)	(0.058)
Senate			0.644
			(0.185)
Decades Served			0.157
			(0.021)
|DW-Nominate D1|			−0.483
			(0.089)
Female			0.029
			(0.044)
African American			−0.090
			(0.069)
Latino			−0.144
			(0.080)
Constant	−0.910	−1.006	−6.181
	(1.047)	(0.918)	(1.080)
N	3565	3565	3556

Dependent variable is total bills sponsored by each MC in each Congress that advanced beyond the committee stage. Bayesian hierarchical negative binomial model, with intercept partially pooled by member committee assignment and Congress.

4.3.3 "Substantive and significant" bills only

For Table 5, we use the same framework we have throughout the section, but change the DV to be only the number of "substantive and significant" legislation that pass through committee. Again, these data come from Volden & Wiseman (2014). The designation "substantive and significant" is reserved for bills that "had been the subject of an end-of-the-year write-up in the Congressional Quarterly Almanac" (Volden & Wiseman 2014, pg. 20), and is considered a more expansive form of Mayhew (1991)'s designation of "landmark" legislation. We would expect that our measure would be even more strongly associated with "substantive and significant" legislation than other non-commemorative legislation because the same underlying latent ability should be more useful that harder the task, and passing "substantive and significant" legislation is among the most challenging tasks in legislative advancement.

We find results that are consistent with the previous models, but with substantially larger coefficients on our measure. The effect of being an effective committee participant seems even more strongly associated with advancing "substantive and significant" legislation than it does any other type.

Table 4: Predicting non-commemorative bills passed through committee

	Model 1	**Model 2**	**Model 3**
Speaker Influence	16.262	2.694	1.827
	(0.761)	(0.706)	(0.704)
Majority		0.967	1.038
		(0.041)	(0.042)
Committee Chair		0.919	0.776
		(0.058)	(0.058)
Senate			0.816
			(0.060)
Decades Served			0.158
			(0.027)
|DW-Nominate D1|			−0.123
			(0.125)
Female			0.057
			(0.069)
African American			−0.256
			(0.107)
Latino			−0.094
			(0.120)
Constant	0.034	−0.376	−0.708
	(0.076)	(0.070)	(0.112)
N	3560	3560	3553

Dependent variable is total bills sponsored by each MC in each Congress that advanced beyond the committee stage, with commemorative bills (as designated by Volden & Wiseman (2014)) excluded. Hierarchical Bayesian negative binomial model, with intercept partially pooled by member committee assignment and Congress.

4.3.4 Legislative Effectiveness Scores

Our final robustness check for predicting legislative productivity is to use Volden & Wiseman (2014)'s Legislative Effectiveness Score (LES) as the stand-in for productivity. The LES is defined as "the proven ability to advance a member's agenda items through the legislative process and into law" (Volden & Wiseman 2014, pg. 20). The LES is essentially a weighted index comprised of the various stages of advancement of a given MC's sponsored bills. From Volden & Wiseman (2014):

> "*With this definition in hand, the next step in developing and assessing a measure of legislative effectiveness is to identify a series of indicators that provide information about such effectiveness. We rely on fifteen such indicators, five for each major stage of the legislative process across each of three categories of legislation. Specifically, we consider: (1) how many bills each legislator introduces (BILL), and how many of those bills (2) receive action in committee (AIC), (3) pass out of committee and receive action on the floor of the House (ABC), (4) pass the House (PASS), and (5) ultimately become law (LAW). These five indicators are constructed separately for bills that are commemorative (C), bills that are substantive (S), and bills that are both substantive and significant (SS), as will be defined below. In combination, these fifteen indicators then form the basis for each member's Legislative Effectiveness Score.*" (Volden & Wiseman 2014, pg. 19)

We use this measure as our final measure of legislative productivity. The primary difference between the

Table 5: Predicting significant bills passed through committee

	Model 1	**Model 2**	**Model 3**
Speaker Influence	29.833	7.723	4.405
	(1.840)	(1.316)	(1.356)
Majority		2.597	2.814
		(0.182)	(0.185)
Committee Chair		1.608	1.006
		(0.105)	(0.107)
Senate			0.635
			(0.123)
Decades Served			0.588
			(0.052)
|DW-Nominate D1|			0.536
			(0.282)
Female			0.068
			(0.151)
African American			−0.786
			(0.296)
Latino			−0.295
			(0.285)
Constant	−2.867	−4.481	−5.852
	(0.135)	(0.210)	(0.285)
N	3560	3560	3553

Dependent variable is the count of "substantive and significant" bills sponsored by each MC in each Congress that advanced beyond the committee stage. Bayesian hierarchical negative binomial model, with intercept partially pooled by Congress and committee assignment indicator variables. "Substantive and significant" distinction is from Volden & Wiseman (2014) and is based on whether or not a piece of legislation is referenced in Congressional Quarterly's Annual Almanac.

models below and the previous models used is that LES is a continuous variable with ratio-level qualities. As a result, we model the dependent variable in this model with a Gaussian distribution in lieu of a beta distribution, but our modeling specification otherwise remains unchanged.

Table 6 shows the results of the three models. Our results remain unchanged from the other robustness checks included in this appendix and in the main paper, though the coefficient estimate associated with our measure of effective committee participation is slightly smaller. This finding is unsurprising. Since the LES is constructed with information from two pre-Committee stages (introducing bills and getting a bill into a committee), we would expect the LES to be less strongly associated with the latent trait we measure compared with a simple count of bills with action in or beyond committee. And it is somewhat questionable if our measure would predict bill passage, since the factors that determine whether a bill becomes law are largely outside of individual members' hands.

Table 6: Predicting Legislative Effectiveness Score

	Model 1	**Model 2**	**Model 3**
Speaker Influence	13.398	2.213	1.519
	(0.754)	(0.744)	(0.750)
Majority		0.596	0.633
		(0.040)	(0.041)
Committee Chair		2.152	2.095
		(0.078)	(0.080)
Senate			−0.265
			(0.063)
Decades Served			0.153
			(0.025)
|DW-Nominate D1|			−0.063
			(0.119)
Female			0.011
			(0.066)
African American			−0.069
			(0.095)
Latino			−0.083
			(0.111)
Constant	0.717	0.463	0.291
	(0.032)	(0.032)	(0.084)
N	3560	3560	3553

Dependent variable is each MC's Legislative Effectiveness Score in each Congress. Bayesian hierarchical Gaussian regression model, with intercept is partially pooled by individual and Congress.

5 Supplemental coefficient tables

Tables 7 and 8 are the full versions of Tables 3 and 4 from the main paper, with all the covariates included.

Table 7: Committee Exile and Lagged Influence

	Model 1	**Model 2**	**Model 3**
	Exiled	**Non-Exiled**	**Both**
Speaker Influence$_{t-1}$	-0.662	5.219	4.874
	(3.050)	(0.525)	(0.480)
Exile			0.122
			(0.122)
Majority			0.840
			(0.042)
Committee Chair			0.540
			(0.066)
Female			0.033
			(0.051)
African American			-0.108
			(0.064)
Latino			0.065
			(0.076)
|DW-Nominate D1|			0.004
			(0.099)
Decades Served Service			0.120
			(0.022)
Speaker Influence$_{t-1}$*Exile			-6.847
			(3.398)
Constant	-5.030	-3.914	-4.738
	(0.123)	(0.032)	(0.091)
N	125	1933	2056

Dependent variable is the mean influence score for each member. Bayesian hierarchical beta regression with Congress random effects included. A member is considered exiled if they involuntarily changed their committee assignment between terms (from Grimmer & Powell (2013) and Powell & Grimmer (2016)).

Table 8: Predicting Number of Bills Passed Through Committee

	Total Bills Beyond Committee		
	Negative Binomial Mixed-Effects		
	Model 1	**Model 2**	**Model 3**
Speaker Influence	14.357	2.549	1.846
	(0.687)	(0.642)	(0.639)
Majority		0.828	0.889
		(0.037)	(0.038)
Committee Chair		0.889	0.763
		(0.053)	(0.053)
Senate			0.746
			(0.055)
Service			0.120
			(0.024)
|DW-Nominate D1|			−0.176
			(0.115)
Female			0.033
			(0.064)
African American			−0.129
			(0.096)
Latino			−0.085
			(0.110)
Constant	0.254	−0.095	−0.335
	(0.071)	(0.066)	(0.100)
N	3560	3560	3560

Dependent variable is total bills sponsored by an MC that advances beyond the committee stage. Bayesian hierarchical negative binomial model, with partial pooling for individual and Congress.

References

Carpenter, Bob, Gelman, Andrew, Hoffman, Matt, Lee, Daniel, Goodrich, Ben, Betancourt, Michael, Brubaker, Michael A, Guo, Jiqiang, Li, Peter, & Riddell, Allen. 2016. Stan: A probabilistic programming language. *J Stat Softw*.

Denny, Matthew J., & Spirling, Arthur. 2018. Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It. *Political Analysis*, **26**(2), 168–189.

Gelman, Andrew, Carlin, John B, Stern, Hal S, Dunson, David B, Vehtari, Aki, & Rubin, Donald B. 2014. *Bayesian data analysis*. 3 edn. Boca Raton, FL: Taylor & Francis.

Ghosh, Joyee, Li, Yingbo, Mitra, Robin, *et al.* 2018. On the use of Cauchy prior distributions for Bayesian logistic regression. *Bayesian Analysis*, **13**(2), 359–383.

Grimmer, Justin, & Powell, Eleanor Neff. 2013. Congressmen in exile: The politics and consequences of involuntary committee removal. *The Journal of Politics*, **75**(4), 907–920.

Mayhew, David R. 1991. *Divided we govern*. Yale University New Haven, CT.

Nguyen, Viet-An, Boyd-Graber, Jordan, & Resnik, Philip. 2012. SITS: A hierarchical nonparametric model using speaker identity for topic segmentation in multiparty conversations. *Pages 78–87 of: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*. Association for Computational Linguistics.

Nguyen, Viet-An, Boyd-Graber, Jordan, Resnik, Philip, Cai, Deborah A, Midberry, Jennifer E, & Wang, Yuanxin. 2014. Modeling topic control to detect influence in conversations using nonparametric topic models. *Machine Learning*, **95**(3), 381–421.

Powell, Eleanor Neff, & Grimmer, Justin. 2016. Money in exile: Campaign contributions and committee access. *The Journal of Politics*, **78**(4), 974–988.

Stewart III, Charles, & Woon, Jonathan. 2017. Congressional committee assignments, 103rd to 114th congresses, 1993–2017. *http://hdl.handle.net/1721.1/18172*.

Volden, Craig, & Wiseman, Alan E. 2014. *Legislative effectiveness in the United States congress: The lawmakers*. Cambridge university press.